frontiers | Research Topics

Psychological studies in the teaching, learning and assessment of mathematics

Edited by
Yiming Cao, Zsolt Lavicza and Shuhua An

Coordinated by
Lianchun Dong

Published in
Frontiers in Psychology
Frontiers in Education

ISSN 1664-8714
ISBN 978-2-8325-4690-1
DOI 10.3389/978-2-8325-4690-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Psychological studies in the teaching, learning and assessment of mathematics

Topic editors
Yiming Cao — Beijing Normal University, China
Zsolt Lavicza — Johannes Kepler University of Linz, Austria
Shuhua An — California State University, Long Beach, United States

Topic coordinator
Lianchun Dong — Minzu University of China, China

Citation
Cao, Y., Lavicza, Z., An, S., Dong, L., eds. (2024). *Psychological studies in the teaching, learning and assessment of mathematics*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-4690-1

Table of contents

TYPE Editorial
PUBLISHED 18 March 2024
DOI 10.3389/fpsyg.2024.1389267

OPEN ACCESS

EDITED AND REVIEWED BY
Douglas F. Kauffman,
Medical University of the Americas,
United States

*CORRESPONDENCE
Lianchun Dong
✉ Lianchun.dong@muc.edu.cn

RECEIVED 21 February 2024
ACCEPTED 05 March 2024
PUBLISHED 18 March 2024

CITATION
Cao Y, An S, Lavicza Z and Dong L (2024) Editorial: Psychological studies in the teaching, learning and assessment of mathematics. *Front. Psychol.* 15:1389267. doi: 10.3389/fpsyg.2024.1389267

Editorial: Psychological studies in the teaching, learning and assessment of mathematics

Yiming Cao[1], Shuhua An[2], Zsolt Lavicza[3] and Lianchun Dong[4]*

[1]School of Mathematical Sciences, Beijing Normal University, Beijing, China, [2]College of Education, California State University Long Beach, Long Beach, CA, United States, [3]School of Education (STEM Education), Johannes Kepler University, Linz, Austria, [4]College of Science, Minzu University of China, Beijing, China

KEYWORDS

mathematics, teaching, learning, assessment, psychological studies

Editorial on the Research Topic
Psychological studies in the teaching, learning and assessment of mathematics

1 Mathematics teachers and teaching

1.1 Mathematics teachers' knowledge and emotion

Jiang, Zhang, Jiang et al. compare how pre-service and in-service mathematics teachers understand mathematics problem solving and its teaching. They reported that pre-service mathematics teachers with postgraduate degrees did not understand mathematics problem solving significantly different from in-service mathematics teachers. They discussed the impacts of postgraduate education, teaching experiences and in-service professional development on the knowledge of mathematics problem solving as well as its teaching.

Wang L. et al. use structural equation modeling to examine how organizational support and job crafting contribute to the occupational wellbeing of the new mathematics teachers in primary and junior high schools. They found that organizational support, job crafting and basic psychological needs have prominent and positive effects on the occupational wellbeing. In addition, there exists a mediating effect of basic psychological needs with organizational support, job crafting, and the teacher wellbeing.

Wang G. et al. investigate secondary school mathematics teachers' autonomous learning capacity from four sub-dimensions: (1) development of study plans, (2) proficiency in using learning methods, (3) learning habit formation, and (4) evaluation of learning outcomes. They reported that secondary school mathematics teachers' overall autonomous learning capacity or the sub-dimensions vary significantly according to teachers' gender, age, years of teaching experience, educational stage, and location (i.e., rural or township). However, teachers with higher qualifications (e.g., doctoral degrees) and job titles (e.g., senior teachers) did not show better autonomous learning capacity than other teachers.

1.2 Mathematics instructional practices

Sim and Mohd Matore examine primary school mathematics teachers' usages of five teaching styles, i.e., personal model teaching style, expert teaching style, formal authority teaching style, delegator teaching style, and facilitator teaching style. They found that mathematics teachers' adoptions of teaching styles are weakly correlated with their teaching experiences, implying that teaching experience in mathematics might influence the adoption of teaching styles.

Jiang, Zhang, Ruan et al. conduct a comparative analysis of the extent to which high school mathematics teachers' audible teaching language was satisfied by students. They found that there were no significant differences regarding students' overall satisfaction on expert, skilled, and novice mathematics teachers' audible teaching language. However, in terms of the tone and adaptability of the audible teaching language, students were reported to be more satisfied with expert mathematics teachers than with novice teachers.

Ayabe et al. design problem-appropriate diagram instruction by integrating appropriate line diagrams, tables, and graphs into high school mathematics problem-solving instruction, and examine the effects on students' ability in mathematics word problem solving. They find that the inclusion of problem-appropriate diagrams helps to improve students' problem solving performance and to reduce perceived cognitive load in problem solving activities. The results highlight the importance of selecting relevant diagrams according to the problem-solving tasks to enhance the quality of mathematics problem solving instruction.

Zhao W. et al. conduct a video analysis of dialogue types in expert and novice mathematics teachers' lessons and presents the characteristics of effective mathematics classroom dialogues. They report some culturally unique aspects of effective mathematics classroom dialogues in Chinese contexts, such as high proportion of discourse about basic knowledge, justification, and probing. The contributes to a better understanding of effective classroom dialogues by including practices and perspectives in Chinese contexts.

2 Processes of mathematics learning

2.1 Origins of affective factors in mathematics learning

Cheng et al. investigate the factors that have impacts on university students' learning satisfaction in the blended learning mode. They find that factors in learning dimension (e.g., learning interest, learning awareness, learning self-efficacy, learning concentration, learning reflection, and the level of communication and interaction) have primary impacts on students' satisfaction with blended learning. It highlights the significance of developing students' learning attitudes in blended learning mode.

Reschke et al. make attempts to explain the gender differences in the early development of students' math self-concepts by examining teachers' judgments of students' mathematics ability and students' perception of teachers' judgments. They find that students' self-concepts can be predicted by both teachers' judgments of students' mathematics ability and students' perception of teachers' judgments. They also find the mediation effect of teachers' judgments on students' self-concepts through students' perceived teachers' judgments. They argue that primary school students tend to have a math-male stereotype when internalizing their teachers' judgments of mathematics ability, leading to girls' lower math self-concepts.

Yesuf et al. examine what factors contribute to high school students' mathematics self-efficacy. They report that living arrangements with parents (i.e., living with single parent or both parents), students' expected grade in the upcoming national exam and expected marks in the semester can significantly predict students' mathematics self-efficacy, whereas students' received tutorial, students' plan for further education and professional aspiration do not significantly influence students' mathematics self-efficacy.

2.2 Relations between affective factors and mathematics learning

Xu et al. use bibliometric analysis techniques to review the existing studies regarding growth mindset and mathematics learning. They find that there is a lack of studies focusing on student's mindset in the specific sub-fields of mathematics, e.g., algebra, calculus, or geometry. In addition, few studies addressed how teachers' mindset and parents' mindset make impacts on students' mathematics learning. The findings provide implications for future explorations about growth mindset and mathematics learning.

Dong et al. examine the functioning processes of growth mindset in mathematics learning by combining students' growth mindsets, failure attributions, intrinsic motivation, mathematics self-efficacy, mathematics anxiety and mathematics achievements in one statistics model. They find that students' growth mindset doesn't directly predict their mathematics achievements, but indirectly influences mathematics achievements through students' intrinsic motivation. Failure attributions, mathematics self-efficacy, and mathematics anxiety play sequentially mediating roles in the relation between students' growth mindset and their academic achievements.

Sakellariou examines the reciprocal relationship between students' academic self-efficacy and mathematics achievements in high school. It is reported that there are robust reciprocal effects between self-efficacy and mathematics achievement for boys, and the dominant effect is from earlier achievement to later self-efficacy. However, there are no strong evidences for such reciprocal effects in the sample of girls. This study supports the existence of higher reciprocal effects for boys than girls, providing evidences for significant gender differences in the context of high school mathematics learning.

Zuo and Wang explore the relationship between mindfulness-based intervention (MBI) and high-risk students' mathematics achievements in middle school by considering the mediating roles of mathematics-specific exam anxiety and mathematics self-efficacy. The results show that mindfulness can significantly improve students' mathematics achievements, and reduce students' mathematics anxiety, as well as problem-solving obstacles resulted from mathematics anxiety. This study provides evidences for mindfulness intervention's efficacy in promoting middle school students' mathematics academic performance.

Brumm and Rathgeb-Schnierer explore the relationship among accuracy in numerosity estimation, mathematics achievement, and mathematics interest in the sample of primary school students. The results indicate that there is no statistically significant association either between accuracy in numerosity estimation and mathematics interest or between accuracy in numerosity estimation and mathematics achievement. This study suggests the needs for further studies regarding the function of numerosity estimation in mathematics learning.

Guo et al. investigate the relations between primary school students' filial piety beliefs and mathematics procrastination by taking into account the mediating role of academic emotions (e.g., enjoyment and anxiety). The results show that, students with reciprocal filial piety tended to have fewer procrastination behaviors in mathematics learning than students endorsing authoritarian filial piety. Reciprocal filial piety leads to more enjoyment in mathematics learning, which may in turn establish a protective mechanism against procrastination. This study contributes to a better understanding of the roles of students' filial piety beliefs in mathematics learning.

2.3 Environmental factors in mathematics learning

Yu et al. compare how primary and middle school students' relationships with parents, teachers, and their peers make impacts on their academic performance in mathematics. They report that, compared with students' relationships with parents, teachers, the quality of peer relationships was more closely associated with academic achievement in mathematics. The results highlight the significant roles of peer relationships in mathematics learning.

2.4 Cognitive processes of mathematics learning

Zhao J. et al. examine students' conflict discourse in mathematics cooperative problem solving, analyzing the discourse style and language characteristics of the three stages of conflict discourse (i.e., Initial stage of conflict, Conflict and Negotiation stage, and the end of conflict stage). They identify twelve categories of conflict discourse, count each categories' frequencies, and examine the processes of how conflicts in mathematics cooperative problem solving can be intensified or resolved by different categories of conflict discourses.

Shang et al. use eye-tracking techniques to examine how structured stepwise presentations make impacts on primary school students' attention in mathematics learning and their learning outcomes in fraction. The results suggest that structured stepwise presentation can better orient student to pay attention to connecting relative elements in fraction learning and contribute to more desirable learning performance in fraction. This study supports the importance of designing structured stepwise presentations in mathematics teaching.

Wan Hussin and Matore explore the impacts of secondary school students' learning styles in mathematics (i.e., visual, auditory, and kinesthetic) on their academic procrastination in mathematics. They find that visual learning styles significantly contribute to academic procrastination in mathematics. Visual learning style requires long-term memory and strong visual skills, which may cause pressure and depression in mathematics learning and in turn lead to academic procrastination in mathematics. In addition, academic procrastination in mathematics decreased for students with kinesthetic learning style and this might be attributed to kinesthetic students' preferences to using hands-on or applications with real situations, which promote their interest and motivation in learning mathematics. This study suggests the necessity of considering students' learning styles when coping with procrastination in mathematics learning.

Jiang and Li examine how secondary school students use mathematics textbooks in mathematics learning. They report that, secondary school students used mathematics textbooks for various reasons, such as developing mathematical knowledge, skills, and abilities, but they were reluctant to believe the connection between mathematics textbooks and their mathematics exam scores. In addition, secondary school students use mathematics textbooks significantly according to their school regions, grade levels, and teachers' profiles. This study contribute to a better understanding of mathematics textbooks by considering students' perspectives.

2.5 Cultivation of high-order thinking skills

Wang T. et al. examine the how middle school students' mathematical modeling competency influence creativity. The results confirm the significant association between mathematical modeling competency and creativity, as well as the mediating roles of curiosity in the association. This study highlights the effects of mathematical modeling competency on creativity, implying the possible way of developing middle school students creativity via the improvement of their mathematical modeling competency.

Ji and Guo conduct a meta-analysis of the relationship between working memory and mathematical problem solving. The results show that, students' ability to solve dressed-up word problems are more strongly correlated with working memory than their ability to cope with intra-mathematical problems. In addition, compared with other components of working memory, the central executive function is more strongly associated with mathematical problem solving ability. Gender ratio shows significant moderating effects and the association is stronger in the samples of boys than in the sample of girls. This study helps to clarify the roles of working memory in mathematical problem solving rather than in the general mathematics learning.

Shi et al. use the functional near-infrared spectroscopy (fNIRS) technique to investigate the effects of middle school students' hands-on experience on geometry learning by considering academic level as an important impacting factor. The results show that, hands-on operation with concrete geometric manipulatives, in contrast to observation, can better enhance the activation of sensorimotor systems, which are important in geometry representation and processing and in turn contribute to better geometry problem-solving performance. This study helps to uncover the mechanism in which hand-on experience promote middle school geometry learning.

3 Assessment in mathematics education

3.1 Assessing contributing factors in mathematics learning

Piccirilli et al. assess the influence of mathematics anxiety on vocational secondary school students' calculus learning. They argue that, assessing secondary school students' level of mathematics anxiety at the beginning of the school year can help to identify the students who have the potential risk of failing in the subsequent calculus learning. The study suggests the consideration of mathematics anxiety assessment as a tool to promote secondary school students' mathematics learning.

Lin and Chen present the processes of develop the students' mathematics self-directed learning scale containing four sub-scales with 50 items and apply the scale to measure high school students' mathematics self-directed learning. The results show the existence of gender differences in mathematics self-directed learning and male students scored higher in mathematics self-directed learning than female students. However, students' mathematics self-directed learning does not increase with the grade level in high school students. This study provides a reliable and valid tool to assess students' self-directed learning in mathematics and helps to include the development of mathematics self-directed learning in mathematics curriculum reform.

Uesaka et al. use item response theory to analyze university students' acquisition and usages of different types of learning strategies, showing that the average levels of students' strategy acquisition are associated with their academic achievement ranking in university learning. This study proposes a framework for assessing university students' learning strategies' levels and suggests to provide university students with learning resources and advice according to their level of acquisition of learning strategies.

3.2 Assessing mathematics competencies

Zhang et al. utilize DINA (Deterministic Inputs, Noisy, and Gate) model to diagnose fourth grade students' mathematical ability in calculation by presenting each student's arithmetic knowledge status. They report students' mastery of ten cognitive attributes in mathematics ability and identify various cognitive error patterns and major cognitive error patterns in arithmetic. This study provides researchers and practitioners with the tool for students' diagnosis in mathematics learning, which is helpful for designing targeted remediation programs in primary mathematics teaching.

Wu et al. employ cognitive diagnostic assessment (CDA) to construct a cognitive model for middle school students' data analysis ability, allowing to accurately detect students' knowledge structure or operational skills in data analysis, and to present students' data analysis ability's learning path and learning progression. This study contributes to the inclusion of a new cognitive diagnostic perspective on the assessment of middle school students' data analysis abilities.

Meng et al. employ bi-factor theory and propose a full-information item bifactor (FIBF) model to measure primary and middle school students' mathematical ability. They also apply the FIBF model in a large-scale data set to examine the performance of the model. The results show that FIBF model is better fitting than other models such as UIRT and MIRT models, and a more reasonable interpretation can be obtained by using the ability scores from the FIBF model. This study supports the feasibility of employing the FIBF model in large-scale mathematics testing projects.

Author contributions

YC: Writing—review & editing. SA: Writing—review & editing. ZL: Writing—review & editing. LD: Writing—original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Social Science Foundation of China (Grant number CHA200260).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 24 August 2022
DOI 10.3389/feduc.2022.921844

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Chongyang Wang,
Beijing Normal University, China
Yi Wang,
Beijing Normal University, China

*CORRESPONDENCE
Yuri Uesaka
yuri.uesaka@ct.u-tokyo.ac.jp

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

RECEIVED 16 April 2022
ACCEPTED 11 July 2022
PUBLISHED 24 August 2022

CITATION
Uesaka Y, Suzuki M and Ichikawa S (2022) Analyzing students' learning strategies using item response theory: Toward assessment and instruction for self-regulated learning.
Front. Educ. 7:921844.
doi: 10.3389/feduc.2022.921844

Analyzing students' learning strategies using item response theory: Toward assessment and instruction for self-regulated learning

Yuri Uesaka[1]*, Masayuki Suzuki[2] and Shin'ichi Ichikawa[1]

[1]Graduate School of Education, The University of Tokyo, Tokyo, Japan, [2]College of Education, Yokohama National University, Yokohama, Japan

Many previous studies related to self-regulated learning have demonstrated that students with higher academic achievement tend to use varied and effective learning strategies. However, they cannot acquire these strategies in a short period. If learning strategies are acquired and used gradually, it may help specify individual levels of use and teach its usage in a reasonable order. Therefore, this study analyzed data on strategy use with item response theory (IRT) to specify students' level of strategy acquisition and determine how they are ordered. Questionnaire data on the frequency of learning strategy use were obtained from students in five universities (N = 472) and analyzed using the IRT model. It was shown that the average levels of students' strategy acquisition were related to their universities' academic achievement rankings.

KEYWORDS

learning strategy use, item response theory (IRT), self-regulated learning, individual differences, students' learning strategy acquisition

Introduction

Life in modern society is highly unpredictable, and people need to be proactive in deciding their behavior. Therefore, fostering self-regulated learners is one of the most important goals in education. Self-regulated learning is defined as involving "metacognitive, motivational, and behavioral processes that are personally initiated to acquire knowledge and skills, such as goal setting, planning, learning strategies, self-reinforcement, self-recording, and self-instruction" (Zimmerman, 2015, p. 541). This implies that these learners have sufficient skills to earn a living or acquire the necessary competencies. To deal with society's unpredictability and decide how to live proactively, students need to develop self-regulated learning skills. This idea is consistent with discussions on the importance of developing learning skills from the perspectives of key competencies, proposed by the Organization for Economic Cooperation and Development (OECD; Rychen and Salganik, 2003) and from the "21st century skills" movement (e.g., Griffin et al., 2012). The key competencies and 21st century skills are related to the cultivation of necessary student skills that will be useful in and out of school and in future contexts.

In considering how self-regulated learners can be cultivated, an important aspect is promoting the use of effective learning strategies (e.g., Uesaka, 2012; Manalo et al., 2017). Many reviews and theoretical papers in the self-regulated learning research area have shown (Zimmerman and Schunk, 1989; Schunk and Zimmerman, 1994, 1998; Dignath et al., 2008) that these learners use effective strategies that might positively affect life-long learning. Based on this idea, various studies have examined factors that promote learning strategy use (Ames and Archer, 1988; Nolen, 1988; Zimmerman and Martinez-Pons, 1990; Lynch, 2006) and proposed educational methods that enhance students' skills by promoting the effective use of these strategies (Harris and Graham, 1999; Cleary and Zimmerman, 2004; Guthrie et al., 2004; Camahalan, 2006; Ramdass and Zimmerman, 2008).

If learning strategy use can be considered an important skill for students to become self-regulated learners, identifying the level of learning strategy use can help in considering how to promote their development. For example, as described later, some studies on knowledge acquisition with worked examples, proposed and examined the necessity of specifying students' knowledge level, which has been used to improve students' knowledge acquisition (e.g., Schwonke et al., 2007; Salden et al., 2010). To achieve these goals, computer-adaptive testing (CAT) has confirmed the effectiveness of many good practices and methodological bases of statistics for analyzing student levels. Item response theory (IRT) is an effective method for analyzing students' levels. For example, their level of English skills in second-language acquisition was analyzed using this method (e.g., Min and Aryadoust, 2021). Previous research in self-regulated learning has not applied such an idea to analyze individual differences in learning strategy use. However, specifying them regarding learning strategies using this method may facilitate the development of more effective instructional strategies for cultivating self-regulated learners.

In detail, learners may be more likely to acquire learning strategies more effectively if they receive advice and guidance that is aligned with their specific learning strategy acquisition. In fact, learning strategies range from relatively easy to difficult, and if they are not taught according to their acquisition level, they are likely to be either too easy or too difficult for the learner to acquire. However, many learning strategy teaching studies, as discussed below, teach the same strategy(ies) to participants, and such level-appropriate teaching is not currently possible. To operationalize learning strategy support in CAT, research is required to identify the underlying learning strategy and learner level. However, to date, no published study has provided the basis for operationalizing the support of such learning strategies in CAT.

The level of acquisition of the learning strategy of a student needs to be identified to take it into account for future initiatives and to advise on appropriate learning strategies. Furthermore, the difficulty level of the learning strategy also needs to be identified (for example, which learning strategies are acquired first and which later). Additionally, it is necessary to map that level to the learner's level of acquisition of the learning strategy. If a new framework is proposed that allows this, it could enable the guidance of strategy instruction by the learner's level of acquisition of the learning strategy. Therefore, this study adopted IRT analysis, which can deal with the difficulty level of the learning strategies and the learner's level of acquisition on a unified scale.

In conducting this study, four considerations were taken into account. First, as this is the first study to determine the extent to which effective learning strategies are acquired, learning strategies that may not necessarily be initially effective were included. Learning strategies such as simple repetition are known to be less effective than those that use metacognition or elaboration organization (for example, Ramsden, 1988); however, the possibility remains that these strategies may be used in the early stages of learning. Initially, they should be seen as integrative, and if they are qualitatively different, a procedure should be followed to exclude them. Second, while using a theoretical framework, we analyzed IRT by preparing items that were in line with specific subject areas. Theoretical strategy studies are often worded in such a way that they are not subject-specific, as will be discussed later in this section. However, it is easier for learners and teachers to give advice on specific learning strategies if they are structured in a subject-specific manner. Therefore, it was decided to list and use learning strategies based on theory but also in a form that is specific to the subject areas.

Third, when analyzing the students' responses to the use of the learning strategies with IRT, we analyzed them with both polytomous and binary IRT and compared the two. Whereas polytomous IRT is an in-depth analysis that uses all the information, it is not always easy to interpret when considering the use of learning strategies. Therefore, if the results of both were found to be similar, we decided to report the results in the binary type, which is more commonly used in IRT. Finally, for this study, participants were drawn from higher-, middle-, and lower-ranked Japanese universities. University level is indicative of academic achievement in a broad sense. Even if it is difficult to verify the nature of the relationship between individual performance data and the level of learning strategies identified in this study, it is possible to argue that learning styles may underlie academic performance if even a small relationship can be demonstrated.

Variation of effective learning strategies and possibility of cultivation

In a seminal study on learning strategy use conducted by Marton and Säljö (1976), the authors interviewed university students to understand their learning processes and discovered two types: surface-level processing and deep-level processing. The former includes simple repetition (such as repeated reading and writing to remember the material), whereas the latter employs extra activities to improve the understanding of the material (such as elaboration of information). Successive studies have shown that learners who use deep-level processing strategies perform better (e.g., Ramsden, 1988).

Since the above-mentioned research, various effective learning strategies have been identified by questionnaire-type studies (Weinstein and Mayer, 1986; Pintrich and de Groot, 1990; Zimmerman and Martinez-Pons, 1990). Of these, the Motivated Strategies for Learning Questionnaire is one of the most well-known and frequently used methods (Pintrich et al., 1993). Prior studies have also attempted to categorize learning strategies. For example, Pintrich et al. (1993) proposed dividing them into three categories: cognitive, metacognitive, and resource management strategies. Accordingly, this study also used the framework of effective strategies considering these three main categories.

The two types of learning strategies categorizations are surface-level processing/deep-level processing and cognitive/metacognitive/resource management. These two categories are closely connected and, therefore, are not independent frameworks. Murayama (2007) indicated that cognitive strategies can be divided into two types: shallow (e.g., rehearsal and repetition) and deep cognitive strategies (e.g., elaboration and organization). Their structure is shown in **Table 1**. Shallow-level processing is a learning method that emphasizes simple repetition (Murayama, 2007). By contrast, here, deep-level processing is a learning method that is based not on simple repetitive processes but rather on cognitive devices in some sense. For example, it includes deep cognitive learning strategies that focus on understanding meaning in relation to existing knowledge, metacognitive strategies that focus on one's own cognitive state, and learning methods that utilize external resources such as other people and diagrams.

Whereas the structure of these learning strategies provides useful information, it is unclear which type is best in the early stages of learning and which in the later stages. Although many studies that have identified these structures also consider learning strategies in a domain-general manner (for example, "When reading I try to connect the things I am reading about with what I already know"), it is difficult to ascertain solely from these items what specific learning strategies should be developed in each subject area.

Another branch of learning strategy research has revealed other variables related to its use (Ames and Archer, 1988; Nolen, 1988; Shell and Husman, 2008; Tabachnick et al., 2008; Kesici et al., 2009). Furthermore, concrete programs to cultivate learning strategy skills to enhance student performance in a specific subject have been conducted. Reading literacy is one of the most active research areas. Concept-oriented reading instruction, proposed by Guthrie et al. (1998, 2004), and self-regulated strategy development, proposed by Sawyer et al. (1992) and Harris and Graham (1999), were seminal works. For example, Guthrie et al. (2004) focused on promoting reading strategies among third-grade students across 12 weeks and demonstrated that their program enhanced reading performance, motivation, and strategy use among the participants. In addition to reading, strategy training has been provided in other subject areas; for example, in mathematics. Ramdass and Zimmerman (2008) provided self-correction strategy training for middle school students and demonstrated improvement in the accuracy of self-evaluation of efficacy and mathematics performance in division learning. These studies showed that learning strategy use can be enhanced by implementing thoughtful programs that are carefully planned by researchers.

Whereas these studies are unique in that they examine the promotion of specific learning strategies in a subject area, leading to actual strategy support, the idea of varying study skills interventions according to a student's level of learning strategy has been overlooked. Most studies that have supported learning strategies, even with targeted groups of participants, have provided the same instructions and materials to all participants (Sawyer et al., 1992; Guthrie et al., 1998, 2004; Harris and Graham, 1999). However, students' spontaneous use of learning strategies differs considerably according to their learning level. Students in the early stages of learning development might use easier strategies, whereas those in the latter stage might use more complex strategies. This was demonstrated by Uchida (2021), who showed that early stage learners mainly used shallow-level strategies in vocabulary learning, whereas latter stage learners used deeper-level strategies. Therefore, students might be placed on a continuum from early to late-stage acquisition of learning strategy use. However, none of the previous studies have been based on the idea of identifying the level of learning strategies within a single scale, and then identifying the learner's level within that same scale. This would make it possible to support level-based learning strategies. Therefore, in this study, based on the structure of **Table 1**, learning strategies for specific subjects are placed on the same scale. In addition, we aim to position the learners on the scale by using IRT, which will be described later. This will make it possible to provide support according to the level of learning strategy.

Computer-adaptive testing and adjusting instruction according to individual levels by item response theory

Although learning strategy research has not analyzed students' levels of learning strategy use, other research areas have conducted such an analysis to adjust for instructions and individual learning circumstances. For example, CAT is a well-known technique for language acquisition among second-language learners, in which items are adapted according to test-takers' responses; for example, following a student's response to the first item (e.g., failure or success in solving the first task), the test identifies the acquisition level and provides one of several alternative types of tasks as the second item. This is rooted in IRT, a unique test theory that calculates individual knowledge levels and hence item difficulties as probabilities based on item responses.

To explain IRT in more detail, the knowledge acquisition dimension is represented by θ, and test responses to each item are modeled by a mathematical item response function; using students' responses to each item, their acquisition levels (called the "trait value" in IRT) are also specified. In the two-parameter model, the function in equation (1) was used. In this analysis, student knowledge acquisition and item difficulty levels are represented using θ. In the two-parameter model, parameter b_i is called "item difficulty" of item i and is defined as the θ score of a population in which 50% of participants pass the item. Higher scores indicate more difficult items. Parameter a_i is called "item discrimination," of item i and refers to the θ score that can be most effectively distinguished using this item. D is a constant 1.702 to approximate the curve to the cumulative normal distribution (see de Ayala, 2009, for basic information on IRT). When analyzing with IRT, a graph of the relationship between θ and the percentage of correct responses (the item characteristic curve, ICC) is drawn for each item. IRT was developed based on the use of binary data, and in the case of binary data, a single ICC is drawn for each item. However, an IRT for analyzing multi-valued data has been proposed (for example, Tang, 1996). In the polytomous-type IRT for analyzing multi-valued data, multiple ICCs are drawn for each item for each response value, such as 1, 2, 3, or 4.

$$p_i(\theta) = \frac{1}{1 + exp(-Da_i(\theta - b_i))} \quad (1)$$

Computer-adaptive testing, as well as various other tools and studies, has used IRT as an assessment of competencies. For example, in the OECD's Programme for International Student Assessment, students' achievement levels were represented by IRT analysis-based scores. Similarly, Brackenbury et al. (2017) developed a screening test using IRT to detect speech-sound disorders, and Balsis et al. (2017) utilized it to conduct complex personality disorder research. IRT also holds promise in many other areas.

A unique feature of IRT is that item difficulty and participant level can be represented by the same dimension (θ). In other words, it is characterized by the ability to select items based on the learner's level. With this feature, it can be used it as a basis for determining which item (in this case, the learning strategy) would be appropriate when teaching a learner at a particular level.

Overview and questions of the present study

This study aims to provide a foundational framework toward a future in which learners can be supported according to their level of use of learning strategies. Currently, the idea of adjusting instruction and assessment according to the level of the student is uncommon. However, identifying students' level of mastery of learning strategies and adjusting instruction accordingly may be important for improving effective learning strategy instruction to foster self-regulated learning.

The IRT framework allows for a unified discussion of the relationship between the difficulty level of a learning strategy and the learner's level of mastery of the learning strategy; when IRT identifies the difficulty level of a strategy and the student's learning strategy level on the same dimension, the result is that appropriate learning strategies can be taught. Therefore, we aim to identify the difficulty level of the learning strategy as a continuum using IRT and represent the learner's use of the learning strategy on the same dimension. These can be used to coordinate instruction to promote the use of learning strategies in the future.

In addition to these objectives, this study examined whether there is a relationship between the level of strategy use and academic performance, which would validate the use of the continuum of individual differences obtained in the IRT analysis in educational settings.

The following steps were taken in this study. First, a list of effective learning strategies for specific subjects was made, using the framework in **Table 1**. If questionnaires were already available, they were used, and if they were missing, we added them. At this stage, even shallow-level learning strategies were included, as they may be used in the early stages of learning. Factor analysis would then be used to confirm the unidimensionality that is the premise of the IRT analysis. If the results of the factor analysis indicated that the shallow-level learning strategies were qualitatively different from the deep-level learning strategies, only deep-level learning strategies were used and shallow-level learning strategies were excluded. If the unidimensionality was not confirmed and the learning strategies were separated by subject area, IRT was used for each one. However, if the items were not grouped by subject but constituted a single dimension, we assumed that effective learning strategies were unidimensional at a deep cross-subject area and proceeded to the next step in the analysis. Second, IRT was applied to items that were confirmed to have a one-factor structure, and acquisition scores were calculated to represent each participant's strategy acquisition level (θ). When analyzing the results with IRT, both polytomous and dichotomous IRT analyses are conducted; if the two results are compared and it is confirmed that there is no significant difference between those two, the results analyzed with dichotomous IRT are reported, giving priority to ease of interpretation. Finally, the relationship between academic achievement and participants' acquisition scores was analyzed. If students who showed higher academic performance demonstrated higher acquisition scores, the strategy acquisition level might be valid for predicting underlying high academic performance. To this end, in this study, students were reminded about the learning methods used in high school. In Japan, the university entrance examinations are very competitive. Students need to use effective learning strategies acquired during their high school years to achieve academic excellence. As Shimoyama (1984) argues, Japanese students select universities that match their academic abilities. Therefore, this study examined learning strategies during high school and university levels to test their validity.

Terminology used in the current study

As shown in **Figure 1**, terminologies different from the traditional IRT framework were used. While IRT analysis has primarily been used for academic achievement tests, some terminologies do not fit the context of learning strategy use. Thus, certain phrases were modified to suit the context of the study. For example, the traditional term "trait value" for the abscissa (θ) was replaced by "strategy acquisition level," which represents the extent to which each student acquired higher strategies. The calculated values representing the students' levels were called "strategy acquisition scores."

We retained the same term, "item difficulty" according to the IRT framework, for parameter *b*. Theoretically, we perceived it to represent an "order of acquisition" (the ranking of strategy items on the θ scale in the IRT analysis) of each learning strategy. However, IRT analysis is not evidence of the temporal and developmental order of strategy acquisition. Therefore, the same term for "item difficulty" was used.

Methods

Participants

To cover a wide range of learning strategy levels of Japanese university students, participants were drawn from a balanced mix of Japan's higher-, middle-, and lower-ranked universities. A total of 472 Japanese university students (female = 180, male = 275, unknown = 17) from five universities in Tokyo and Kanagawa prefectures (mean age = 19.81 years, *SD* = 2.07) were recruited. The universities represented different rankings in Japan; Universities A and B were part of the higher-ranked universities, Universities C and D were in the middle ranks, and University E was in the lower-ranked category (Benesse Corporation, 2017). The number of students in each school was as follows: University A had 165 students, University B had 67 students, University C had 59 students, University D had 127 students, and University E had 54 students. All participants were recruited from educational psychology courses, as these courses are compulsory for students aiming for teacher qualification. Furthermore, those studying educational psychology belong to diverse disciplines, from science and technology to the humanities and economics. All students in this study participated voluntarily. We obtained ethics approval to conduct this study from the Life Sciences and Research Ethics and Safety Committee of the University of Tokyo (approval number: 18–16).

Materials and procedure

As stated at the beginning of this manuscript, the purpose of this study is to create a foundation for specifically advising students on what learning strategies to learn next, depending on their level of learning strategy. To achieve this objective, it is necessary to list specific subject learning strategies while maintaining a theoretical foundation. Therefore, we have listed specific subject learning strategies, using the structure of **Table 1** as a foundation. Specifically, we took five subjects (mathematics, English, Japanese, social studies, and science) that are relatively commonly taught in Japanese junior high and high schools and selected items in an unbiased manner, considering how the learning strategies of cognition, metacognition, and resource management can be embodied in each. In selecting items, we referred to previous studies on learning strategies (for example, Inuzuka, 2002; Murayama, 2003; Ichihara and Arai, 2006; Oyama, 2009), case reports on cognitive counseling, and educational practices using psychology (for example, Ichikawa, 1998; Uesaka et al., 2017). Where studies could not be found, we added items ourselves, taking theory into account.

Table 2 lists labels, definitions, and specific examples for each of the subcategories in **Table 1**. The first group is the (shallow) cognitive learning strategy group, which is a learning method that emphasizes simple repetition (Murayama, 2007). Examples of the items are "When trying to remember English words, I memorized them by writing them repeatedly" (English) and "When I learned mathematics, I remembered the procedure as much as possible" (mathematics). The second group is the (deep) cognitive learning strategy group, which is a learning method that focuses on understanding meaning in relation to existing knowledge (Murayama, 2007). Strategies such as "elaboration" and "organization" as discussed in psychology are also included in this category. Examples of the items are "When trying to remember mathematical formulas, I considered how to get the formula" (mathematics) and "When I learned social

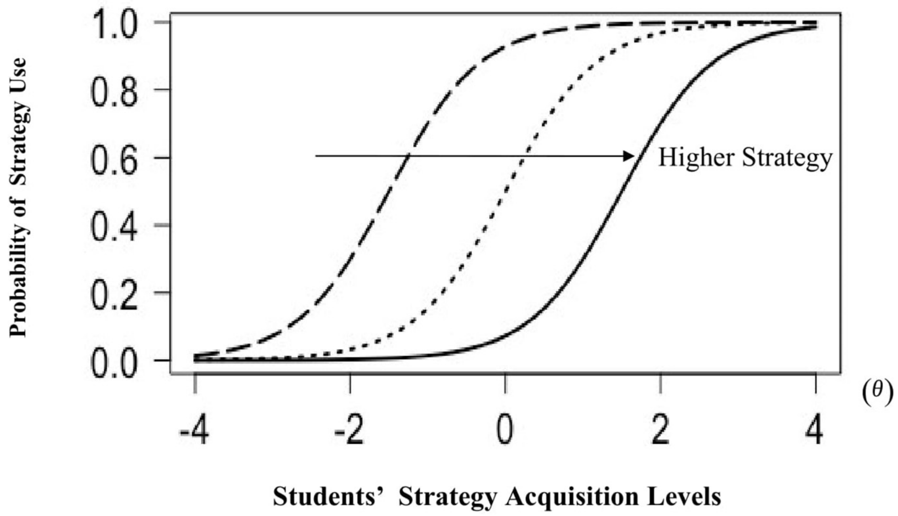


FIGURE 1
Basic framework and terminology.

science, I considered why such a historical event happened" (social science). The third group is the metacognitive learning strategy group, which is a learning method that monitors one's own cognitive state and controls the learning behavior based on it. Examples of the items are "When I learned science, I tried to make clear the points that I did not understand" (science) and "When learning mathematics, to prevent failure, I wrote points for attention in my notebook" (mathematics). The final category is the external resource utilizing strategy, which is a method of learning that involves actively using things and people around oneself. Examples of the items are "When I found difficult words while reading (Japanese), I looked them up in the dictionary" (language learning), and "I tried to participate in scientific learning programmes held in the area or in the school" (science). All items used are listed in **Table 3**.

Participants answered the questionnaires on a four-point Likert-type scale to reflect their learning strategies in high school: "1: do not use at all," "2: use sometimes," "3: use often," and "4: always use." Another choice, "not taking the course," was also included for those who had not taken a course in the subject area.

The questionnaire included other items that were not relevant to this study's aim, such as items on beliefs about learning. Along with these and the focal items, the booklet additionally asked the participants to report to their university to check the validity of the connection between the level of strategy acquisition and academic achievement. Participants responded at their pace; the questionnaire took approximately 20–30 min to complete.

TABLE 1 Relationship between the two models.

Type of processing	Type of learning strategy
Shallow-Level Processing	(Shallow) Cognitive Learning Strategy
Deep-Level Processing	(Deep) Cognitive Learning Strategy
	Meta-cognitive Learning Strategy
	Resource Management Strategy

Results

Results of factor analysis to check one-dimensionality

A factor analysis was preliminarily conducted to choose one-dimensional items before the IRT analysis. R 3.4.4 with the "psych" package was employed to perform the factor analysis (Revelle, 2018). Two items that many participants responded to without having taken the course (over 10%) were excluded.

A factor analysis was conducted to determine the items to be analyzed in the IRT. Specifically, we tested the possibility that the items of the learning strategies prepared for this study would be grouped by subject area, theoretical grouping, and shallow and deep level, or as one cohesive whole. The decay of eigenvalues was as follows: 10.09, 5.10, 2.99, 2.67, 1.97, and so on. Harris–Kaiser independent cluster rotation (Harris and Kaiser, 1964) was used as the rotation method to obtain a simple structure. As shown in **Table 3**, a two-factor model was understandable from the perspective of theories and findings in cognitive psychology. As shown in **Figure 2**, the results of the factor analysis demonstrated that learning strategies do not cohere by five subject areas. It was also shown that they do not cohere according to theoretical groupings (shallow cognitive learning,

TABLE 2 Subcategories, definitions, and examples of items used in this study.

Subcategories	Definition	Examples of items
(Shallow) Cognitive Learning Strategy	A learning method that emphasizes simple repetition.	-"When trying to remember English words, I memorized them by writing them repeatedly" (English). -"When I learned mathematics, I remembered the procedure as much as possible"(mathematics).
(Deep) Cognitive Learning Strategy	A learning method that focuses on understanding meaning in relation to existing knowledge. Strategies such as "elaboration" and "organization" as discussed in psychology are also included in this category.	-"When trying to remember mathematical formulas, I considered how to get the formula" (mathematics). -"When I learned social science, I considered why such a historical event happened" (social science).
Metacognitive Learning Strategy	A learning method that monitors one's own cognitive state and controls learning behavior based on it.	-"When I learned science, I tried to make clear the points that I did not understand" (science). -"When learning mathematics, to prevent failure, I wrote points for attention in my notebook" (mathematics).
External Resource Utilizing Strategy	A way to learn by actively using things and people around oneself.	-"When I found difficult words while reading (Japanese), I looked them up in the dictionary" (language learning). -"I tried to participate in scientific learning programmes held in the area or in the school" (science).

deep cognitive learning, metacognitive learning, and resource managing strategies). However, it is difficult to understand them as a single factor as a whole; it is more appropriate to understand them as two cohesive groups: deep-level and shallow-level learning strategies, as will be discussed below. Therefore, two factors were considered appropriate and adopted in this study.

A more detailed look at the results of the factor analysis shows that items that load highly on the first factor are generally deeper-level learning strategies. For example, "When trying to remember mathematical formulas, I considered how to get the formula" and "When I learned social science, I considered why such a historical event happened" were among the items categorized as deep cognitive learning strategies; "When I learned mathematics, I tried to conside why I failed to solve the problems" and "When I read (Japanese), I set myself some questions to check my understanding" were metacognitive learning strategies; and items such as "When I read English text, I looked words up in the dictionary if I found unknown words" and "When I found unclear points when learning science, I asked my friends and teachers" were categorized as utilizing external resource learning strategies. As indicated in the classification in **Table 1**, these are all positioned as deep-level learning strategies.

However, items that load highly on the second factor are deemed shallow-level learning strategies. For example, "When I learned science, I memorized the content without considering the reason" and "When I learned history, I memorized years of historical events or the events." As shown by Ramsden (1988), deep and metacognitive learning strategies are more effective than shallow ones. The correlation between the first and second factors was very low (r = 0.02).

The results indicate that shallow-level learning strategies are qualitatively different from deep-level learning strategies. Therefore, since the purpose of this study was to identify the level of effective learning strategies, rather than using a shallow-level in the early stages of learning that gradually transforms into a deep level, shallow-level learning strategies that were determined not to constitute unidimensionality with deep-level learning strategies were excluded from subsequent analyses.

Again, the factor analysis was also conducted for the 45 items loading on the first factor, with three excluded due to the small factor loadings (the criterion for exclusion was set as a factor loading value under 0.25). The decay of eigenvalues was as follows: 9.84, 2.82, 2.41, 2.07, 1.84, 1.65, and so on. The factor loading values, which are highly correlated with the first-factor loading values obtained from all items, are shown in the right column of **Table 3**. Thus, only those items were used for subsequent IRT analysis. The alpha coefficient of these 45 items was 0.92, indicating high internal consistency.

Item response theory analysis with learning strategy use data

Based on the results of the factor analysis, which showed that items loading on the first factor were deep-level processing learning strategies, 45 items with high loadings on the first factor were used in the IRT analysis. The analysis used both binary and polytomous IRTs. When using a binary IRT, the Likert-type data were converted into binary data. If participants answered "3: use often" or "4: always use" for a given learning strategy, it was coded as "1"; all other responses were coded as "0." The analysis of both IRTs was conducted using R 3.4.4 and the "irtoys" package (Partchev et al., 2017).

Correlations between students' mastery scores obtained from the binary IRT and polytomous IRT were analyzed. The results are presented in **Figure 3**. As shown in the results, the acquisition scores obtained from both analyses are very similar. However, in the case of polytomous IRT,

TABLE 3 Results of factor analysis and basic statistics.

	Item	*M*	*SD*	Factor loading		Results analyzing only one factor
				F1	F2	
59	When I learned science, I tried to consider why I failed to solve problems.	2.59	1.02	0.67	−0.14	0.66
56	When I learned science, I tried to make clear the points that I did not understand.	2.59	0.93	0.64	−0.06	0.63
61	When I learned science, I tried to focus on learning parts I did not understand.	2.62	0.97	0.62	0.05	0.61
22	When I learned mathematics, I tried to consider why I failed to solve problems.	2.91	0.92	0.60	−0.22	0.59
57	When I found new concepts in learning science, I considered the definition and example.	2.52	0.94	0.59	−0.19	0.58
58	When I learned science, I summarized what I learned with tables and figures.	2.50	1.03	0.58	−0.04	0.58
17	When I found maths problems I could not solve, I tried to do so by drawing diagrams.	2.77	1.01	0.57	−0.25	0.55
18	When I learned mathematics, I tried again to solve the problems that I had failed to solve.	2.89	0.94	0.56	−0.10	0.55
64	When I found unclear points when learning science, I asked my friends and teachers.	2.76	1.01	0.54	0.11	0.54
48	When I learned social science, I considered why such a historical event happened.	2.60	1.03	0.53	0.02	0.54
43	When I learned social science, I tried to consider how different events connect to each other.	2.62	1.06	0.53	0.04	0.54
46	When I read textbooks concerning social science, I tried to find unclear points.	2.34	1.02	0.53	0.22	0.54
7	When I read English text, I paid attention to points where I got confused.	2.44	0.95	0.51	0.05	0.52
34	When I read (Japanese) text and found a new word, I tried to consider its meaning.	2.71	0.84	0.50	−0.01	0.50
19	When trying to remember mathematics formulas, I considered how to get the formulas.	2.50	0.98	0.50	−0.37	0.48
49	When learning social science, I applied mental resources to points I did not remember.	2.71	0.96	0.49	0.23	0.49
45	When I learned history, I summarized how events happened in my notes or my mind.	2.76	1.05	0.47	0.13	0.48
15	When trying to remember mathematics formulas, I memorized situations where they were used.	2.65	0.96	0.47	−0.10	0.46
37	When I read (Japanese) text, I read slowly if I found unclear points.	3.04	0.92	0.45	0.05	0.45
8	When trying to remember English words, I concentrated on words I did not remember.	2.77	1.01	0.45	0.19	0.45
4	When trying to remember English words, I paid attention to situations where the word was used.	2.38	0.94	0.44	−0.28	0.43
30	When reading (Japanese) text, I tried to summarize the meaning of each sentence.	2.44	0.96	0.43	0.09	0.44
6	When I found unknown words, I took a guess based on the context where they were used.	2.85	0.97	0.43	−0.11	0.43
32	When I read (Japanese) text, I considered how content related to knowledge I already had.	2.37	0.95	0.43	0.04	0.44
63	When learning science, I used study aids.	2.52	1.11	0.42	0.03	0.42
23	When learning mathematics, to prevent failure, I wrote points for attention in my notebook.	1.95	0.95	0.42	0.03	0.42
35	When finding difficult words in (Japanese) text, I paraphrased using easy expressions.	2.47	0.93	0.41	0.03	0.41
24	When I found unclear points while learning mathematics, I asked my friends and teachers.	2.86	0.97	0.39	0.18	0.39
36	When I read (Japanese), I set myself some questions to check my understanding.	1.77	0.84	0.38	0.18	0.39
20	When I learned mathematics, I checked whether answers were correct after solving.	3.11	0.98	0.38	−0.23	0.37
5	When learning English, I remembered words by identifying prefixes and suffixes.	2.39	1.06	0.38	0.05	0.37
44	I talked with my friends or family about history.	2.39	1.11	0.37	0.19	0.38
29	When I found difficult words while reading (Japanese), I looked them up in the dictionary.	2.73	0.96	0.37	0.10	0.37
13	When I learned mathematics, I used study aids.	2.43	1.02	0.36	0.17	0.37
47	When I learned history, I tried to understand the big sweep of time rather than the details.	2.83	0.97	0.36	0.19	0.37

(Continued)

TABLE 3 (Continued)

	Item	*M*	*SD*	Factor loading		Results analyzing only one factor
				F1	F2	
60	When I learned science, I watched TV programs or read journal concerning science.	2.01	1.05	0.36	0.05	0.35
31	When I found unclear meanings in conversation with others, I asked questions about them.	2.27	0.95	0.34	0.13	0.34
9	When I read English text, I looked words up in the dictionary if I found unknown words.	2.87	0.94	0.33	0.19	0.33
28	When trying to remember Chinese characters, I memorized similar or opposite ones too.	2.11	0.92	0.32	0.15	0.32
21	When I learned mathematics, I copied correct answers in my notebook or on exam papers.	2.87	1.06	0.30	0.08	0.30
16	When I learned mathematics, I considered how answers were gotten by reading correct ones.	2.76	0.91	0.30	−0.07	0.30
27	When trying to remember Chinese characters, I memorized radicals or origins of the words.	2.01	0.94	0.29	0.06	0.30
38	When I learned social science, I read books or articles on topics I found interesting.	2.77	1.05	0.29	0.11	0.30
10	When I learned English, I read English text in books, magazines, or newspapers.	1.66	0.89	0.29	−0.15	0.29
62	I tried to participate in scientific learning programmes held in the area or in the school.	1.50	0.89	0.27	0.06	0.27
53	When I learned science, I memorized the content without considering the reason for it.	2.12	1.01	−0.17	0.64	
14	When trying to remember mathematics formulas, I did not consider why they were like that.	2.11	1.03	−0.19	0.61	
52	When I found unclear points when learning science, I just memorized those.	2.51	1.03	0.13	0.58	
39	When I learned history, I remembered points that seemed important using rote memorization.	2.91	0.98	0.20	0.54	
40	When I learned history, I rehearsed years of historical events or the events themselves.	2.44	1.06	0.12	0.54	
12	When I learned mathematics, I remembered the procedure as much as possible.	2.48	0.92	0.05	0.53	
26	When trying to remember Chinese characters, I memorized them by writing repeatedly.	3.03	1.02	0.12	0.47	
1	When trying to remember English words, I memorized them by writing them repeatedly.	2.53	0.96	0.10	0.46	
3	When trying to remember English words, I used Japanese translations and no example sentences.	1.91	0.92	−0.04	0.45	
41	When trying to remember the years of historical events, I memorized them using rhyming words.	2.56	1.04	0.22	0.42	
25	When trying to remember Chinese characters, I memorized them using rehearsing.	1.90	1.04	0.09	0.42	
2	When trying to remember English words, I memorized only Japanese translations.	1.76	0.87	−0.02	0.36	
33	When learning Japanese, I used rote memorization without understanding difficult content.	1.63	0.85	−0.11	0.36	
42	When I learned history, I used comic books, TV programmes, or games concerning history.	2.10	1.08	0.17	0.33	

interpretation is difficult because an ICC is given for each response value. Therefore, the present study gives priority to ease of interpretation and reports the results of the binary IRT.

Parameters were estimated using a total of 45 items and IRT. The results are presented in **Table 4**, grouped by subject area for readability; note that IRT is not carried out separately for each subject area. Examples of the ICCs obtained are shown in **Figure 4**. The acquisition scores were also estimated. The value of the discrimination parameter in all items was greater than 0.55, suggesting that there were no items with a low discrimination parameter value.

Examination of learning strategies used by each university revealed that students at the higher-performing universities actively used deep cognitive and metacognitive strategies; however, they used strategies that utilized external resources less frequently. For example, learning strategies such as "I tried to participate in scientific learning programmes held in the area or school" or "When I learned English, I read English text in books,

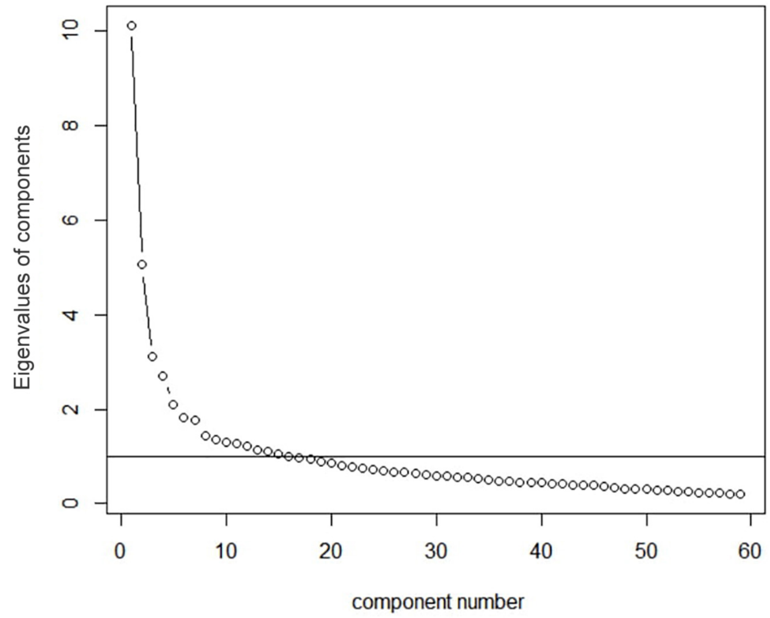


FIGURE 2
Scree plot of factor analysis.

magazines, or newspapers" had item difficulty values above 2.0. These strategies were also used less frequently by students at universities with higher academic achievement. However, this should be examined in future studies.

Relationship between learning strategy acquisition and academic achievement

If learning strategy acquisition level highly correlates with academic achievement, it might provide evidence for its validity. The average acquisition scores for each university were calculated to examine the relationship between acquisition level and academic achievement. Note that the study included students from several universities and grades. Therefore, it is not possible to analyze academic achievements using a unified index. However, the university itself is considered to broadly reflect the learner's level of study. In particular, it has been suggested that in Japan, higher education plans are determined mainly by academic ability and less by other influences (Kariya and Rosenbaum, 1987). Kariya and Rosenbaum (1987) collected data for junior high schools, but the idea is that Japanese students do not change significantly, even when they are aiming to go on to university. Therefore, individual differences in the level of acquisition of learning strategies are related to the level of university, which is considered a manifestation of academic ability in a broad sense. The purpose of this analysis is to show that the theta resulting from this study may provide a background for a certain type of academic ability.

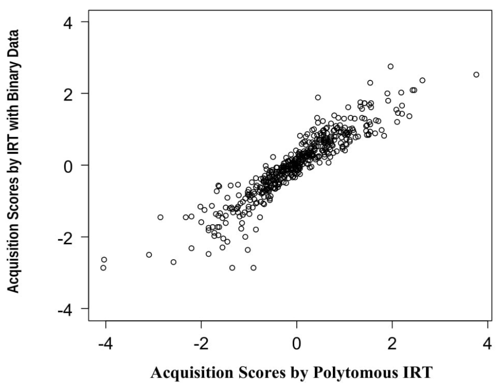


FIGURE 3
The correlation between students' acquisition scores obtained from item response theory (IRT) with binary data and polytomous IRT.

As shown in **Figure 5**, average acquisition scores, arranged in order, correspond to academic achievement as estimated by the Japanese university rankings (i.e., Universities A and B = comparatively higher-ranked, Universities C and D = middle-ranked, University E = comparatively lower-ranked). The result suggests that the participants' level of mastery of the use of deep-processing learning strategies may be part of the background to their high academic performance.

TABLE 4 Results of item response theory (IRT) analysis.

Item		Item difficulty		Item discrimination	
		Estimates	*SE*	Estimates	*SE*
Learning Strategies in Mathematics					
20	When I learned mathematics, I checked whether answers were correct after solving.	−1.42	0.21	0.96	0.15
22	When I learned mathematics, I tried to consider why I failed to solve problems.	−0.67	0.10	1.66	0.21
18	When I learned mathematics, I tried again to solve the problems that I had failed to solve.	−0.51	0.10	1.51	0.19
24	When I found unclear points while learning mathematics, I asked my friends and teachers.	−0.82	0.17	0.84	0.13
21	When I learned mathematics, I copied correct answers in my notebook or exam papers.	−1.09	0.27	0.57	0.12
17	When I found math problems I could not solve, I tried to do so by drawing diagrams.	−0.39	0.10	1.36	0.17
16	When I learned mathematics, I considered how answers were gotten by reading correct ones.	−0.83	0.24	0.55	0.11
15	When trying to remember mathematics formulas, I memorized situations where they were used.	−0.36	0.12	1.03	0.14
13	When I learned mathematics, I used study aids.	0.09	0.14	0.77	0.13
19	When trying to remember mathematics formulas, I considered how to get the formulas.	0.07	0.10	1.07	0.15
23	When learning mathematics, to prevent failure, I wrote points for attention in my notebook.	1.15	0.18	1.05	0.16
Learning Strategies in Science					
64	When I found unclear points when learning science, I asked my friends and teachers.	−0.41	0.11	1.18	0.16
61	When I learned science, I tried to focus on learning parts I did not understand.	−0.19	0.08	1.68	0.21
63	When learning science, I used study aids.	−0.16	0.11	0.99	0.14
56	When I learned science, I tried to make clear the points that I did not understand.	−0.10	0.07	1.94	0.23
59	When I learned science, I tried to consider why I failed to solve problems.	−0.07	0.07	2.11	0.25
58	When I learned science, I summarized what I learned with tables and figures.	−0.02	0.09	1.42	0.18
57	When I found some new concepts in learning science, I considered the definition and example.	0.04	0.08	1.77	0.22
60	When I learned science, I watched TV programs or read journals concerning science.	1.14	0.21	0.81	0.14
62	I tried to participate in scientific learning programmes held in the area or in the school.	2.41	0.46	0.81	0.17
Learning Strategies in Language Learning					
37	When I read (Japanese) text, I read slowly if I found unclear points.	−1.26	0.21	0.84	0.14
34	When I read (Japanese) text and found a new word, I tried to consider its meaning.	−0.39	0.11	1.09	0.15
29	When I found difficult words while reading (Japanese), I looked them up in the dictionary.	−0.53	0.16	0.75	0.13
35	When finding difficult words in (Japanese) text, I paraphrased using easy expressions.	0.12	0.13	0.81	0.13
30	When reading (Japanese) text, I tried to summarize the meaning of each sentence.	0.35	0.13	0.82	0.13
32	When I read (Japanese) text, I considered how content related to knowledge I already had.	0.43	0.12	1.00	0.15
31	When I found unclear meanings in conversation with others, I asked questions about them.	0.88	0.21	0.66	0.13
28	When trying to remember Chinese characters, I memorized similar or opposite ones too.	1.26	0.26	0.67	0.13
36	When I read (Japanese), I set myself some questions to check my understanding.	1.67	0.25	1.05	0.18
27	When trying to remember Chinese characters, I memorized radicals or origins of the words.	1.71	0.36	0.63	0.14
Learning Strategies in Social Science					
47	When I learned history, I tried to understand the big sweep of time rather than the details.	−0.78	0.17	0.80	0.13
38	When I learned social science, I read books or articles on topics I found interesting.	−0.75	0.23	0.55	0.11
45	When I learned history, I summarized how events happened in my notes or in my mind.	−0.43	0.12	1.00	0.14
49	When learning social science, I applied mental resources to points I did not remember.	−0.40	0.11	1.10	0.15
43	When I learned social science, I tried to consider how different events connect to each other.	−0.14	0.09	1.36	0.18
48	When I learned social science, I considered why such a historical event happened.	−0.12	0.09	1.35	0.17
44	I talked with my friends or family about history.	0.27	0.14	0.76	0.13
46	When I read textbooks concerning social science, I tried to find unclear points.	0.33	0.10	1.26	0.17
Learning Strategies in English Learning					
9	When I read English text, I looked words up in the dictionary if I found unknown words.	−1.34	0.35	0.49	0.11
6	When I found unknown words, I took a guess based on the context where they were used.	−0.70	0.15	0.91	0.14
8	When trying to remember English words, I concentrated on words I did not remember.	−0.60	0.14	0.91	0.14
7	When I read English text, I paid attention to points where I got confused.	0.13	0.10	1.17	0.16
4	When trying to remember English words, I paid attention to situations where the word was used.	0.21	0.12	0.96	0.14
5	When learning English, I remembered words by identifying prefixes and suffixes.	0.27	0.13	0.81	0.13
10	When I learned English, I read English text in books, magazines, or newspapers.	2.73	0.61	0.66	0.16

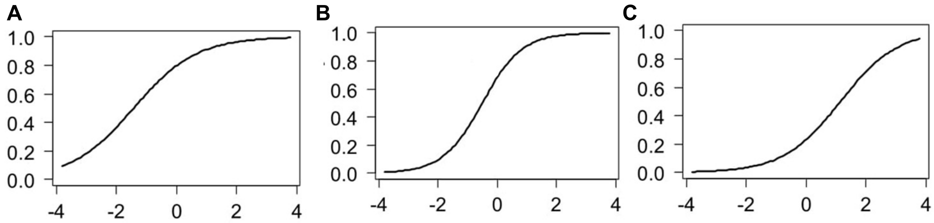


FIGURE 4
Examples of item characteristic curves (ICCs) in items concerning mathematics. **(A)** ICC of Item 20 "When I learned mathematics, I checked whether answers were correct after solving." **(B)** ICC of Item 18 "When I learned mathematics, I tried again to solve the problems that I had failed to solve." **(C)** ICC of Item 23 "When learning mathematics, to prevent failure, I wrote points for attention in my notebook."

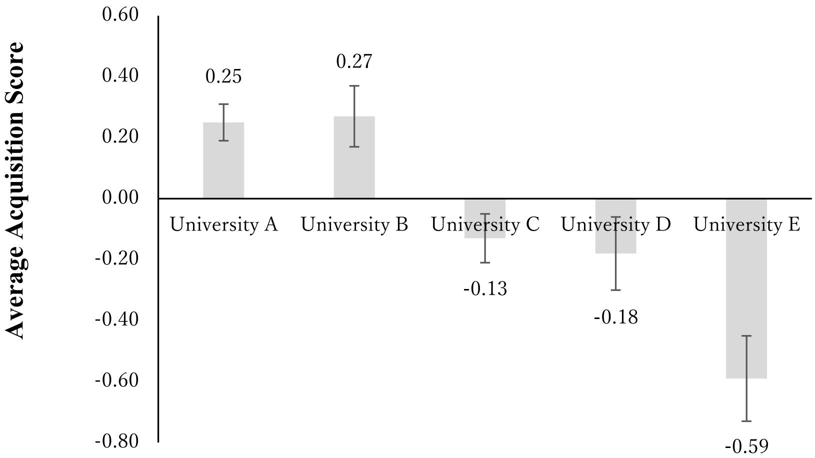


FIGURE 5
Average acquisition scores for students in each university. The vertical bars show the standard deviations.

Discussion

Summary of the findings in the present research

This study analyzed students' levels of learning strategies using IRT and specified the ease or difficulty level of each strategy. Traditionally, learning strategy research did not consider the level of learning strategies among students. Moreover, most research, which tried to promote students' academic performance by encouraging the use of learning strategies, usually provided instructions that were not adjusted to individual levels. This study attempted to shed light on this aspect.

After confirming the one-dimensionality, IRT analysis was applied to all deep-level processing types of items. Moreover, the positive relationship between the acquisition score retrieved from this analysis and academic performance was confirmed. We also specified the appropriate learning strategy for a specific learner level. Although academic achievement was evaluated by the university's rank, a clear relationship was demonstrated. This means that the θ assessed by the questionnaires might be the basis of academic achievement.

How to utilize the findings of the current study

This manuscript proposes a framework as a basis for ensuring that students can be advised according to their level of acquisition of learning strategies. Although it is limited in that it only proposes a basic framework that has not yet led to actual support, it proposes a direction for the use of this framework in the future.

First, applying this idea in educational practice should involve utilizing it in assessment as well as in the results of personalized interventions. Although current personal tutoring research mainly focuses on instruction in specific subject-related knowledge, such as biology (see the review of Chi et al., 2001; Roscoe and Chi, 2008), some studies and practices focus on improving learning strategy use through

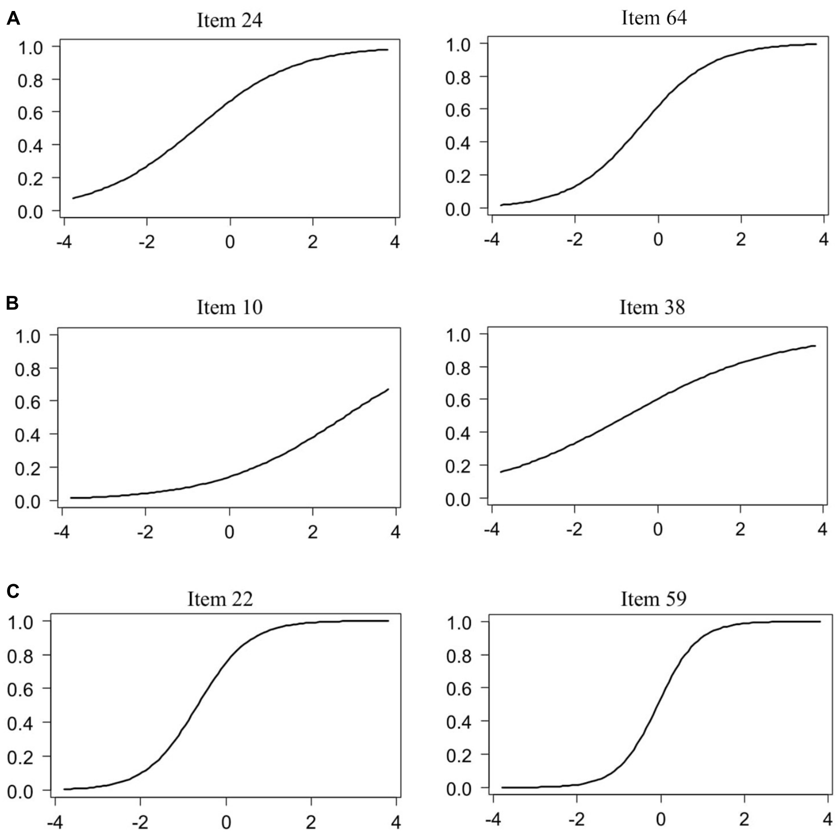


FIGURE 6
Comparison of item characteristic curves (ICCs) in different subjects. **(A)** ICCs of Item 24 ("When I found unclear points while learning mathematics, I asked my friends and teachers") and Item 64 ("When I found unclear points when learning science, I asked my friends and teacher"). **(B)** ICCs of Item 10 ("When I learned English, I read English text in books, magazines, or newspapers") and Item 38 ("When I learned social science, I read books or articles on topics I found interesting"). **(C)** ICCs of Item 22 ("When I learned mathematics, I tried to consider why I failed to solve problems") and Item 59 ("When I learned science, I tried to consider why I failed to solve problems").

personal tutoring; for instance, Craig et al. (2000) demonstrated the enhancement of questioning skills by tutoring. Another personal tutoring study conducted in Japan, referred to as "cognitive counseling" (Ichikawa, 2005; Ichikawa et al., 2017), aimed to develop autonomous learners and utilized cognitive psychology. Currently, counselors with a strong background in cognitive psychology manually assess clients' learning strategy acquisition levels based on their materials. Based on the findings of this study, tutors or counselors can assess learners' strategy acquisition levels and teach appropriate learning strategies.

Second, this framework can also be used in classroom instructions to support the complex task of adjusting them according to each student's needs. This can be done in several ways. For instance, the framework can provide teachers with information on what kinds of learning strategies are used. They can also assess average strategy acquisition levels among students and help teachers adjust their instructions accordingly. Moreover, effective classroom interventions can be implemented, for example, to construct subgroups that share similar levels of learning strategies, and provide tailored instruction to the levels in those respective groups.

Based on the assumption that learning strategies with item difficulties at the same level as acquisition scores are recommended, more concrete suggestions to apply these findings in classroom instruction are reflected below. When students' average acquisition score is approximately $\theta = -1.0$ (not so high), the following learning strategies might be effective (many of these provide ways to deal with failure in problem-solving or perceived difficulty in learning).

- Item 20: When I learned mathematics, I checked whether answers were correct after solving.
- Item 37: When I read (Japanese) text, I read slowly if I found unclear points.
- Item 18: When I learned mathematics, I tried again to solve the problems that I had failed to solve.
- Item 24: When I found unclear points while learning mathematics, I asked my friends and teachers.

- Item 9: When I read English text, I looked words up in the dictionary if I found unknown words.

If the students' average acquisition score is approximately $\theta = 0$ (moderate), the following learning strategies might be effective.

- Item 58: When I learned science, I summarized what I learned with tables and figures.
- Item 35: When finding difficult words in (Japanese) text, I paraphrased using easy expressions.
- Item 5: When learning English, I remembered words by identifying prefixes and suffixes.
- Item 4: When trying to remember English words, I paid attention to situations where the words were used.

If students' average acquisition score is approximately $\theta = 1.0$ (comparatively high), more advanced learning strategies might be appropriate. At this level, advanced strategies that utilize engagement with authentic resources, such as journal articles or community-provided learning programs outside of school, would be useful. The following items provide such examples.

- Item 60: When I learned science, I watched TV programs or read journals concerning science.
- Item 10: When I learned English, I read English text in books, magazines, or newspapers.
- Item 62: I tried to participate in scientific learning programmes held in the area or in the school.

Future directions and questions arising from this study

It is worth discussing the future research directions that emerge from this study. One such direction is to integrate more students as participants. The θ calculated in this study was based on university students' reflections on their practices and learning during high school. However, high school is not the only stage relevant to this model. For example, Uesaka (2010) instructed junior high school students (eighth grade) in metacognitive learning strategies, reviewing how to analyze failures regarding cognitive counseling, which succeeded in inducing the transfer of learning strategies. Studies in the personal tutoring context and meta-analysis by Dignath et al. (2008) demonstrated the effectiveness of learning strategy instruction at younger school levels, specifically, elementary school. University students should also be instructed in effective learning strategies (Manalo and Henning, 2017). Integrating data right from elementary school to the university level would reveal the learning strategy acquisition order, contributing to high-quality education over the entire learning period.

The second direction is to collect a wider range of items for the analysis. This study applied IRT analysis to items that could be retrieved from previous work and classified them as cognitive, metacognitive, or external resource utilizing strategies. The study could not integrate all types of learning strategies due to time limitation in implementing the questionnaire. However, if the range of participants were expanded, it would be necessary to integrate an increasing number of learning strategies; for example, if we want to cover elementary school students, more items with easier acquisition orders should be integrated. Although this study did not do so, the IRT equating method might be worth using, wherein common items are included for every participant and item parameters are calculated consistently for every age group.

The third direction is to examine in more detail the relationship between data on the level of learning strategies and academic achievement. In this study, we divided the university levels into three categories and examined their relationship to the learning strategy levels obtained. In the future, it will be important to obtain data on individual academic achievement and verify its relationship with learning strategy levels.

Finally, it is worth examining how the use of similar strategies varies across subjects. This study used various types of learning strategies so that the findings could help teachers gain concrete suggestions regarding which strategies should be employed in different subject areas. The results raised the question of whether learning strategy use differed depending on the subject area. For example, as shown in **Figure 6**, the pair of ICCs comprising Item 24 ("When I found unclear points while learning mathematics, I asked my friends and teachers") and Item 64 ("When I found unclear points when learning science, I asked my friends and teachers"), and the pair of Item 10 ("When I learned English, I read English text in books, magazines, or newspapers") and Item 38 ("When I learned social science, I read books or articles on topics I found interesting") were different, even though the types of learning strategies were similar. However, a pair of ICCs for Item 22 ("When I learned mathematics, I tried to consider why I failed to solve the problems") and Item 59 ("When I learned science, I tried to consider why I failed to solve the problems") were similar. A more detailed examination of the relationship between learning strategy use and subject areas is one of the most important topics for future investigation.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.org/10.6084/m9.figshare.12502028.v1.

Ethics statement

The studies involving human participants were reviewed and approved by the Life Sciences and Research Ethics and Safety Committee of the University of Tokyo. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YU prepared the questionnaire, collected the data, and wrote the first draft. MS analyzed the data. SI initiated the original idea for this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by grants-in-aid (15H01976, 19H00616, and 20K20516) received from the Japan Society for the Promotion of Science.

Acknowledgments

We would like to thank Mikiko Seo of Japan Women's University and Yoshinori Oyama of Chiba University for supporting data collection, and Emmanuel Manalo of Kyoto University for giving helpful advice on this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ames, C., and Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *J. Educ. Psychol.* 80, 260–267. doi: 10.1037/0022-0663.80.3.260

Balsis, S., Ruchensky, J. R., and Busch, A. J. (2017). Item response theory applications in personality disorder research. *Personal. Disor.* 8, 298–308. doi: 10.1037/per0000209

Benesse Corporation (2017). *Check from the list of deviation values of the university*. Available online at: https://manabi.benesse.ne.jp/ap/daigaku/search/nanido/ (accessed April 1, 2022).

Brackenbury, T., Zickar, M. J., Munson, B., and Storkel, H. L. (2017). Applying item response theory to the development of a screening adaptation of the Goldman-Fristoe Test of Articulation-Second Edition. *J. Speech Lang. Hear. Res.* 60, 2672–2679. doi: 10.1044/2017_JSLHR-L-16-0392

Camahalan, F. M. G. (2006). Effects of self-regulated learning on mathematics achievement of selected Southeast Asian children. *.Instruc. Psychol.* 33, 194–205.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., and Hausmann, R. G. (2001). Learning from human tutoring. *Cogn. Sci.* 25, 471–534. doi: 10.1016/S0364-0213(01)00044-1

Cleary, T. J., and Zimmerman, B. J. (2004). Self-regulation empowerment program: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychol. Schools* 41, 537–550. doi: 10.1002/pits.10177

Craig, S. D., Gholson, B., Ventura, M., Graesser, A. C., and The, T. R. G. (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *Int. J. Artif. Intellig. Educ.* 11, 242–253. doi: 10.2190/Q8CM-FH7L-6HJU-DT9W

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford.

Dignath, C., Buettner, G., and Langfeldt, H. P. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educ. Res. Rev.* 3, 101–129. doi: 10.1016/j.edurev.2008.02.003

Griffin, P., McGaw, B., and Care, E. (eds) (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer.doi: 10.1007/978-94-007-2324-5

Guthrie, J. T., Van Meter, P., Hancock, G. R., Alao, S., Anderson, E., and McCann, A. (1998). Does concept-oriented reading instruction increase strategy use and conceptual learning from text? *J. Educ. Psychol.* 90, 261–278. doi: 10.1037/0022-0663.90.2.261

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., et al. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *J. Educ. Psychol.* 96, 403–423. doi: 10.1037/0022-0663.96.3.403

Harris, C. W., and Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika* 29, 347–362. doi: 10.1007/BF02289601

Harris, K. R., and Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learn. Disabil. Q.* 22, 251–262. doi: 10.2307/1511259

Ichihara, M., and Arai, K. (2006). Moderator effects of meta-cognition: A test in math of a motivational model Japanese. *J. Educ. Psychol.* 54, 199–210. doi: 10.5926/jjep1953.54.2_199

Ichikawa, S. (2005). "Cognitive counseling to improve students' metacognition and cognitive skills," in *Applied developmental psychology: Theory, practice, and research from Japan*, eds D. W. Shwalb, J. Nakazawa, and B. J. Shwalb (Charlotte: Information Age Publishing), 67–87. doi: 10.5014/ajot.2019.031450

Ichikawa, S. (ed.) (1998). *Consultation and guidance on learning methods from the perspective of cognitive counseling*. New York: Brain-shuppan Inc.

Ichikawa, S., Uesaka, Y., and Manalo, E. (2017). "Three approaches to promoting spontaneous use of learning strategies: Bridging the gap between research and school practice," in *Promoting spontaneous use of learning and reasoning strategies: Theory, research, and practice for effective transfer*, eds E. Manalo, Y. Uesaka, and C. A. Chinn (Park Drive: Routledge), 195–210. doi: 10.4324/9781315564029-13

Inuzuka, M. (2002). The structure of reading strategies for understanding expository text. *Jap. J. Educ. Psychol.* 50, 152–162. doi: 10.5926/jjep1953.50.2_152

Kariya, T., and Rosenbaum, J. E. (1987). Self-selection in Japanese junior high schools: A longitudinal study of students' educational plans. *Sociol. Educ.* 60, 168–180. doi: 10.2307/2112274

Kesici, Ş, Erdoğan, A., and Ahmet, K. (2009). Predicting college students' mathematics anxiety by motivational beliefs and self-regulated learning strategies. *College Stud. J.* 43, 631–642.

Lynch, D. J. (2006). Motivational factors, learning strategies and resource management as predictors of course grades. *College Stud. J.* 40, 423–428. doi: 10.1186/s12913-016-1423-5

Manalo, E., and Henning, M. (2017). "Second language vocabulary learning: Are students cognitive misers and if so, why?," in *Promoting spontaneous use of learning and reasoning strategies: Theory, research and practice for effective transfer*, eds E. Manalo, Y. Uesaka, and C. A. Chinn (Park Drive: Routledge), 91–102. doi: 10.4324/9781315564029-7

Manalo, E., Uesaka, Y., and Chinn, C. A. (eds) (2017). *Promoting spontaneous use of learning and reasoning strategies: Theory, research, and practice for effective transfer*. Park Drive: Routledge.doi: 10.4324/978131556 4029

Marton, F., and Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *Br. J. Educ. Psychol.* 46, 4–11. doi: 10.1111/j.2044-8279.1976.tb02980.x

Min, S., and Aryadoust, V. (2021). A systematic review of item response theory in language assessment: Implications for the dimensionality of language ability. *Stud. Educat. Evaluat.* 68:100963. doi: 10.1016/j.stueduc.2020.100963

Murayama, K. (2003). Test format and learning strategy use. *Jap. J. Educ. Psychol.* 51, 1–12. doi: 10.5926/jjep1953.51.1_1

Murayama, K. (2007). "Gakushu horyaku: kodomo jishin no jiritsuteki na gakushu wo mezashi te [Learning Strategies: for children's self-regulated learning]," in *Zettai yakudatsu kyoiku shinri gaku*, ed. T. Fujita (Kyoto: Mineruvua Shoboo), 85–100

Nolen, S. B. (1988). Reasons for studying: Motivational orientations and study strategies. *Cogn. Instruct.* 5, 269–287. doi: 10.1207/s1532690xci0504_2

Oyama, S. (2009). Strategies used for learning English vocabulary and occurrence of errors in English. *Jap. J. Educ. Psychol.* 57, 73–85. doi: 10.5926/jjep.57.73

Partchev, I., Maris, G., and Hattori, T. (2017). *irtoys: A collection of functions related to item response theory (IRT)*. R package version 0. 2. 1. Available online at: https://CRAN.R-project.org/package=irtoys (accessed April 1, 2022).

Pintrich, P. R., and de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *J. Educ. Psychol.* 82, 33–40. doi: 10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D. A. F., Garcia, T., and McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educ. Psychol. Measur.* 53, 801–813. doi: 10.1177/0013164493053003024

Ramdass, D., and Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *J. Adv. Acad.* 20, 18–41. doi: 10.4219/jaa-2008-869

Ramsden, P. (1988). "Context and strategy," in *Learning strategies and learning styles*, ed. R. R. Schmeck (New York, NY: Plenum), 159–184. doi: 10.1007/978-1-4899-2118-5_7

Revelle, W. (2018). *psych: Procedures for personality and psychological research. R package version 1.8.4*. Available online at: https://CRAN.R-project.org/package=psych (accessed April 1, 2022).

Roscoe, R. D., and Chi, M. T. H. (2008). Tutor learning: The role of explaining and responding to questions. *Instruct. Sci.* 36, 321–350. doi: 10.1007/s11251-007-9034-5

Rychen, D. S., and Salganik, L. H. (eds) (2003). *Key competencies for a successful life and a well-functioning society*. Göttingen: Hogrefe.

Salden, R. J. C. M., Aleven, V., Schwonke, R., and Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instruct. Sci.* 38, 289–307. doi: 10.1007/s11251-009-9107-8

Sawyer, R. J., Graham, S., and Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *J. Educ. Psychol.* 84, 340–352. doi: 10.1037/0022-0663.84.3.340

Schunk, D. H., and Zimmerman, B. J. (eds) (1994). *Self-regulation of learning and performance: Issues and educational applications*. New York: Lawrence Erlbaum Associates, Inc.

Schunk, D. H., and Zimmerman, B. J. (eds) (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York, NY: Guilford Publications.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R., Krieg, C., and Renkl, A. (2007). "Can tutored problem solving benefit from faded worked-out examples?," in *Proceedings of EuroCogSci 07. The European Cognitive Science Conference 2007*, eds S. Vosniadou, D. Kayser, and A. Protopapas (Erlbaum), 59–64.

Shell, D. F., and Husman, J. (2008). Control, motivation, affect, and strategic self-regulation in the college classroom: A multidimensional phenomenon. *J. Educat. Psychol.* 100, 443–459. doi: 10.1037/0022-0663.100.2.443

Shimoyama, H. (1984). A longitudinal study on the career decision-making process of a high school. *Jap. J. Educ. Psychol.* 32, 206–211. doi: 10.5926/jjep1953.32.3_206

Tabachnick, S. E., Miller, R. B., and Relyea, G. E. (2008). The relationships among students' future-oriented goals and subgoals, perceived task instrumentality, and task-oriented self-regulation strategies in an academic environment. *J. Educ. Psychol.* 100, 629–642. doi: 10.1037/0022-0663.100.3.629

Tang, K. L. (1996). *Polytomous item response theory models and their applications in large-scale testing programs: Review of literature*. Princeton: Educational Testing Service.

Uchida, N. (2021). Developmental differences in English vocabulary learning strategies: A cross-sectional study of secondary-level students. *Jap. J. Educ. Psychol.* 69, 366–381. doi: 10.5926/jjep.69.366

Uesaka, Y. (2010). How learning strategy use transfers across different school subjects: A case study on promotion of spontaneous use of "lesson induction.". *Jap. J. Educ. Psychol.* 58, 80–94. doi: 10.5926/jjep.58.80

Uesaka, Y. (2012). Reviews of psychological studies and educational practices focusing on improving student learning skills: Suggestions for addressing the gap between academic studies and classroom practices. *Annu. Rep. Educ. Psychol. Japan* 51, 105–117. doi: 10.5926/arepj.51.105

Uesaka, Y., Shiba, S., and Ichikawa, S. (eds) (2017). *Case studies of learning support at 2017 academic year: Cognitive counseling practices for fostering independent learners*. Tokyo: The University of Tokyo.

Weinstein, C. E., and Mayer, R. E. (1986). "The teaching of learning strategies," in *Handbook of research on teaching*, ed. M. C. Wittrock (London: Macmillan), 315–327.

Zimmerman, B. J. (2015). "Self-regulated learning: Theories, measures, and outcome," in *International Encyclopedia of the Social* & cpsjrn*Behavioral Sciences* (Second Edition), ed. J. D. Wright (Amsterdam: Elsevier), 541–546. doi: 10.1016/B978-0-08-097086-8.26060-1

Zimmerman, B. J., and Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *J. Educ. Psychol.* 82, 51–59. doi: 10.1037/0022-0663.82.1.51

Zimmerman, B. J., and Schunk, D. H. (eds) (1989). *Self-regulated learning and academic achievement: Theory, research, and practice*. Berlin: Springer-Verlag Publishing, doi: 10.1007/978-1-4612-3618-4

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fpsyg.2022.945909

Check for updates

OPEN ACCESS

EDITED BY
Shuhua An,
California State University, Long Beach, United States

REVIEWED BY
Lianchun Dong,
Minzu University of China, China
Wenjun Zhao,
Sichuan Normal University, China

*CORRESPONDENCE
Jingbo Zhao
359129668@qq.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 17 May 2022
ACCEPTED 18 August 2022
PUBLISHED 20 September 2022

CITATION
Zhao J, Song T, Song X and Bai Y (2022) Analysis on the linguistic features of conflict discourse in mathematical cooperation problem solving in China. *Front. Psychol.* 13:945909. doi: 10.3389/fpsyg.2022.945909

Analysis on the linguistic features of conflict discourse in mathematical cooperation problem solving in China

Jingbo Zhao*, Tingting Song, Xiaoying Song and Yuanmeng Bai

Key Laboratory of Data Science and Smart Education, Ministry of Education, Hainan Normal University, Haikou, China

Classroom teaching is a kind of social activity system. Thus, as a form of classroom learning, collaborative problem solving has a strong social attribute. It is extremely common to choose the conflict discourse in the context of cooperation. The verbal characteristics of the conflicting discourse level in cooperative mathematics problem solving directly affects the cooperative learning between students and the classroom teaching of teachers. This article focuses on the overall linguistic characteristics of conflict discourse in solving cooperative problems and the discourse style and language characteristics of the three stages of conflict discourse. The main research conclusions are as follows: (1) The classification of language features of conflict discourse includes extreme summaries, negation, discourse markers, and so on. Among them, the frequency of Indexical 2nd-person pronouns is the highest. (2) The language expressions at the "initial stage of conflict" include Explanatory statement Negative response, instruct refuse and Seditious inquiry Confrontational answer. The language shows the characteristics of using emphatic words or phrases, negative words, imperative sentences and so on. Meanwhile, rebuttal questions, direct responses, explanations, and negative avoidance are the main forms language expressions at the "conflict stage." It also exhibits the verbal characteristics of rhetorical questions, negative comments, and direct negation. Lastly, topic-shifting, compromise, third-party intervention, and one-sided wins are the linguistic expressions at the "end of conflict." The language features are the appearance of tone relaxation and language easing, and the conflict ending utterances reflect cooperation.

KEYWORDS

mathematical problem solving, collaborative learning, conflict talk, discourse analysis, linguistic features

Introduction

Cooperation is an important driving force behind the development of human society, and cooperation skills have become one of the most essential qualities for citizens the 21st century. As an important teaching method, various countries have paid increasing attention to cooperative problem-solving in recent years. Students start by realizing social interaction in the process of problem solving, which subsequently improves cooperative cognition. As early as the beginning of the 21st century, the United Nations Educational, Scientific, and Cultural Organization (UNESO) regarded teamwork and problem-solving skills as essential qualities for citizens of the 21st century (OECD, 2020). In 2013, the PISA 2015 Cooperative Problem Solving Framework Draft published by the International Student Assessment Project introduced the Collaborative Problem Solving (CPS) test for the first time (Lin, 2016). "China's Compulsory Education Mathematics Curriculum Standard (2011 Edition page11)" pointed out: "Experience the process of cooperating and communicating with others to solve problems." The latest 2022 edition (page 1) also pointed out: "Help students develop the habit of independent thinking and the willingness to cooperate and communicate."

This article mainly studies the verbal characteristics of the conflicting discourse in cooperative mathematics problem solving. It refers to Scott's (2002) language classification of conflicting discourse in combination with the language characteristics of middle school students in the process of group cooperation in a Chinese classroom environment to identify the characteristics of conflict words and their distribution trends in the process of cooperative mathematics problem solving to lay a foundation for the further analysis of the impact of speech and distribution trends with different characteristics on cooperative problem-solving. Conflict discourse is a catalyst to help solve cooperative problems (Cazden, 2004). Effective conflict in the classroom can prompt group members to re-examine their initial opinions, promote the verification of a student's point of view, effectively strengthen students' motivation to obtain information, and enhance students' motivation to learn. At the same time, it can encourage learning to actively communicate with others and improve interpersonal communication skills. Conflicts caused by different cognitions among students are very beneficial for them to acquire real knowledge. Solving the conflict in the process of discussion drives the further development of students.

At the same time, Conflict discourse is the normal state of group members' discourse choice and negotiation in the context of cooperation. Studying Conflict Discourse in students' cooperative problem solving will help group members further grasp the rules of verbal communication and better understand the power of the team in cooperative learning; Teachers can understand the dynamic process of students' discourse conflict according to the occurrence, development and ending process of conflict discourse, grasp the interaction law, and provide targeted intervention guidance for group cooperative learning, which has important theoretical and practical significance.

Literature review

Status quo of research on cooperative learning based on problem solving

In 1926, some scholars discussed the cooperative problem solving (Alschuler et al., 1977). Cooperative problem-solving integrates the two capabilities of problem solving and cooperation. Since the 1980s, there has been more and more research on problem-solving, especially in the field of education in the United States and Australia. There is a deep interest in this field (Goos and Galbraith, 1996). Since the 21st century, team-based problem solving has been increasingly recognized as having a vital impact on society. Enterprises and society have paid more and more attention to the power of teamwork as well as how interpersonal relationships and the problem-solving abilities of staff are related. In 2013, the PISA 2015 Collaborative Problem Solving Framework Draft included the cooperative problem solving test as part of their test content, with problem solving ability being an important marker of mathematical ability. The problem solving ability has been valued by the United States (Li and Cao, 2019), Finland (Yu and Cao, 2017), and other countries that have achieved good results in the PISA test. In the newly announced PISA 2021 math test framework, the definition of mathematics literacy is clearly defined, with mathematics communication and problem solving an important part of it (Cao and Zhu, 2019).

Yang and Wang (2019) racked and studied the dynamics of cooperative problem-solving research on the international front and came to a series of important conclusions. At present, cooperative problem-solving mainly focuses on the following six hot topics: distributed cognition, cooperative learning, co-building scaffolding, experimental teaching, peer guidance and chemistry. The relatively new research content in the field of mathematics education currently includes multidisciplinary research, early childhood education, and green chemistry research. It is more common about holistic cognition and sequential cognition in the research of cognitive style, and there are some new development trends in this research field. In terms of the cognitive approach, the theory of situational cognition has gradually shifted to the theory of distributed cognition; in terms of education, it has shifted from a single subject to multiple cooperation; in terms of content, it has shifted from traditional subjects to emerging subjects (Yang and Wang, 2019). In China, the research on cooperative problem solving mainly focuses on the macro level, such as the evaluation of cooperative problem solving ability (Hao and He, 2019), the cultivation of cooperative ability, and the model of cooperative learning (Cao and Bai, 2018).

Research on conflict discourse in mathematical cooperation problem solving

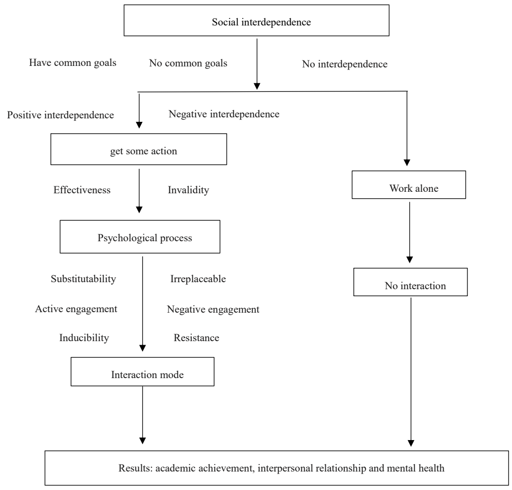
In the process of cooperative problem solving, conflict is inevitable. This conflict usually manifests in the form of "cognitive conflict," which is caused by different views on the problem rather than other forms of social conflict such as emotional conflict (Shi and Du, 2008). Johnson claims that conflict occurs when "the two parties in the communication hold different views or opinions, and they want to reach an agreement (Zhang, 2020)." The purpose of cooperative learning is to solve problems and for team members to reach a consensus. Team members use various forms of interaction to achieve a new balance among themselves in the form of conflict and negotiation.

The cognitive construction process of cooperative learning is divided into two stages: the individual knowledge construction stage and the cooperative joint structure stage (Liu and Wang, 2008). The stage of individual knowledge construction is the process by which individuals compare their own mental model with the presented material and with other mental models after being presented the learning materials. The cooperative joint structure stage can coordinate differences and is a process of refinement, discussion, and conflict (Wang, 2004). Alexopoulou and Driver (1997) proved the dual nature of conflict in cooperative learning and concluded that "arguments are composed of other effective types of expressions."

To sum up, the current international research on conflict discourse mainly includes theoretical research on the definition and function of conflict discourse. The empirical research mainly focuses on the structure and language characteristics of conflict discourse, but most of them are conducted from the perspective of daily conversation. However, there is still a certain research space to explore the types and styles of conflict in classroom teaching from the perspective of mathematics classroom based on real corpus. Research on the conflict and negotiation process of cooperative mathematics problem solving is our focus. This article mainly analyzes the conflict language characteristics in cooperative mathematical problem solving, and empirically studies the conflict and negotiation events in the context of mathematical problem solving based on the existing research.

Theoretical framework

Social interdependence theory is an important theoretical basis of cooperative learning. It mainly discusses the action efficiency, internal psychological process, interaction mode and results of individuals when they interact in cooperative and competitive social situations. In 1935, Kurt Kafka, the founder of Gestalt psychology in Germany, first put forward the view of "the integrity of group dynamics." In 1949, the disciple of Lewin, Dodge, developed his theory and proposed two types of positive and negative social interdependence, which directly affected the interaction mode and psychological process of both sides of communication. Then the Johnson Brothers, Dodge's disciples, constructed a systematic "social interdependence theory" based on their social interdependence theory, forming a series of operable procedures, As shown below:



In the process of cooperative learning, positive social interaction should have a good performance in both action and emotion. Conflict is common in cooperative situations. No matter positive or negative interdependence, different individuals will conflict about how to achieve the goal of reciprocity. Facing such a problem, we should face the conflict directly, trust and help each other, and deal with the conflict constructively through the circular system thinking mode of "difference coordination unity." In the process of positive interaction, individual self-education and self-improvement, so as to complete self transcendence and finally realize self-worth.

In short, from the content of cooperative mathematics problem-solving research, cooperative learning, cognitive interaction, social interaction, peer guidance, etc. are currently the hot issues in the international research on Cooperative problem-solving. At present, the relatively new research contents in the field of mathematics education include multidisciplinary integration research, education at an early age of children, and cognitive style research. There are more holistic cognition and sequential cognition, and there are some new development trends as a whole. In terms of cognitive approach, it gradually turns from situational cognitive theory to distributed cognitive theory; In the aspect of educational subject, from single subject to multiple cooperation; In terms of problem content, it has shifted from traditional disciplines to emerging disciplines. The research on conflict and negotiation in the process of cooperative mathematics problem solving has a certain research space. Based on the existing research, this paper empirically studies the conflict and negotiation events in the context of mathematical problem solving. Based on the analysis of the existing literature, it is found that there are few studies on the types of Conflict Discourse in classroom teaching, especially in the process of group cooperative problem-solving learning. In mathematics education research, the Conflict Discourse Research of cooperative learning is still a blank. Through this research, we can deeply analyze the social interaction in the process of students' cooperative problem-solving, and then further analyze its impact on the effect of cooperation.

Research design

Research questions

Existing research mainly analyzes students' "cooperative problem-solving ability" from the perspective of assessment ability, but there have been few instances of interactive research on cooperative problem solving in classroom learning, especially in the mathematics classroom. Therefore, this topic will focus on the following issues: (1) What are the overall linguistic characteristics and distribution trends of conflicting discourse in solving mathematical cooperation problems? (2) What are the language expression methods and language characteristics at the three stages of "conflict initiation, conflict and negotiation, and conflict end" in solving mathematical cooperation problems?

Participants

The research data of this paper comes from the Sino Australian cooperation project "The Social Essentials of Learning (SEL): An experimental investigation of collaborative problem solving and knowledge construction in mathematics classrooms in Australia and China," Specifically, it comes from the 2018 national general topic of the 13th 5 years plan of China's Educational Science (BHA180157): "empirical research on cognitive interaction and social interaction and their relationship in middle school students' cooperative problem solving," Researchers have been involved in the collection and collation of project data since 2018. All the participants are Chinese students, these students get used to group work in their everyday mathematics classroom learning. In the collected video data, the video data of four people's cooperative problem-solving is selected as the data source, because the participation of the four-person cooperation group is relatively high, and the conflict discourse phenomenon is relatively prominent.

The data of this study used a total of 67 groups in eight classes of grade one in L middle school and Y middle school in B city. The group with blurred image and sound and the group with more than 4 people were removed. 48 groups were used in this study.

A study by Bruxelles and Kerbrat-Orecchioni (2004) pointed out that the distinguishing feature of multi-person conversation is "alliances" where people with the same opinion will unite to refute another person's point of view, which is more likely to cause conflict. The themed task of the four-person group is: "Xiao Ming's Apartment" (**Figure 1**).

In this study, the 48 groups with no or little intervention by the teacher was selected as the research object. In this state, the verbal interaction between group members is free and real.

Research methods

This research focuses on text analysis, and the research steps are as follows: First, the collected video data is text-transcribed to obtain the most basic text data and build the corpus of this research. Secondly, the corpus is further analyzed after using Python software to segment the text. The specific operations are as follows:

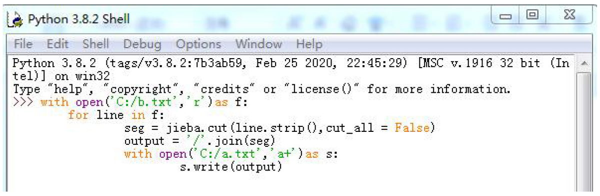
Adopting the combination of the retrieval function of the corpus retrieval software Antconc and the retrieval and screening of Excel to realize the analysis of text data, so as to realize the collection and sorting of the characteristics of the

Xiao Ming's Apartment: There are five rooms in Xiao Ming's Apartment and the total area of the apartment is 60 square meters.(each group has 20 minutes to discuss)

1. Please draw a picture showing Xiao Ming's apartment.
2. Label each room in the apartment and show the function and dimensions (length and width) of all rooms.

FIGURE 1
The math problem of "Xiao Ming's Apartment."

hedge sudden language, and then use the jieba in Python to segment the corpus, and use the antconc software to count the word frequency. The python program is as follows:



Specific analysis methods also include classroom video observation and case analysis.

Classroom video observation method takes real-time video as a technical means, and has unique advantages in analyzing the speech behaviors of teachers and students in classroom teaching. It can repeatedly observe and record the target characteristics, and analyze the speech behaviors of students scientifically and systematically through group and individual comparative comprehensive research. For different research problems, pre class interviews with teachers and students were conducted before and after the data were collected in the class, mainly to examine the understanding and participation of teachers and students in cooperative problem-solving activities. After all the preparatory work is completed, based on the recording classroom and video equipment, collect and summarize the data of students' communication audio, video, task list and so on in the process of solving cooperative problems, and carry out coding analysis.

Case analysis method, also known as case study method, is to study specific individuals, units, phenomena and themes, collect relevant data, determine research cases, sort out and analyze the generation and development process of research objects, analyze internal and external factors and mutual relations, and strengthen a more in-depth and comprehensive understanding of the problem. When analyzing the Conflict Discourse Structure in the argumentation process, this study mainly adopts case studies, selects representative focus cases for analysis, and makes a comprehensive and systematic in-depth study on the conflict in the argumentation process in cooperative learning.

Construction of coding system

When dividing conflict events, we should start from two aspects: task conflict and relationship conflict, and mainly consider three aspects. First, during the discussion, the panel members had different views; Second, the dialogue among the group members is confrontational; Third, the group members have strong emotional expressions when the conflict occurs. Conflict events mainly have three links: initial stage of conflict, Conflict and Negotiation stage and the end of conflict stage. When dividing conflict events, we should pay attention to whether these three links are complete. In particular, the end of conflict in the third link has a variety of forms, including stalemate, compromise, concession, successful negotiation, transfer conflict, etc (Table 1).

Results

The overall linguistic features of conflict discourse in mathematical cooperation problem solving

Verbal conflict is the main form of conflict in the process of cooperative problem solving. This paper uses the collected video of 48 teams (each team consists of four people) engaged in cooperation problem solving as a corpus database to analyze the verbal characteristics of conflicting language in cooperative problem solving. The classification of discourse features in this article is based on the work of Scott (2002). The language classification of Scott's conflict discourse is divided into 12 different types. On this basis, this article combines the language characteristics of middle school students in the process of group cooperation and constructs the language characteristics and distribution trends of conflict words in the process of middle school students' cooperative mathematics problem solving in the classroom environment of our country. We then further analyze the influence of different characteristics of speech and their distribution trends on the effectiveness of cooperative problem solving.

Classification of conflict language features

Based on the prominence of the conflict, Scott (2002) classified the verbal characteristics of conflicting discourse into twelve categories: extreme generalizations, negative forms, discourse markers, emphasis, turn-taking, discourse fluency markers, second-person pronouns, modality vocabulary, repetition, question sentence, turn length, and topic avoidance (As shown in Table 2).

This study is based on Scott's classification of language features of conflict words. Through the preliminary analysis of the research object, combined with the speech features of Chinese middle school students in cooperative mathematics problem solving, a speech feature classification suitable for conflict words in this study is constructed. It is mainly embodied in 10 types: extreme generalizations, negative forms, discourse markers, emphatic words, turn-taking words, second-person pronouns, modal words, repetition, rebuttal questions, and topic avoidance. We further combine the conflict fragments in the corpus to determine the analysis unit, determine the similarity pattern in the use of language features in the problem task environment, and analyze its meaning and main feature words.

TABLE 1 Conflict stage and coding.

Conflict stage	Student	Speech characteristics	Code
	S2	Multiply by what, equal to 60?	M1
Initial stage of conflict	S3	tagging.	M2
	S1	First think, first think wide. First think about the possibilities of	M3
		length and width	
	S2	2,3	M4
Conflict and Negotiation stage	S1	It's impossible to multiply by 60. It's impossible to be 2 m wide.	M5
	S3	If it's at least 5 m or 10 m wide, right? If it's 10 m wide, it's 12 m	M6
	S1	long.	M7
	S2	There are five rooms.	M8
	S1	Five rooms, yes, first of all, if he, if the length is calculated	M9
	S4	according to the integer solution, then 10, if the width is 6, the	M10
	S1	length is 10, right, if the width is 8, can 8 be divided?	M11
		It can be done.	
		8 can be divided.	
		Don't you think it's too square.	
		15.5	
	S2	In fact, the formula is better. Let's use the 10. 10 and 6 are better,	M12
The end of conflict stage	S1	the smaller the difference, the better	M13
	S2	Yes, the smaller the difference, the better. The bigger the	M14
	S1	difference, the bigger you think it is.	M15
	S2	Yes	M16
	S4	But it can't be too different. It's absolutely impossible to be thin.	M17
	S2	Then 6 and 10.	M18
	S1	6 and 10, 8 and 7.5 are the best.	M19
		Width is 6.	
		Yes, you two draw 6 and 10, and we draw 7 and 15.	

TABLE 2 Classification of language features of conflict discourse.

Classification	Expressions of conflicting discourse
Extreme summaries (Absolutes)	All, anyone, anything, anywhere, all, every, every person, everything, no matter where, never, no one, no, no place, impossible, must, absolute, sure, certain, definite, only
Negation	No, wrong, false, can not, no way, not at all
Discourse markers	But, now, ok, then
Emphatics	There are indeed, many, more, most, real, true, for example, in case, if
Floor bids	Let me come to him/her/us + verbs (for example, speak), wait a moment
Flow	
Indexical 2nd-person pronouns	You, all of you, yourself, your selves
Modals	Possibility: be able to, can, maybe, probably Necessity: must, should Predictive: shall Semi-modal words: have to do something
Repetition	Restatement of words and sentences
Questions	Interrogative sentence
Turn length	In number of words per turn
Uptake avoidance	Avoidance of previous topic

Absolutes

"Everyone, anyone, anything, anywhere, every, every, everybody, everything, no matter where, never, no one, no, no place" etc. are common extreme generalizations in students' cooperative learning. When the words "every," "all," and so on appear in the students' words, it means that the students are very confident in their words and believe that all situations are under their control, and hope that all group members are able to accept their own opinions. In the follow-up discussion, they will continue to emphasize these points. Such extreme generalizations are often refuted by others because each student's experience and knowledge level are inconsistent and they will have different views on different situations, which may lead to conflicting discourse.

Fragment 1:

S3 Brother, how did you draw it?
S1 It is drawn crooked.
S2 Actually, you can do. . .
S3 **All** your overall straight lines are drawn crookedly.
S1 Because I watched it backwards.
S2 No, they can actually have a three-dimensional one.
S1 **Does not.**
S2 You can take a look first.
S1 Paper is too small.

Extreme summaries are a form that often appears in cooperative problem solving. As in the above two segments, S3 proposed that "all" straight lines were drawn crookedly, and S1 retorted "No." Because the third party S2 intervened to let S1 draw one first, it ended the conflict between S1 and S3. As far as the conflicting parties are concerned, extreme generalizations can often stimulate one of the communicators to think about whether there are other possibilities in the plan, and to a certain extent can promote the divergence of the communicator's thinking.

Negation

Negative forms are a main way of causing conflicting discourse, which appears in the form of "no, inability, wrong, not necessary" and other expressions. The negative form is the most direct rebuttal. It shows that the two sides of the communicator are on two sides that are obviously opposite, and they hold two completely different views on the same topic. When the negative form appears, the tone is often stronger. Based on the face theory, the refuted party may feel that his face is "threatened" and will further refute and emphasize his own point of view, which escalates the conflict. If the rebutted party does not conduct another rebuttal but instead asks for the other party's opinions and negotiates with the other party, then the conflict may be eased as in the following dialogue:

Fragment 1:

S4 His house is small, you can see that it is big as a whole, so why do you think our house is so small. Do you know why the house is small? It's because of the furniture.
S1 Yes, yes, there are still messes in the house.
S4 This is because it is empty.
S1 Then this living room, the cloakroom must be in front of the living room?
S4 **It is wrong, the cloakroom shouldn't be here.**
S1 Where can the cloakroom be? Is the cloakroom in the bedroom? You wipe it, I painted it first.
S4 **It is not used.** For example, this is the living room, this opens a door, this is the door, this is the door, and this is the cloakroom.

The negative form expresses disagreement with the other party's point of view. S1 proposes that the cloakroom is in front of the living room but S4 directly denies it. S1 does not refute it further but instead asks for S4's opinion and proposes to draw first. S4 denies it again and then puts forward its own ideas. In this process, S4 is in a dominant position who hopes that the other party will accept his/her own point of view and draw according to his/her own ideas, which is why he/she used negative words at the beginning of the sentence. In addition to expressing a clear negation, the tone used for opposition is also stronger, expressing a clearer attitude of the individual.

Discourse markers

Sun and Fang (2011) conducted a comprehensive study on the types and functions of Chinese discourse markers and divided Chinese discourse markers into at least 17 types according to the word segmentation in this research, in the cooperative problem solving, the students most commonly used discourse markers such as "right, then, but" and so on. When discussing a topic, the two parties will often respond with "right," "yes," and so on, which can play a role in connecting the dialogue and making the dialogue more coherent. After the communicator is approved or affirmed, they will further discuss their views in a more harmonious atmosphere; "Then" can play the role of supplementary explanation to develop the topic further and promote a new argumentation process; "but" can play the role of "contrast" or "turning." Both parties can think and discuss from multiple levels and aspects, making the discussion process more comprehensive and efficient.

Fragment 1:

S2 In fact, the square is better, let's use the 10, 10 and 6 are better, the smaller the difference, the better.
S1 **Yes,** the smaller the difference, the greater the difference you think the bigger it is.
S2 **Yes.**
S1 However, the difference should not be particularly large.
S2 Let's choose 10 m in length, 6 m in width.
S4 6 and 10, 8 and 7.5 are the best.
S2 It is 6 m wide.
S1 **Yes,** you draw 6 and 10, then you two draw 6 and 10, and we draw 7 and 15.
S4 I feel special square.
S2 **Yes,** let's see who is better, and then design it yourself.

After S2 put forward his own ideas, he got S1's affirmative answer "Yes." The two parties in the communication reached a preliminary agreement, and then further discussed and improved from the same point of view, proposed their ideas, and negotiated with each other in a harmonious discussion atmosphere. Discourse markers can keep the dialogue connected in meaning and promote efficient discussions between the two parties, which have a certain promoting effect on the final formation of a high-quality plan.

Emphatics

It is used to emphasize one's own point of view, mainly in the form of "does, many, more, most, true, true, for example, such as, if, true" and other verbal expressions. In the process of group discussion, students often emphasize their own opinions to convince others to get on the same page. Whether other people respond to their own opinions determines whether the advocates further emphasize their opinions. If they get a positive response, then the two sides of the communicator reach a preliminary agreement. If a negative response is received, the advocate will re-examine his point of view and give examples to support his claim.

Fragment 1:

S4 10 meters in the kitchen is a bit too long, so don't talk about it first, and finish the question first.
S2 can only be 2 and 5. **Really,** my sister's house is very long and very narrow.
S4 Does 3 × 3 work?
S2 3 × 3, 9.
S4 let's still choose 3 m in length and 3 m in width.

The use of emphasized language may escalate or ease the conflicting discourse, which depends on the attitude of the two parties in the communication. S2 claimed that "length and width can only be 2 and 5" before giving an example of "my sister's house is. . ." to supplement his own point of view. S4 did not directly refute S2's claim but he tried to negotiate instead. The attitude and tone of both parties were not strong, and the discussion atmosphere was more harmonious, which promoted the initial consensus of the two parties.

Floor bids

The words used to fight for the right to speak are mainly expressions such as "let me/he/she/we come + verbs (for example: speak, say)" "wait a moment" and so on. The fight for the right to speak often occurs when the communicator does not get a response after putting forward an idea, or when the topic is changed and he wants to demonstrate his point of view again. It may also happen when you were unsatisfied with your first argument, did not express it clearly, and want to attract the attention of others again in the follow-up. It is more in line with the thinking development level of students in this age group who want to defend their sovereignty and status, are self-centered, and do not consider whether others are making speeches. The language skills of students of this age group are not perfect, and there may be unclear expression which renders them unable to persuade others. After thinking again, students will want to further add to their arguments. However, a sudden interruption of other people's speech may cause confusion in the scene and cause conflict.

Fragment 1:

S4 draw one less.
S1 **wait a moment.**
S4 Paint a little less, and paint this for the toilet.
S1 Let me take a look. What is 22 minus 8?

The topic of S4 is "Where is the toilet painted and how big is the painting," but S1 interrupted S4's topic "Wait a moment" "What is 22 minus 8?" They are not on the same topic - S1 is still immersed in the previous topic and wants to attract the attention of others. At this time, the discussion is in a chaotic situation, which will cause further conflict between the two parties. If one of the parties does not give up on their topic, then the discussion will not proceed, which will have a certain impact on the quality of the final plan.

Indexical 2nd-person pronouns

The Indexical 2nd-person pronouns mainly include "you" and "yourself." In the process of group cooperation, when the other person's point of view is inconsistent with one's own point of view, the Indexical 2nd-person pronouns are often used to highlight the inconsistent point of view, to show that one has an opposing point of view, and to emphasize one's own point of view. The tone of the Indexical 2nd-person pronouns is sometimes strong, and there are cases where one party in the communication orders the other party about. Students at this age have strong self-awareness and self-esteem, and the commanding tone may cause students to think that they are in a disadvantaged position. As such, they are more likely to raise conflicts in the discussion process to defend their status.

Fragment 1:

S3 Oh. Why don't **you** separate it?
S1 Do **you** think that painting is appropriate?
S3 then **you** close the window.
S1 study room must be the quietest.
S4 **You** will soon make the wall thicker.

Fragment 2:

S4 What is it now, do you know? An apartment belonging to the villa category.
S1 If it gives us three layers, right? Three-storey apartment. Each floor is 60 square meters.
S4 three sixty eight.
S1 OK, who fills it in, who fills it in.
S2 Let her draw later.
S1 **You** paint, **you** paint. It's okay to use mine.
S3 No, I haven't finished it yet.
S1 OK, **you** don't do it. Okay, I'll do it.

In the first segment, S3 asks "Why don't you?", S1 responds with "Do you think. is appropriate?" to express their own rebuttal. To which S3 proposes "You close the window" with a strong tone. Both parties of the communicator use Indexical 2nd-person pronouns which are inconsistent with each other. If one of the parties insists on their views, the conflict will escalate further. In Fragment 2, "Let's draw" and "None of you do it" highlights the problem of small group labor. The use of Indexical 2nd-person pronouns expresses dissatisfaction and helplessness with the division of labor. The discussion atmosphere of the group subtly descends into disharmony, which has a certain impact on the quality of the final plan.

Modals

Modal verbs mainly express possibility (can and may), necessity (must and should), prophecy (about to), while semi-modal words express the need to do something. Modal verbs expressing possibility are weaker than modal verbs expressing necessity, indicating that the speaker is not very sure, which leaves room for one's own expression and others' response. Modal verbs and semi-modal verbs that express necessity and prophecy are stronger, indicating that the speaker has a very positive attitude and wants others to accept his own point of view. Words that express possibility show a willingness to negotiate, which makes the discussion atmosphere more harmonious and can promote initial consensus among the communicators. Modal verbs that express more necessity may cause strong rebuttal from the other party and escalate the conflict. It is not conducive to the formation of the final plan.

Fragment 1:

S2 The toilet **must** be written smaller, the toilet, really, someone really has a two-square-meter toilet. Then write a larger bedroom, write a smaller kitchen, and write a larger living room. Make a plan. How many now? How about five square meters?
S4 It **should** be in circles.
S2 **may** have two toilets. toilet.
S4 **should** have a living room.

This is the discussion of the team members when planning the function and size of the room. S2 proposed that "the toilet must be smaller" and expressed his affirmative attitude. S4 replied that "it should be in a circle," which supplements S2's point of view and affirms their own view. S2 once again proposed that there are two toilets possibly with a weaker tone than before, indicating that he is now more open toward the opinions of others. To which S4 again replied "should have a living room" to indicate that he is more affirmative and his desire for a response from others. The use of modal verbs can promote the further deepening of the discussion and allow the team members to form a more satisfactory plan in the continuous argumentation.

Repetition

Repetition mainly refers to the repetition of words and sentences, with the purpose of emphasizing one's own point of view and attracting the attention of others. If there is no response, the communicator will re-emphasize it until the other person responds to his point of view. If a negative response is received, the communicator will look for a new basis again, re-emphasize it, and the discussion will be further advanced.

Fragment 1:

S1 You don't need to draw special, just mark each room, why use it, five rooms, three rooms, two rooms, **two rooms and one living room, one kitchen and one bathroom.**
S3 **Two rooms and one hall, one kitchen and one bathroom.**
S1 It is just right, **two rooms, one hall, one kitchen and one bathroom,** just right.
S3 **Two rooms, one living room, one kitchen and one bathroom,** almost the same.

This segment is a discussion on the function of the room. S1 first proposes "two rooms and one hall, one kitchen and one bathroom," S3 gives a repeated response before S1 and S3 repeat it again. The two emphasized the plan many times, highlighting their affirmative attitude toward the plan, hoping to get the attention and approval of others. In line with the language expression characteristics of students at this stage, they hope to emphasize their views through constant repetition before finding new grounds for argumentation.

Rebuttal questions

One of the group members raised rhetorical questions about the views or speech behavior of the other party. Such rhetorical questions often have a provocative tone which can easily lead to conflict. After one party in the communication raises a rebuttal question, the other party will feel that his views have been questioned and actively defend his position, which will cause conflict. At this time, it may turn from task conflict to relationship conflict, accompanied by emotional dissatisfaction of students. If both parties in the communication keep repeating rebuttal questions, the discussion will be put on hold, which delays the process and is not conducive to the formation of the final plan. If the rebuttal question is based on the task itself, it will encourage both parties to consider their views more comprehensively and demonstrate them better, which can promote discussion to a certain extent.

Fragment 1:

S2 If it is me, if I paint, just draw a three-dimensional one.
S3 1:100, 1:10,000, 1:10,000.
S1 You can't draw three-dimensional.
S2 **Why can't draw three-dimensionally?**
S1 Just can't draw three-dimensional.
S3 Just can't draw three-dimensional, because yours is a floor plan, **how do you draw three-dimensional?**
S1 This is a floor plan.
S4 **Isn't it the floor plan I just drew?**

This clip is a conflict between the team members about the drawing scale of the plan. S1 put forward "Can't draw three-dimensional" to which S2 asked "Why can't draw?" The tone was provocative because S2 thought that he could draw three-dimensional pictures. This triggered S1 and S3's rebuttal "You just can't draw because it is a floor plan" to which S4 again refuted "What I just drew is a floor plan." The rebuttal questions that appeared in the dialogue caused emotional dissatisfaction among other group members, which led to conflict. The use of rebuttal questions is more in line with the thinking development level of students of this age. Based on their existing learning and life experience, they already have their own way of thinking and will remain suspicious of the opinions of others. In the process of discussion, rebuttal questions can develop students' dialectical thinking to a certain extent.

Uptake avoidance

When a conflict occurs and the two sides are deadlocked, the team members will try to avoid the conflict entirely. The main manifestations of this include topic change, silence, and third-party intervention. Changing the topic is the most commonly used method. When the two parties in the communication cannot reach an agreement and one of them does not want to cause an argument, that party will choose to initiate the next topic to divert attention. When the other party feels that the dispute has not received a response, it automatically ends the argument. The use of topic avoidance can promote the process of group discussion to a certain extent, avoid the emotional dissatisfaction of group members due to disputes, and have a certain positive effect on the final formation of a high-quality plan.

Fragment 1:

S4 6 and 10, 8, and 7.5 are the best.
S2 The width of is 6.
S1 Yes, you draw 6 and 10, then you two draw 6 and 10, and we draw 7 and 15.
S4 **I feel special square.**
S2 Yes, let's see who is better, and then design our own design.
S1 Draw the frame first. You can use the pencil. I don't have a scale. Who can lend me? I don't have a scale. Thank you.

When S4 put forward "I think it's a special square," other team members ignored it and continued to determine the "overall area of the room" based on the opinions of other team members. If only one group member puts forward a different opinion, the topic will often be avoided and the discussion continued, which has a positive effect on maintaining the feelings between the group members.

The overall distribution trend of language features of conflict discourse

The general trend analysis of the language characteristics of conflict discourse takes the lexical level analysis as the object. So seven types of conflict discourse are selected for analysis. Based on the statistical results of computer software analysis, the linguistic feature statistics of the vocabulary level of conflict discourse in group cooperation are shown in the following table (**Table 3**).

TABLE 3 Statistical table of language features at the lexical level of conflict discourse.

Language features at the lexical level	Frequency
Extreme summaries	483
Negation	886
Discourse markers	529
Emphatics	68
Floor bids	16
Indexical 2nd-person pronouns	1154
Modals	622

The above data shows that at the word level, the Indexical 2nd-person pronouns occurred at the highest frequency. In the process of interaction, "you" was used the most, the use of these words not only shows that the individual has a clear direction as well as the opposition and distance between oneself and the other party, but also accuses the other party (Connor Linton, 1989); followed by negative form words. The application of negative form is mainly to express dissent and negate the other party's point of view. The application of these words is most likely to cause conflicts between group members; The third is the application of modal words such as can, possibly, must, should, will, etc. The application of these words reflects some of the negotiating significance of the team members in the problem-solving process; There is also the application of discourse markers such as right, then, but, and so on. The analysis found that Floor bids occurred the least. In actual classrooms, it is possible that some students compete for the right to speak but do not necessarily use floor bids.

In summary, although the general trend of language features in conflict discourse is a more general overview of the language features of group cooperation, it can also reflect some of the students' language preferences in this process, which can serve as reference for teachers looking to intervene in cooperative classroom teaching. In the 1950s, the speech act theory was first proposed by the British philosopher J. L. Austin. After this theory, Zhao (2009) analyzed the conflict discourse and proposed three stages: initial stage of conflict, Conflict and Negotiation stage and the end of conflict stage. The following will analyze the language characteristics from these three aspects.

Linguistic features of "the initial stage of conflict" in mathematical cooperative problem solving

This section mainly analyzes the initial stage of conflict discourse and uses fragments to explain its language expression and characteristics.

Language expression in the initial stage of conflict discourse

Explanatory statement ↔ Negative response

In verbal interaction, the description of objective facts is a more frequent verbal act. In the course of the presentation, the two interacting parties will respond accordingly to their different views on the content of the presentation.

Fragment 1:

S1 **Our door is usually on the corner, right? let's confirm the door first, this corner.**
S2 **No, you don't need to draw this one inside.**
S1 No, you have to draw like this.
S2 You have to calculate the length and width.
S1 No, what you are asking for is the total area. Look, look at me, the door is facing this side, and draw this side. In this way, it will have a small corridor, right? The small corridor is here, so here is the locker, right next to the door is the locker.

The conflicting discourse that begins with the explanatory statement is caused by the speaker's different views on people and things. According to corpus research, the speaker and responder in conflicting discourse are mainly in the mode of "explanatory expression and negative response." In the form of narrative, affirmation, appreciation, and approval of the speaker, what is waiting is the direct opposition, denial, questioning, or interruption of the other party's response, which leads to conflict.

Instruct↔refuse

In verbal interaction, the speaker issues instructions, suggestions, requests, and orders to the other party to engage in a certain behavior. After the other party responds with confrontation, rejection, questioning etc., the verbal conflict will begin. The expression of this "instruction rejection" mode occurs in the form of imperative and declarative sentences. E.g.:

Fragment 1:

S4 There are only five rooms, so save a bit.
S1 No need to save.

In this dialogue, S4 suggested to save a little while allocating area, and S1 directly rejected it. In the verbal interaction, the caller's instruction hopes to get an affirmative and accepting answer, and the recipient S1 can either obey the instruction of S4 or refuse. Obviously, in the above case, the recipient refuses to accept the other party's instructions, which leads to the beginning of a frontal conflict between the two sides.

Seditious inquiry ↔Confrontational answer

In verbal interaction, one question and one answer often leads to conflict. Questions are trigger words, and answers are responses to inflammatory or provocative questions from the speaker, as per the following example:

S4 Zhang Yang, what kind of ghost do you paint?
S2 Go! Don't insult my creativity.

In the above-mentioned conflicts, S4 issued an inflammatory question that was discriminatory to a certain extent. The other party, S2, was not to be outdone, leading to a frontal conflict between the two parties. The inflammatory questioning was also an important reason for conflicting discourse.

The linguistic features of "the initial stage of conflict" in conflict discourse

The initial stage of conflict is the origin of the entire conflict process and plays a vital role in the occurrence of conflict. By categorizing the expressions of conflicting discourse in this stage, this article finds that the linguistic characteristics of the "initial stage of conflict" have the following points:

The use of accents or phrases

Emphasized words or phrases are used to enhance the degree of expression, which can be adjectives, adverbs, interjections, pronouns, or modal particles. In conflict conversations, both women and men tend to use emphasized words or phrases to enhance their feelings, but women are better at using such words because they have a rich emotional network. Men often use emphasized words or phrases to express their differences. In the initial stage of the conflict, the two parties in the conflict often choose modal particles, such as "ba," "le," "of," "ah," "ma," "ah," etc.; there are also pronouns, such as the first person "I." When they have a conflicting conversation, they use these words to express their strong emotions.

The use of negative words

Most conflicts are caused by disagreements between communicators, and they often use the negative word "no" to express their disagreements. In the initial stage of the conflict, both parties to the conflict usually choose the negative words "no," "can't," "impossible," and so on. According to the face threat theory, the use of negative words by one of the two parties will inevitably threaten the "face" of the other party, and the conflict will further escalate.

The use of imperative sentences

The imperative sentence has the function of being a command, instruction, or warning. If there is "please" before the imperative sentence, it will be regarded as a request. Negative imperative sentences are used to prohibit someone from doing something or to discourage someone from doing something. In dominant interactive networks, imperative sentences are highly used. In the initial stage of conflict, some group leaders will use their authority to use imperative sentences to command group members, which further intensifies the conflict. For example, "You paint, don't make trouble, you don't make trouble, take it away," "You give up," "You don't, you don't," "The living room must be the largest."

10.3389/fpsyg.2022.945909

Language features of "conflict and negotiation stage" in mathematical cooperation problem solving

In the process of group cooperative learning, the conflict stage is the core part of a conflicting verbal event, and it is also the main part of the conflict from the beginning to the final agreement. After the conflict begins, it usually takes only one or two rounds for the members to reach a consensus, but sometimes it takes many interaction rounds between the group members to reach a consensus.

Based on the analysis of text data, this article mainly discusses the language expression methods and language characteristics of the "conflict phase" of cooperative problem resolution. The language expression of the "conflict stage" mainly includes several types such as the rebuttal questions, the direct response, the explanation, and the negative avoidance.

The language expression of "conflict and negotiation stage"

Rebuttal questions

In the process of group cooperation problem-solving conversation, when the listener expresses opposition to the speaker's point of view, provocative counter-questioning is the most common way of expressing conflicting discourse.

This form of rebuttal question can easily lead to re-rebuttal by the speaker, leading to further intensification of conflicts.

Fragment 1:

S3 now has four.
S2 four, plus one.
S3 storage room.
S2 Storage room, let me think about it.
S4 balcony.
S1 balcony.
S3 **Does the balcony count? (With a provocative tone)**
S2 Balcony is counted.
S4 Balcony does not count as a room.
S3 Balcony does not count as a room.
S3 Balcony does not count as a room.
S4 It is not a room, it just occupies an area.
S3 Yes, it occupies an area, but not a room.

Direct response

In the process of discourse conflict, facing the conflict caused by the speaker, other team members responded in a straightforward manner. Based on the face theory, the straightforward denial or questioning of one party's point of view further aggravated the conflict.

For example, in the case of "explanatory statement and negative response," S1 and S2 had a fierce discourse conflict on the issue of "door." The dialogue was conducted in a straightforward manner such as "No, you don't need to paint this one," "No, you have to paint like this," etc. The attitudes of both sides were very straightforward, which led to further intensification of the conflict between the two.

Explanation

After a conflict occurs, in the face of questions and inquiries from the speaker, the listener gives the reason for the objection, which help resolves the discourse conflict. The explanation refers to the dialogue formed by the listener facing the inquiry or question raised by the speaker during the utterance interaction of the group members.

Fragment 1:

S4 Isn't it awkward to put the kitchen next to the toilet?
S2 Not awkward, anyway, it's not your own home, really.
S1 No, no, the kitchen and the toilet should not be next to each other, the kitchen should be next to the bedroom, and the living room should be next to the kitchen.

When S4 put forward a point: "Is it awkward to put the kitchen next to the toilet?" Both S1 and S2 gave their own opinions and explained their reasons. S2 thinks "not awkward" because "it is not your own home anyway," which means you can design whatever you want. However, S1 thinks this design is unreasonable, because "the kitchen and the toilet can't be next to each other," the kitchen should be next to the bedroom, and the living room should be next to the kitchen.

Negative avoidance

In the process of discourse conflict, when the listener disagrees with the speaker's point of view, he will often adopt an avoidance method to express his negative point of view instead of making a positive reply. There are three main forms of avoidance: intentional avoidance, temporary avoidance, and unintentional avoidance. After the listener avoids the topic, if the speaker insists on turning the topic back, it may lead to an aggravation of the conflict. But on the whole, when one party has entered negative avoidance mode, it indicates that one party wants to end the discourse conflict, which transitions the entire conflict event toward the end.

Fragment 1:

S2 wait a minute, one, two, three, four, one more time.
S1 restaurant.
S3 The kitchen should be with the dining room.
S1 restaurant and kitchen, kitchen has not been written yet.
S2 It is enough, the kitchen is 10 square meters, is 60 enough?
S1 The rest is the kitchen.
S3 Why do you want 10 square meters if you don't eat?
S2 20, 30, 35, It's still 15 square meters short, almost, the long one is almost the same as the bedroom.

S3 Our home is also a long strip.
S4 yours too.
S2 draw a picture.
S4 draw it.

S2 advocated "10 square meters in the kitchen" to which S3 asked "Why do you want 10 square meters if you don't eat?" – expressing denial. S2 did not respond to this question directly and said "Draw a picture" instead, which aggravated the conflict no further.

The linguistic features of "conflict and negotiation stage" in conflict discourse

This stage is the key stage in the entire conflict. During this process, the conflict between the two interacting parties escalates rapidly, and a large number of conflicting words will be produced, showing strong antagonistic characteristics and possibly even some offensive words. The research in this paper finds that the language characteristics of the "conflict phase" are as follows:

Rhetorical questions

Rhetorical questioning is a way of expressing negation through rhetorical questioning using truths or facts with distinctive features to achieve the purpose of strengthening tone, emphasizing semantics, and expressing blame. In the "conflict phase," both parties will usually use rhetorical questions to express their strong doubts about each other. This will lead to the climax of the conflict where the two parties will refuse to give in to each other, which will escalate the conflict. For example, "Is your house made of walls?" "The row is so full, can you still paint?" "Then what do you say this is?" "A two-meter bedroom, are you still sleeping?"

Negative comments

In the process of group cooperative learning, group members' opinions on the problem are evaluated negatively through negative words, which imply accusations and criticisms from the speaker. In the "conflict phase," both interacting parties will express their disagreement through the use of negative evaluation terms. According to the theory of adaptation, both parties are neglecting to adapt to the contextual factors which will escalate the conflict and eventually result in a stalemate. For example, "Shut up, you shut up." "Can you stop being so funny?" "Okay, you." "Can we still play happily?"

Direct negation

Direct negative words are often expressed as negative words "no/not + ……" and other structures. That is, to refute by directly negating the proposition stated by the other party, which is manifested when A says proposition P, and B says¬P. In the "conflict phase," the two interacting parties will express their own confrontation by directly denying the other's views, which will escalate the contradiction. For example, "No, I think about it. This bedroom is connected to the living room, because my bedroom is connected to the living room." "No, it needs a balcony." "No, like me, the bedroom, the kitchen." "No, the balcony cannot live in."

Linguistic features of "end of conflict" in mathematical cooperation problem solving

In cooperative learning, the ultimate goal of discourse conflicts among group members is to form a unity of viewpoints and achieve the purpose of problem solving. No matter how intense the process of discourse conflict is, the conflict on a certain issue will eventually end. This section mainly introduces the expression methods and language characteristics of the ending stage of discourse conflict.

Ways of language expression at the end of the conflict

How the conversation ends in discourse analysis is also an important analysis content. In the discourse conflict of group cooperation problem solving, the group members finally reached a unity of opinions after several rounds of disputes and negotiations, which ended the discourse conflict. The way the group members end their conflict and their language characteristics are the focus of this section.

Topic-shifting

When the group members argue about a certain point of view, and the two sides are in a stalemate with each other, one of them tries to terminate the topic and find another way to talk about the topic instead. After a period of arguing between the two parties, one party changes the subject and the conflict ends. This not only preserves the face of both parties in the conflict, but also easily achieved a win-win situation for both parties in the conflict.

Fragment 1:

S1 Shouldn't we list it first, why do we have to draw it first. The house doesn't have to be square.
S2 We didn't say to draw a square.
S1 Our opinions cannot be unified at all.
S1 What else is there besides the living room, toilet, and bedroom?
S3 kitchen.
S1 Yes.
S2 What to add, as for the specific range, 1, 2, 3, 4, 5, a total of five.
S1 One more.

S1 and S2 had different opinions on the overall shape of the house at first, and a conflict occurred. Then S1 changed the subject and questioned the functions of the five rooms. The resolution of the previous conflict resulted in a new round of discussions.

Compromise

Fragment 1:

S2 The toilet must be written smaller. The toilet, really, some people have a two-square-meter toilet. The bedroom should be larger, the kitchen should be smaller, and the living room should be larger. Make a plan. There are a few now. How about five square meters?
S2 This is not enough, these five are so big, plus this, so big, stand on the toilet and take a shower.
S3 10 square meters.
S1 10 square meters are too big.
S3 9.
S2 7 square meters.
S1 Made do with it.

S2 advocates that the toilet should be written smaller, S3 said "10 square meters," S1 said "10 square meters is too big," S3 said "9," and finally S2 said "7 square meters" to which S1 agreed to "make do with it." The three people had different ideas about the size of the toilet, and finally compromised on "7 square meters."

Third-party intervention

When two parties in conflict are arguing about a certain point of view, the intervention of the third party can play a coordinating role to end the conflict. In many cases, this type of ending plays a regulatory role that can promote the rapid unification of opinions.

Fragment 1:

S4 Is your length drawn longer? Is your width drawn longer? Is your width drawn longer?
S3 2 cm. 2.
S4 Shrink it.
S3 You cannot change the number of square meters.
S4 Toilet is actually suitable by this. Then you are directly here, and the bedroom is directly here. I should rely on this.
S4 How big is the bedroom?
S1 Let's not discuss it yet.

S3 and S4 were in conflict over the length and width of the room, and they also had different opinions on the layout of the room. As a third party, S1 intervened in their conflict and asked them to "not discuss it yet" to resolve the conflict.

One-sided win

Compared with the conflict ending method in which participants voluntarily compromise, a one-sided win is achieved when one party conveys a message through verbal expression or physical action and compels the other party to accept his own point of view. Although this ending temporarily ends the discourse conflict, it may also create hidden dangers for the subsequent relationship between the two parties, which leads to the emergence of new conflicts.

Fragment 1:

S2 The toilet can be smaller.
S4 Toilet is at least bigger than the balcony, right? Is not it?
S1 The toilet is bigger than the balcony. Our house is 40 or 50 square meters.
S1 Our public toilet, can you manage it?
S3 10 square meters public toilet.

S2 advocated that the toilet should be smaller, while S1 advocated that the toilet should be larger than the balcony, and said "Can you manage it?" The attitude was tough, so other students had to accept his idea. This temporarily ended the conflicting dialogue.

The linguistic features of "the end of conflict" in conflict discourse

The "end phase" of the conflicting discourse is ended by the two sides of the conversation changing the topic, making a compromise, or reaching an agreement through negotiation. Past research has shown that language in interpersonal communication has two pragmatic purposes: convergence and divergence orientation, and harmony orientation is a basic orientation for group members in cooperative learning. Harmony orientation will affect the choice of speech forms and interaction strategies (Spencer-Oatey, 2000). At this stage, the two parties in the communication pass through the previous conflict initiation and escalation stages, and some of the opinions are agreed while some are still deadlocked. However, both parties will also compromise or evade to end the conflict. The tone of the two parties or one of them is not as strong as before, and they are in a state of wanting to end the argument. According to the theory of adaptation, both parties have the awareness of adapting to the context and no longer overemphasize their own ideas, so that conflicts are gradually resolved. The language at this stage has the following characteristics:

Moderation of tone and the appearance of mitigator

The moderation of tone and language are to reduce and weaken the intensity of certain language factors in verbal interaction, and to reduce the risks of interpersonal conflicts and face-threatening behaviors in the interaction, so as to ensure the smooth progress of the interaction (Yu, 2015). At the end of the conflict discourse, in order to achieve the goal of harmony, the two parties in the conflict often use mitigation methods or words such as changing the topic and compromising to alleviate the conflict. Moderation language has the function of realizing effective communication and constructing a clear interpersonal identity. It reflects the strategic and practical characteristics of moderation language. Moreover, moderation language generally follows the basic principle of "face threat" proposed by Brown and Levinson (1987). In the process of discussion and exchange in the cooperation group, the other party's "face" will also be taken into consideration and compromise is adopted to end the conflict. In this study, the emergence of a moderating tone was mainly accompanied by facial expressions and body language. Members often used "laughing" to ease conflicts.

Conflict end discourse reflects cooperation

Cooperative learning is to use students as human resources in teaching to make up for the inadequacy of teachers and to cultivate their cooperative spirit, rather than simply shrinking the self and enlarging the collective. In cooperative learning for the purpose of solving mathematical problems, conflicting discourse is the result of the combined effect of many factors such as context and discourse structure. This stage can best reflect the cooperative tendency of "seeking common ground while reserving differences." To achieve cooperation tasks, even if there are differences in opinions between the two parties in the conflict, one of the parties will choose to compromise, avoid the topic, or stay silent to reach an agreement with the other party. Through the sharing of resources among students during the interaction of group members. Students have their own subjectivity, the interaction between the subject and the subject using dialogue as a means and understanding as the goal, with the goal of reaching a consensus. The different ways of thinking and different understandings of the problem among the team members make the conflict dialogues constructive.

Summary

This study mainly analyzes the general linguistic features of Conflict Discourse in Chinese students' cooperative problem solving, including the classification of conflict discourse language characteristics and the conflict discourse vocabulary level before exploring its distribution trend. The linguistic features of conflict discourse include extreme generalizations, negative forms, discourse markers, emphatic words, turn-taking words, second-person pronouns, modal words, repetition, rebuttal questions, and topic avoidance. Among them, the frequency of Indexical 2nd-person pronouns is the highest; the frequency of negation, extreme summaries, discourse markers, and modals lie somewhere in the middle with little differentiation; the frequency of emphatics and floor bids is the least.

The frequency of language features at the lexical level can reflect the language features of conflicting discourse in the process of cooperative mathematical problem solving. However, there are many factors affecting it, such as social factors, participant status, role, scene, and discourse structure factors such as discourse type. These deep-seated reasons will affect participants' choice of language forms such as negation, modals, discourse markers and questions.

Secondly, based on the three stages of conflict discourse, the discourse style and language characteristic analysis were carried out respectively. Among them, the language expressions of the "initial stage of conflict" include Explanatory statement ↔Negative response, instruct↔refuse and Seditious inquiry ↔Confrontational answer. The language shows the characteristics of using emphatic words or phrases, negative words, imperative sentences and so on. These types of sentences or phrases will further intensify the conflict. In conflict conversation, both boys and girls tend to use emphatic words or phrases to enhance their feelings, but girls are better at using this kind of words. Boys often use emphatic words or phrases to express their differences, and often use the negative word "no" to express dissent. Some leaders of small groups will use their authority to order group members with imperative sentences, which further intensifies the conflict. Rebutting interrogative, straightforward, explanatory, and negative avoidance type are the language expressions of the "conflict stage." The conflict between the two sides of communication escalates, resulting in the largest number of conflict words. The language shows a strong antagonism, and some offensive words will appear It also exhibits the verbal characteristics of rhetorical question, negative comments, and direct negation. The use of these types of language will bring the conflict to a climax, and the two parties will be deadlocked. Topic-shifting, compromise, third-party intervention, and one side wins are the linguistic expressions of the "end of conflict." At this stage, both parties involved in the communication begin to have a sense of compliance, their tone of voice is relaxed, and mutual compromise or concession. As a result, conflicts are gradually resolved, tend to end, gradually reach consensus, and tend to be completed through cooperation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee School of Mathematical Sciences, Beijing Normal University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

JZ: research design, data collection, data interpretation, and wrote the main manuscript text. TS, XS, and YB: data collection and data interpretation. All authors reviewed the manuscript.

Funding

This study was supported by Hainan Provincial Natural Science Foundation of China (622RC670) and Project of the 13th Five Year Plan of Educational Science in Hainan Province of China (QJY20191071).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alexopoulou, E., and Driver, R. (1997). Gender differences in small group discussion in physics. *Int. J. Sci. Educ.* 19, 393–406. doi: 10.1080/0950069970190403

Alschuler, A., Atkins, S., Irons, R., McMullen, R., and Santiago-Wolpow, N. (1977). Social literacy: A discipline game without losers. *Phi Delta Kappan* 58, 606–609.

Brown, P., and Levinson, S. C. (1987). *Politeness: Some universal sin language usage*. Cambridge: Cambridge University Press.

Bruxelles, S., and Kerbrat-Orecchioni, C. (2004). Coalitions in polylogues. *J. Pragmat.* 36, 75–113. doi: 10.1016/S0378-2166(03)00037-7

Cao, M., and Bai, L. (2018). Cooperative learning process model for mathematical problem solving and its application. *Audio Vis. Educ. Res.* 39, 85–91.

Cao, Y., and Zhu, Z. (2019). Change and invariance: Evolution of PISA 2000-2021 mathematics evaluation framework. *J. Math. Educ.* 28, 1–5.

Cazden, C. B. (2004). What goals should schools be trying to meet? Essay review of psychological theory and educational reform: How school remakes mind and society by David R. Olson. *Hum. Dev.* 40, 327–330. doi: 10.1159/000079162

Connor Linton, J. (1989). *Crosstalk: A multi-feature analysis of soviet-american spacebridges.* PhD dissertation. Los Angeles, CA: University of Southern California.

Goos, M., and Galbraith, P. (1996). Do it this way! Metacognitive strategies in collaborative mathematical problem solving. *Educ. Stud. Math.* 30, 229–260. doi: 10.1007/BF00304567

Hao, J., and He, Q. (2019). Evaluation and practice of cooperative problem solving ability. *China Exam.* 11–21.

Li, X., and Cao, Y. (2019). Cultivation of cooperative problem solving ability – Research and Enlightenment Based on high-quality mathematics teaching in the United States. *Res. Educ. Sci.* 79–84.

Liu, D., and Wang, X. (2008). Foreign research on group cognitive process – new ideas for cooperative learning research. *Glob. Educ. Outlook* 41–45.

Lin, D. (2016). *Research on the core quality of student development in the 21st century.* Beijing: Beijing Normal University Press, 106–109

OECD. (2020). *Multilingual Summaries PISA 2015 Results (Volume V): Collaborative Problem Solving [EB/OL].* Available online at: http://www.keepeek.com/Digital-Asset-Management/oecd/education/pisa-2015-results-volumev/summary/chinese_2a7898c2-zh#page1 (accessed June 26, 2018).

Scott, S. (2002). Linguistic feature variation within disagreement: An empirical investigation. *Interdiscip. J. Study Discourse* 2, 310–318.

Shi, L., and Du, W. (2008). Cognitive conflict: Multilateral dialogue in cooperative learning. *Educ. Acad. Mon.* 85–87.

Spencer-Oatey, H. (2000). "Introduction: Language, culture and rapport management," in *Culturally speaking: Managing rapport through talk across cultures*, ed. H. Spencer-Oatey (London: Continuum).

Sun, L., and Fang, Q. (2011). A comprehensive study on the types and functions of Chinese discourse markers. *Chin. Learn.* 76–84.

Wang, H. (2004). Review of cooperative learning cognitive research. *Psychol. Sci.* 438–440.

Yang, R., and Wang, J. (2019). International Research progress and trend of collaborative problem solving:based on bibliometric analysis of educational studies in the web of science database. *For. Prim. Second. Educ.* 10–18.

Yu, G., and Cao, Y. (2017). Interdisciplinary teaching research: Taking finnish phenomenon teaching as an example. *For. Prim. Second. Educ.* 57–63.

Yu, Z. (2015). Pragmatic function of discourse marker "what do you say". *J. Guangdong Univ. For. Stud.* 26, 44–48.

Zhang, L. (2020). Partner learning under the young gentleman system: Relationship and connotation. *Educ. Theory Pract.* 40, 47–51.

Zhao, Y. (2009). Argumentation discourse analysis in Chinese conflict speech events. *J. Northeast Univ. (Soc. Sci. Edn.)* 11, 449–454+460.

TYPE Original Research
PUBLISHED 22 September 2022
DOI 10.3389/fpsyg.2022.941253

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Lietta Marie Scott,
Arizona Department of Education,
United States
Animasaun I. L.,
Federal University of Technology,
Nigeria

*CORRESPONDENCE
Chris Sakellariou
acsake@ntu.edu.sg

SPECIALTY SECTION
This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 11 May 2022
ACCEPTED 18 August 2022
PUBLISHED 22 September 2022

CITATION
Sakellariou C (2022) The reciprocal relationship between mathematics self-efficacy and mathematics performance in US high school students: Instrumental variables estimates and gender differences. *Front. Psychol.* 13:941253.
doi: 10.3389/fpsyg.2022.941253

The reciprocal relationship between mathematics self-efficacy and mathematics performance in US high school students: Instrumental variables estimates and gender differences

Chris Sakellariou*

Nanyang Technological University, Singapore, Singapore

Objective: To investigate the reciprocal relationship between high school students' academic self-efficacy and achievement in mathematics using US data from the HSLS:2009 and first follow-up longitudinal surveys, while accounting for biases in effect estimates due to unobserved heterogeneity.

Methods: Instrumental Variables (IV) regressions were estimated, to derive causal effect estimates of earlier math self-efficacy on later math achievement and vice versa. Particular attention was paid to testing the validity of instruments used. Models were estimated separately by gender, to uncover gender differences in effects.

Results: Evidence of robust reciprocal effects between self-efficacy and achievement for male students is presented, with the dominant effect from earlier achievement to later self-efficacy. For girls, evidence of such effects is weak. Generally, IV estimates are higher than OLS estimates for males, but not for females. As opposed to earlier correlational studies which did not find significant gender differences despite theoretical expectations for their existence, the findings support higher effects for male students.

KEYWORDS

mathematics achievement, mathematics self-efficacy, longitudinal data, unobserved heterogeneity, gender differences

Introduction

The theoretical and empirical literature has examined the role of non-cognitive, domain-specific constructs, referred to as student self-beliefs (such as self-efficacy and self-concept), in predicting behavior, choices, and practices which can affect achievement. Bandura (1982) defines perceived general self-efficacy as a personal judgment of "how well one can execute courses of action required to deal with prospective situations." Self-efficacy is domain-specific and multidimensional, and beliefs vary across wide ranges of activities. Academic self-efficacy, as opposed to general self-efficacy, relates to students' confidence in their ability to complete academic tasks like studying for examinations and writing term papers, and should be measured in academic settings (e.g., Animasaun and Abegunrin, 2018). Developmental origins of self-efficacy perceptions include familial sources, peer influences, and transitional influences of adolescence. Since academic self-efficacy influences students' approach to confronting educational challenges, a higher academic achievement is expected for students with higher self-efficacy (Pajares and Schunk, 2001). Past studies have described self-efficacy as a positive predictor of performance outcomes (e.g., Schunk et al., 2008). According to Usher and Pajares (2008), self-efficacy "predicts students' academic achievement across academic areas and levels. Past successes in (non-trivial) tasks consistent with mastery in a particular domain (e.g., mathematics) increase students' confidence in succeeding again. In other words, self-efficacy is likely both the cause and the effect of academic achievement (see Pajares and Schunk, 2002). Student motivation is interlinked with specific student self-efficacy beliefs (for example, in doing well in a mathematics test) and such beliefs do not coincide with students' academic ability, since they are self-constructed.

It has been proposed that there may be too much or too little self-efficacy and that the optimum level seems to be slightly above a student's true capacity (Bandura, 1977). Students may and often do misperceive their true skills, which could have complex effects on students' motivations (Pajares, 1996, p. 565). Grossly overestimating self-efficacy might lead a student not preparing for a task properly, which will impair performance. Emerging evidence suggests that men overestimate both their ability as well as their perception of past performance, while women either underestimate both or at least rate them more accurately, while in most situations, actual performance in mathematics does not differ significantly between genders (e.g., Bench et al., 2015). Chatard et al. (2007) found that, the more students believed in gender stereotypes prior to recalling their marks, the more female students underestimated their marks in mathematics and male students underestimated their marks in arts. With respect to empirical evidence on how biases in self-evaluations of math competence relate to achievement, there is no consensus. Advocates of the view that being optimistic about one's efficacy is required for a variety of accomplishments and is associated with higher motivation, persistence and performance include Taylor and Brown (1988) and Bandura (1993). The opposite view states that overly optimistic or overly pessimistic self-assessments may lead disappointment and loss of morale, following repeated failures (Schunk, 1981). Related empirical findings are mixed. Some studies found that overstating one's competence is associated with higher academic performance (e.g., Martin and Debus, 1998). On the other hand, Robins and Beer (2001) found no association between overstating competence and receiving better grades or likelihood of graduation. Dupeyrat et al. (2011) found that overrating one's mathematics competence was related to better mathematics achievement, although the findings were open to interpretation.

Section "Self-efficacy and mathematics achievement using longitudinal data" reviews the more recent literature investigating reciprocal effects between self-efficacy and achievement using longitudinal—repeated measure data. Section "Contribution of study and objectives" outlines the contribution and objectives of the study.

Self-efficacy and mathematics achievement using longitudinal data

Studies on the relationship between self-beliefs and performance often refer to the notion of reciprocal determinism at a theoretical level (Bandura, 1986). However, the reciprocal relationship between self-efficacy and performance has not found direct empirical support; this could be due to the relative scarcity of longitudinal, repeated-measure data, as opposed to cross-sectional data. For example, Williams and Williams (2010) investigated the reciprocal determinism of self-efficacy and mathematics achievement in 15-year-old students using cross-sectional data.

Of the models which assume a causal relationship between self-efficacy and achievement, the *self-enhancement model* proposes that the dominant effect is through high prior self-efficacy enhancing achievement in a particular domain. The *skill development model*, on the other hand, proposes that there is a causal association between achievement and future self-efficacy, because academic success (failure) improves (diminishes) self-efficacy (Usher and Pajares, 2008). The *reciprocal-effects model* stresses the likely role of self-efficacy as both a cause of and an effect of academic achievement and intends to integrate the causal relationships proposed by the two alternative models (Marsh, 1990).

Various methodological approaches have been used in assessing the relationship between self-efficacy and achievement, including meta-analytic studies, studies using longitudinal-repeated measures data and approaches such as structural equation modeling. Valentine et al. (2004) synthesized the findings of longitudinal studies investigating the relationship between self-beliefs and achievement and found a small influence of generalized positive self-beliefs on academic achievement after controlling for initial levels of achievement, and somewhat stronger effects when measures of self-beliefs and achievement are matched by academic domain. Valentine and

DuBois (2005) found stronger evidence for the *skill development model* compared to the *self-enhancement model*.

More recent studies using longitudinal data include Hannula et al. (2014), Hwang et al. (2016), and Schöber et al. (2018). Hannula et al. (2014) used Finish longitudinal data and autoregressive/cross-lagged models to estimate the reciprocal causal relationship between mathematics enjoyment, mathematics self-efficacy, and mathematics achievement. They found that mathematics achievement and self-efficacy have a reciprocal relation and that the dominant effect is from achievement to self-efficacy. Hwang et al. (2016) used data on Korean students in the liberal arts track and an autoregressive cross-lagged model to assess the casual ordering of self-efficacy beliefs and academic achievement over short time intervals. Performance scores were from Korean, English, mathematics, and social studies subjects. They found a reciprocal relationship, with the effect of past academic achievement on self-efficacy beliefs being stronger than the effect of self-efficacy beliefs on academic achievement. One of the reported limitations of the study is that the design of the study does not allow for a clear conclusion regarding the reciprocal causal influence between self-efficacy and academic achievement, as it does not account for unobserved confounding factors. Schöber et al. (2018), did not find support for the *skill development model*, while their estimate of the positive effect of mathematics self-efficacy on later mathematics achievement was small and in line with that of Valentine et al. (2004). The authors state that, since the influence of self-efficacy on achievement is mediated by other variables and such variables were not available, the processes underlying the causal link from self-efficacy to later achievement and vice versa could not be examined. Furthermore, the findings of the study (lack of consistent reciprocal effects) may be due to the short interval between measurement occasions.

One recent study by Zakariya (2021) aimed to derive evidence of a causal relationship between self-efficacy and performance in mathematics, using an instrumental variable approach to structural equation modeling. Unfortunately, besides the small sample size, the findings (large bidirectional effects between math self-efficacy and math performance) are not generalizable, as the focus is on first-year engineering students following a mathematics course at a Norwegian university. Finally, the above cited studies were not designed to derive gender-specific evidence on reciprocal effects.

Contribution of study and objectives

Reflecting on the empirical literature, estimation of the reciprocal causal relationship between self-efficacy and achievement benefits from the use of longitudinal data with multiple measurement occasions and adequate time interval between measurements. Furthermore, the methodology, needs to account for the potential endogeneity of students' self-efficacy perceptions and mathematics performance, when attempting to estimate bidirectional causal effects between mathematics self-efficacy perceptions and mathematics achievement. By endogeneity, I refer to a covariate appearing in a model, which is correlated with unobserved student characteristics which also affect the outcome. While the terms endogeneity and unobserved heterogeneity often used interchangeably, unobserved heterogeneity refers to variation/differences among subject participants which are unmeasured, hence omitted from the model.[1] Endogeneity, on the other hand, can arise from three sources[2]: omitted variables, measurement error, and simultaneity. Presence of any of the three, or a combination of them can bias effect estimates. Hence one needs to use an estimation method which accounts for potential endogeneity of covariates of interest, to derive bias-corrected effect estimates. In this investigation, mathematics self-efficacy perceptions (when assessing the effect of earlier self-efficacy on later performance) and achievement (when assessing the effect of earlier performance on later self-efficacy) are treated as potentially endogenous covariates.

The following research questions were investigated:

> **RQ1**: Is there evidence that students' self-efficacy perceptions (Model 1) and/or students' mathematics achievement (Model 2) are endogenous covariates? If this is the case, the OLS effect estimates will be biased, and one needs to use an estimation method which accounts for such biases.

> ***RQ2***: Is there evidence in support of a reciprocal relationship between academic self-efficacy and academic achievement?

> ***RQ3***: Are there gender differences in relation to RQ1 and RQ2?

Materials and methods

Section "Data: The HSLS:09 and first follow-up" provides information on the HSLS:09 and follow up surveys and associated datasets, while Section "Methodological approach" outlines the methodological approach.

1 For example, unobserved mediators such as familial environment and peer influences may be involved when investigating earlier self-efficacy as a predictor of later achievement, while unobserved student innate academic ability or effort may be involved when investigating earlier achievement as a predictor of later self-efficacy.

2 While it is helpful to identify these "sources" separately, they are not truly distinct. For example, if a variable which truly measures self-efficacy is not available, this can be described as an omitted variables problem, or a problem associated with the available measure of self-efficacy being mismeasured/misreported. Understanding unobserved heterogeneity is facilitated by understanding endogeneity in a regression context.

Data: The HSLS:09 and first follow-up

I used data from the US High School Longitudinal Study (HSLS:09) and the 2012 follow up, in which students who were in grade 9 during the 2009–2010 school year were followed for 7 years into high school graduation, and college or employment. The survey, implemented by the U.S. National Center for Education Statistics (NCES) contains measures of students' mathematics performance (mathematics assessment in algebraic reasoning) and mathematics self-efficacy beliefs, both at two points in time (at grades 9 and 11), which is needed for testing reciprocal effects. The HSLS:09 has been described as the ideal dataset for examining determinants of student outcomes, such as who majors in STEM fields (Cimpian et al., 2020).

The survey followed a nationally representative sample of fall-term 9th-grade students in 2009 in more than 900 public and private high schools to 2012 (first follow-up) when most sample members were in 11th grade, and into higher education or the workplace (second follow-up). Students completed an in-person mathematics assessment focused on algebraic reasoning, as well as a survey that included items on educational experiences, sociodemographic background, educational expectations, mathematics and science student self-beliefs and interests, among other items. Students' parents, principals, teachers, and school counselors participated in the surveys. In the first follow-up, which took place in the spring of 2012, each module from the base year was fielded again, except for the teacher questionnaires (U. S. Department of Education, 2018). The student-level files contain student responses and associated composite variables from the student, parent, and school administrator survey instruments. The school-level variables contain responses and associated derived variables from the school administrator and counselor instruments. In the public-use datafile, school and counselor/administrator survey data has already been merged into the student-level file.

The data files contain school and student level weights, including longitudinal weights appropriate for analyses which use variables from both the base year and the first follow-up (as is the case in this study) when conducting survey data analysis.

Methodological approach

To estimate bidirectional effects between mathematics self-efficacy beliefs and mathematics achievement, two models were estimated (Model 1 and Model 2) separately by gender. In Model 1, earlier math self-efficacy is a potentially endogenous predictor of later math achievement. In Model 2, earlier math achievement is a potentially endogenous predictor of later math self-efficacy.

To address complications posed by endogenous covariates, we need to model potentially endogenous covariates, along with the outcome equation. This requires using one or more variables ("instruments") that affect the endogenous covariate but can be excluded from the outcome equation. In estimating Models 1 and 2, I used instrumental variables (IV) estimation. Section "Model specification: Outcomes and covariates" discusses the excluded instruments used in each equation and testing of exclusion restrictions.

Measures and model specification

Section "Measures" provides information on the derivation of measures which are central to the investigation, i.e., the self-efficacy composite scales, and the mathematics assessment scores. Section "Model specification: Outcomes and covariates" describes the covariates used in model estimation. It also discusses the excluded instruments used in the IV regressions, and how the relevance and validity of these instruments was tested.

Measures

The self-efficacy composites in HSLS:09, given at two points in time (at grade 9 and 2.5 years later, at grade 11), are composite scales of the sample member's perceived mathematics self-efficacy in a continuum of values from negative to positive, with higher values representing higher self-efficacy. They were generated by HSLS:09 project scientists through weighted principal component factor analysis and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were survey questions on how confident the student was in doing an excellent job on fall 2009 math tests, how certain the student was that he/she can understand math textbook, how certain the student was that he/she can master skills in fall 2009 math course, and how confident the student was on doing an excellent job on fall 2009 math assignments. There were four possible answers to these questions ranging from "Strongly Disagree" to "Strongly Agree." The reported reliability (Cronbach's α) of the math self-efficacy variable was 0.90 in the first wave and 0.89 in the second wave, which indicate a high level of internal consistency of the self-efficacy composite.

The mathematics assessment variables in the HSLS:09, provide a measure of student achievement in algebraic content domains and reasoning, upon entry to high school in fall 2009 and 2.5 years later. The scores used to assess students' performance were based on Item Response Theory (IRT). The IRT model used patterns of correct, incorrect, and omitted responses to obtain ability estimates that are comparable across the low-, moderate-, and high-difficulty test forms (see, Ingels and Dalton, 2013, pp. A9-A11). This IRT estimate was then transformed to derive a final score (standardized *theta* score), the standard unit of the item response theory (IRT) model that represents the level of the domain being measured. The final achievement *theta* scores at grades 9 and 11 provide a summary measure of student mathematics achievement. The *theta* (ability) estimate at each point in time provides a summary measure of achievement. Scores from HSLS:09 first follow-up can be equated to the scale of HSLS:09 base year so that scores may be compared longitudinally. This is the case due to the common items between the HSLS:09 base year and first follow-up. The tests were equated using a procedure[3] which allowed the base-year *thetas* to remain unchanged while the first follow-up *thetas* were equated to the existing base-year scale (U. S. Department of Education, 2018).

Across countries, boys generally (but not always) tend to outperform girls in mathematics tests, although differences are small; on the other hand, girls report lower mathematics self-efficacy than boys (e.g., Louis and Mistele, 2012). **Table 1** presents summary statistics from the HSLS:09 for mean student math achievement, and math self-efficacy by grade and gender. In both the base year and the first follow-up, there were no statistically significant gender differences in mathematics achievement (*theta* ability score). On the other hand, statistically significant and non-trivial differences in perceived mathematics self-efficacy in favor of boys were observed in both years. Therefore, based on the raw data, while boys reported higher mathematics self-efficacy, boys and girls performed equally well.

Model specification: Outcomes and covariates

The mathematics (*theta*) score and the math self-efficacy composite are standardized (mean = 0, SD = 1). Model 1, in which Grade 11 math *theta* score is the outcome, includes the following categories of determinants of achievement: (1) Student demographic characteristics (age; location; minority race; immigrant status); (2) Student's family characteristics (socioeconomic status; mother/father in a STEM occupation) (3) School characteristics (school ownership; school problems as reported by principal, as a proxy of school quality); (4) Teacher characteristics (math teacher's sex; math teacher's experience in teaching math); (5) Student's interest in mathematics composite index; and (6) Student's Grade 9 math-specific self-efficacy composite index, as a potentially endogenous determined covariate. I also control for early math ability using student's earlier math performance (grade 9 math score) as proxy. It is important to properly control for earlier performance; while the survey contains an earlier measure of math performance, student's grade 8 performance in the most advanced math course enrolled, performance is based on the course grade, not on a standardized test.

In the equation endogenously determining math self-efficacy, the excluded instruments are dummy variables which relate to students' general beliefs on which gender is better in mathematics. These are: (1) "boys are better in math" and (2) "girls are better in math," with "girls and boys are equally good in math" as the reference category. Such beliefs correlated with students' assessment of their own efficacy in math. Specifically, believing that males are better in math are positively correlated with math self-efficacy for boys and negatively correlated for girls, while the opposite correlation pattern applied when believing that girls are better in math. Potential determinants of such student beliefs were investigated, by examining their relationship with various student background characteristics, such as parental beliefs on which gender is better in mathematics (based on related survey questions), parents' education, and gender of math teacher. Statistically significant positive correlations were found between parents' and students' beliefs that males are better in math. Similarly, having a male math teacher is positively correlated with both male and female students' belief that males are better in math. Furthermore, higher parental education (especially father's education) is positively correlated with students believing that males are better in math.

Exclusion of these instruments from the outcome equation requires that, while such general beliefs correlate with students' individual self-efficacy perceptions, they are not significant determinants of the outcome. From IV estimation of the Model 1, I assessed the relevance and exclusion restrictions associated with the excluded instruments using tests based on Sargan's J-statistic. In addition, I used the recently developed *kinky least-squares* (KLS) approach by Jan Kiviet,[4] which yields statistical inference on the validity of exclusion restrictions regarding candidate external instrument/s (single, or as a set) when a plausible range of endogeneity correlations is available, whereas these unavoidable restrictions were always supposed to be non-testable (Kiviet, 2020; Kripfganz and Kiviet, 2021).

Model 2 includes an outcome equation with the 11th grade self-efficacy composite as the outcome and the same controls as in the endogenous self-efficacy equation of Model 1. The equation modeling the potentially endogenous earlier mathematics achievement (mathematics *theta* score reported at grade 9), contains the following fixed characteristics as instruments: age of 9-grader (born in 1995 or later, vs. born in 1994); student/school location (town/village vs. larger location). The validity of the instruments requires that birth year and student/school location correlates math achievement but not with math-self-efficacy. I assessed the relevance and exclusion restrictions associated with the excluded instruments as in Model 1.

3 The procedure used is the Stocking-Lord procedure (Stocking and Lord, 1983).

4 The main objective *kinky least-squares* estimation is to provide an instrument free alternative method of identifying coefficients and correcting ordinary least squares (OLS) biases, when a plausible range of postulated endogeneity correlation between a regressor and the error term (degree of endogeneity) is available.

TABLE 1 Self-efficacy and mathematics achievement by gender and grade.

Means	Grade 9			Grade 11		
	Male	Female	Δ (Male-Female)	Male	Female	Δ (Male-Female)
Math *theta* score	0.092 (0.966)	0.087 (0.885)	0.005	0.685 (1.16)	0.657 (1.07)	0.027
Math self-efficacy composite	0.131 (1.00)	-0.054 (0.999)	**0.185**	0.119 (1.00)	-0.073 (0.994)	**0.192**

Standard deviations in parentheses; Bold indicates that differences significant at the 1% level or lower.

Estimation and results

Sections "Model 1: The effect of prior math self-efficacy on later math achievement" and "Model 2: The effect of earlier achievement on later self-efficacy" present the estimation results for the effect of prior math self-efficacy on later math achievement (Section "Model 1: The effect of prior math self-efficacy on later math achievement") and for the effect of earlier achievement on later self-efficacy (Section "Model 2: The effect of earlier achievement on later self-efficacy"). Section "Including teacher variables: comparison of estimates" compares estimates with and without inclusion of teacher variables in the vector of covariates.

Model 1: The effect of prior math self-efficacy on later math achievement

Models were estimated with and without teacher variables (teacher sex and teacher years of experience teaching math) in the vector of exogenous covariates. Teacher variables contain a substantially higher proportion of missing values compared to the student information variables, resulting in about 25% more attrition; furthermore, it is not possible to impute missing values for teacher sex and experience. For example, the sample sizes in estimating Model 1 (6,813 and 7,045 for males and females) would have been 5,229 and 5,417, respectively. I, therefore, present the estimation results from models without teacher variables.[5] A comparison of findings without and with teacher variables is given in Section "Including teacher variables: comparison of estimates."

Table 2 contains the coefficient estimates for Model 1 by gender. The first column contains the Ordinary Least Squares (OLS) estimates (naïve model), with math self-efficacy an exogenous covariate. The second column reports the corresponding IV estimates, with a math self-efficacy the endogenous covariate.

5 The detailed results from models which include the two teacher variables are available upon request.

Findings

In reporting the findings, the effect size metric, Beta (β), is the standardized regression coefficient, expressing the amount of expected change in the standardized outcome variable, associated with one standard deviation change in the predictor variable of interest.

The Ordinary Least Squares (OLS) estimate of the effect of higher earlier self-efficacy on later mathematics achievement is small; for males, it is estimated at nearly 0.1 SD higher Grade 11 math achievement for one SD increase in Grade 9 self-efficacy (β = 0.089, 95% CI: 0.05, 0.13), and slightly smaller for females (β = 0.08; 95% CI: 0.04, 0.11); the difference in point estimates by gender is not significant based on 95% confidence intervals. From these estimates, higher earlier mathematics self-efficacy is associated with higher later achievement, but the effects are small, approaching Cohen (1988) threshold of small effect size in social sciences. This is a general finding in the related empirical literature which is based on methodologies not suitable to derive estimates of causal effects.

Looking at the effect of other covariates with math achievement, earlier math interest does not display any association with late achievement. Significant associations are found for fixed characteristics such as race, socioeconomic status, parents' occupation, school quality, along with earlier math achievement as a proxy of math ability. Black and Hispanic male students performed worse than White students (β = -0.14; p-value = 0.002 for Black and β = -0.1; p-value = 0.03 for Hispanic students), while Asian females performed better (β = 0.26; p-value = 0.000) and to a lesser extent Asian males (β = 0.1; p-value = 0.075). Male students born outside the US outperformed native students (β = 0.26; p-value = 0.000) but interestingly, there is no statistically significant difference by immigrant status among females. The positive effect of better socioeconomic status is small (β = 0.1; p-value = 0.000), while having a father in a STEM occupation is associated with higher achievement for boys (β = 0.12; p-value = 0.016), while having a mother in a STEM occupation is associated with higher achievement for girls (β = 0.11; p-value = 0.07). Finally, there is a strong association between earlier (Grade 9) and later (Grade 11) math achievement for both genders (β = 0.67 p-value = 0.000).

The lower panel of Table 2 contains the first-stage results from the IV regressions. The coefficient estimates

TABLE 2 Effect of earlier mathematics self-efficacy on later mathematics achievement by gender.

	Male		Female	
Outcome: Grade 11 Math achievement (stand.)	**OLS**	**IV**	**OLS**	**IV**
Grade 9 math self-efficacy (stand.)	**0.089** (0.020)	**0.185** (0.087)	**0.080** (0.019)	0.095 (0.093)
Grade 9 math interest (stand.)	0.026 (0.015)	-0.018 (0.070)	0.014 (0.019)	-0.022 (0.080)
Born in 1995 or later	**0.092** (0.026)	**0.091** (0.027)	**0.082** (0.029)	**0.082** (0.029)
Town school	-0.054 (0.042)	-0.052 (0.042)	0.037 (0.038)	0.037 (0.038)
Village school	**-0.060** (0.029)	**-0.059** (0.029)	-0.032 (0.027)	-0.031 (0.027)
Black	**-0.138** (0.043)	**-0.152** (0.052)	**-0.100** (0.048)	-0.105 (0.072)
Hispanic	**-0.100** (0.046)	**-0.104** (0.044)	-0.077 (0.044)	-0.077 (0.045)
Asian	0.105 (0.059)	0.109 (0.059)	**0.264** (0.063)	**0.265** (0.065)
Another race	0.030 (0.180)	0.012 (0.186)	-0.064 (0.111)	-0.063 (0.111)
Born outside the US	**0.263** (0.067)	**0.260** (0.071)	-0.021 (0.084)	-0.021 (0.085)
Socioeconomic status index (stand.)	**0.105** (0.017)	**0.102** (0.017)	**0.098** (0.017)	**0.097** (0.017)
Father in STEM	**0.118** (0.049)	0.103 (0.055)	0.024 (0.049)	0.024 (0.049)
Mother in STEM	0.020 (0.054)	0.029 (0.056)	0.111 (0.061)	0.111 (0.062)
Public school	**-0.081** (0.033)	**-0.087** (0.034)	**-0.092** (0.032)	**-0.092** (0.033)
School problems index (stand.)	**-0.051** (0.014)	**-0.051** (0.014)	**-0.054** (0.016)	**-0.054** (0.016)
Grade 9 Math achievement (stand.)	**0.673** (0.016)	**0.653** (0.032)	**0.670** (0.017)	**0.667** (0.038)
Constant	**-0.092** (0.037)	**-0.091** (0.037)	**-0.112** (0.037)	**-0.110** (0.043)
F-value	251.6	250.7	209.4	207.8
[*p*-value]	[0.000]	[0.000]	[0.000]	[0.000]
First stage:				
Females better in math	–	**-0.167** (0.045)	–	**0.133** (0.043)
Males better in math	–	**0.129** (0.036)	–	**-0.154** (0.043)
F test of excluded instruments:				
F-value	–	44.62	–	38.2
[*p*-value]		[0.000]		[0.000]
Overidentification test				
Sargan's J-statistic	–	0.46	–	0.01
[*p*-value]		[0.498]		[0.920]
Test of endogeneity of Grade 9 math self-efficacy (H_0: Grade 9 math self-efficacy exogenous)				
Wu-Hausman F-statistic	–	1.24	–	0.028
[*p*-value]		[0.265]		[0.868]
N	6,813	6,813	7,045	7,045

Standard errors in parentheses. Bold indicates significance at the 5% level or lower.

of the two binary instruments show that, boys believing that boys (girls) are better in math correlates positively (negatively) with boys' individual self-efficacy perceptions; similarly, girls believing that girls (boys) are better in math correlates positively (negatively) with girls' individual self-efficacy perceptions. Based on the F-values in the first-stage test of excluded instruments (F = 44.6 in the male regression and 38.2 in the female regression), the instrument set is not weak. The overidentifying restrictions test (J-test) provides evidence that the excluded instruments are exogenous with high associated *p*-values. Charts 1, 2 provide additional evidence for the excludability of the instruments, using the *kinky least-squares* (KLS) approach (see Section "Model specification: Outcomes and covariates"). After deriving plausible ranges of endogeneity correlations (based on error correlation coefficients between the outcome and the self-efficacy equations), the charts depict the associated *p*-values for the validity of the exclusion restriction for each instrument alone, as well as the combination of instruments for various values of postulated endogeneity correlations. Finally, tests the endogeneity of math self-efficacy (Wu-Hausman F-test) suggest that the null hypothesis that math self-efficacy is exogenous is clearly accepted for females, while accepted with associated *p*-value = 0.26 for males.

TABLE 3 Effect of earlier mathematics achievement on later mathematics self-efficacy by gender.

	Males		Females	
Outcome: Grade 11 Math self-efficacy (stand.)	**OLS**	**IV**	**OLS**	**IV**
Grade 9 math achievement (stand.)	**0.194** (0.018)	**0.316** (0.122)	**0.207** (0.022)	0.164 (0.176)
Grade 9 math interest (stand.)	**0.099** (0.020)	**0.090** (0.023)	**0.058** (0.023)	**0.059** (0.025)
Black	**0.217** (0.051)	**0.269** (0.075)	**0.248** (0.064)	**0.233** (0.090)
Hispanic	0.041 (0.052)	0.059 (0.056)	0.047 (0.060)	0.042 (0.065)
Asian	-0.061 (0.076)	-0.115 (0.091)	-0.054 (0.084)	-0.029 (0.129)
Another race	-0.006 (0.188)	0.014 (0.202)	0.122 (0.204)	0.118 (0.206)
Born outside the US	-0.008 (0.078)	-0.046 (0.094)	0.094 (0.075)	0.092 (0.076)
Socioeconomic status index (stand.)	0.013 (0.020)	-0.022 (0.040)	-0.014 (0.020)	0.001 (0.054)
Females better in math	-0.052 (0.046)	-0.028 (0.052)	-0.045 (0.048)	-0.048 (0.050)
Males better in math	**0.117** (0.038)	**0.115** (0.040)	-0.081 (0.046)	-0.079 (0.047)
School problems index (stand.)	0.004 (0.016)	0.019 (0.021)	-0.017 (0.020)	-0.022 (0.029)
Grade 9 Math self-efficacy (stand.)	**0.232** (0.022)	**0.204** (0.038)	**0.227** (0.023)	**0.236** (0.045)
Constant	0.032 (0.022)	0.033 (0.022)	**-0.128** (0.023)	**-0.129** (0.024)
F-value	56.2	47.4	36.0	28.0
[*p*-value]	[0.000]	[0.000]	[0.000]	[0.000]
First stage				
Born in 1995 or later	–	**0.183** (0.032)	–	**0.209** (0.035)
Town school	–	**-0.206** (0.050)	–	-
Village school	–	**-0.143** (0.032)	–	**-0.061** (0.032)
F test of excluded instruments:				
F-value		46.8		39.6
[*p*-value]		[0.000]		[0.000]
Overidentification test: H_0				
Sargan's J-statistic	-	1.77	-	0.45
[*p*-value]		[0.412]		[0.500]
Test of endogeneity of Grade 9 math achievement (H_0: Grade 9 math achievement exogenous)				
Wu-Hausman F-statistic		1.67		1.36
[*p*-value]		[0.196]		[0.243]
N	7,031	7,031	7,260	7,260

Standard errors in parentheses. Bold indicates significance at the 5% level or lower.

Turning to the findings from the IV regressions (upper panel of Table 2), the point estimate for the effect of grade 9 math self-efficacy on Grade 11 math achievement for males is twice that from the OLS regression (β = 0.185; *p*-value = 0.034; 95% CI: 0.014, 0.36). However, the corresponding IV-point estimate for females is smaller (β = 0.095; *p*-value = 0.54; 95% CI: -0.21, 0.40), less precisely estimated, and of similar size to the OLS estimate. Due to the higher standards errors associated with IV estimates compared to those from the OLS estimates, the difference in IV estimates by gender is not statistically significant based on 95% confidence intervals. However, if the comparison is between the male IV estimate and the female OLS estimate (given that the null hypothesis of exogeneity of self-efficacy is accepted at high *p*-values for females), a difference in estimates by gender is established. From these findings, there is evidence in support of the *self-enhancement model* for males, while such evidence on females is weak.

Model 2: The effect of earlier achievement on later self-efficacy

Table 3 contains the coefficient estimates for Model 2 by gender. From the OLS regression, the point estimate of Grade 9 math achievement on Grade 11 math self-efficacy (while controlling for Grade 9 math self-efficacy) is modest, but precisely estimated and of similar magnitude for males (β = 0.194; *p*-value = 0.000) and females (β = 0.21; *p*-value = 0.000).

Before adjusting for math performance and other characteristics (i.e., based on summary statistics), Black and Asian students reported higher self-efficacy than White students. After adjusting for earlier math performance and other covariates, the Black-White difference in reported math self-efficacy increases further for both males (β = 0.22; *p*-value = 0.000) and females (β = 0.25; *p*-value = 0.000). On the other hand, after conditioning for covariates there is not statistically significant difference in reported self-efficacy between Asian and White students. In other words, Black students' perceived efficacy in mathematics is higher than what would be consistent with their past math performance, while White and Asian students' perceived efficacy is more in line with past performance. These findings are consistent with those in Bachman et al. (2011) on race/ethnic differences in self-esteem. They used nationally representative data of 8th- 10th- and 12th-grade students in the US, to find that African American students score highest, and Asian Americans score lowest; controlling for grades and college plans, heighten these race/ethnic differences. Furthermore, the findings are highly consistent over time. It is difficult to reconcile these race/ethnic differences in self-esteem using potential explanations based on social comparison processes (see Gray-Little and Hafdahl, 2000), or the reflected appraisals theory (see Crocker and Major, 1989), which would predict that ethnic minorities (such as African Americans) will have lower self-esteem than the ethnic majority. A more promising potential explanation relates to differences in cultural traditions and associated patterns of behavior. For example, Cai et al. (2007) suggest East Asian individuals, while feeling positively about themselves, cultural norms (such as modesty) account for observed differences in self-esteem scores. On the other hand, it has been suggested that African American families strive to instill self-esteem in youth, as a defense mechanism in coping with discrimination (e.g., Hughes et al., 2006; Bachman et al., 2011).

Turning to the findings from IV regressions, from the first-stage results, slightly younger students (born in 1995) are associated with about 0.2 SD higher math achievement, while going to school in smaller communities is associated with about 0.15–0.2 SD lower math achievement. Based on the F-value in the first-stage test excluded instruments, the instrument sets are not weak (F = 46.8 in the male regression and F = 39.6 in the female regression). From the overidentifying restrictions test (J-test), there is evidence that the excluded instruments are exogenous, with associated *p*-values in the 0.4–0.5 range. **Charts 3, 4** provide corroborative evidence for the excludability of the instruments, using the *kinky least-squares* (KLS) approach. Finally, tests the endogeneity of math self-efficacy (Wu-Hausman F-test) suggest that the null hypothesis that math self-efficacy is exogenous is accepted for both males and females; however, the associated *p*-values are relatively low, at about 0.2.

In the male regression, the IV estimate of the effect of higher earlier math achievement on later math self-efficacy (β = 0.32; *p*-value = 0.01; 95% CI: 0.08, 0.55), is larger than the OLS point estimate (β = 0.19; 95% CI: 0.16, 0.23). This IV estimate exceeds the threshold for weak effect size. However, the corresponding IV estimate for females (β = 0.16; *p*-value = 0.35; 95% CI: -0.18, 0.51) is smaller, less precisely estimated compared to the male estimate, and similar in size to the OLS estimate for females (β = 0.21; 95% CI: 0.16, 0.25). Due to higher standard errors of the IV estimates, gender differences in effect estimates can be established only at 80% or lower confidence intervals; however, the findings are at least suggestive of a higher effect size for males.

In conclusion, after considering potential endogeneity biases associated with earlier math achievement, based on the moderate size of the effect of earlier math achievement on later math self-efficacy, the findings provide support for the *skills development* hypothesis for males, but the evidence for females is weak.

Including teacher variables: Comparison of estimates

Since the reported results are based on models without teacher variables, the coefficient estimates of interest were compared between the models without and with inclusion of teacher variables. With the teacher variables included in the vector of exogenous covariates in the model, having a more experienced math teacher is associated with a small and marginally statistically significant positive effect on students' math achievement. Having a male math teacher was negatively associated with girls' math performance but the association was weak, while no association between math performance and teacher's gender was detected for males. Finally, no association between math self-efficacy and the two teacher variables was detected.

The coefficient estimates of earlier math self-efficacy in Model 1 and earlier math performance in Model 2 from the IV regressions without and with teacher variables, at least qualitatively; however, the estimated effects for earlier self-efficacy on later achievement (Model 1) are somewhat smaller for both males and females and less precisely estimated. For males, the estimated effect (β = 0.09; *p*-value = 0.46) is of similar magnitude to the OLS estimate, while the corresponding effect estimate for females (β = -0.03; *p*-value = 0.85) suggests that earlier math self-efficacy is not predictive of later achievement. In Model 2, the estimates from models with and without inclusion of teacher variables allow for the same conclusions; for males, the estimated effect mirrors that from the model without teacher variables (β = 0.33; *p*-value = 0.01) and is statistically significant at the 1% level. For females, the estimated effect is small and statistically insignificant (β = .04; *p*-value = 0.78).

Furthermore, from 95% confidence intervals, the estimated effect for males is statistically larger than the corresponding effect for females. Concluding, based on the findings from the models with teacher variables, there is evidence of reciprocal effects for males, with the dominant effect from earlier math achievement to later math self-efficacy, while reciprocal effects cannot be established for females.

Discussion

When assessing RQ1, findings suggest that in Model 1, weak evidence of endogeneity of earlier self-efficacy was found, but only for males. Comparing OLS and IV estimates using confidence intervals, the IV estimate (β = 0.185) is larger than the OLS estimate (β = 0.089) at conventional levels of significance. For females, on the other hand, exogeneity of earlier self-efficacy is clearly accepted, with both estimates essentially identical, at less than 0.1 SD higher math achievement for one SD increase in earlier math self-efficacy. With respect to RQ2 (is self-efficacy both a cause of and an effect of academic achievement?), reciprocal effects were established for male students, with the dominant effect from earlier achievement to later self-efficacy. For females, evidence in support of reciprocal effects is weak, especially in relation to the *self-enhancement* hypothesis. Finally, when assessing gender differences in effects (RQ3), comparing the OLS confidence interval for females to the IV confidence interval for males allows for the conclusion that the effect of earlier math self-efficacy on later math achievement is higher for males. However, gender differences in the effect of earlier math achievement on later math self-efficacy could not be established.

Before comparing the findings to those from earlier studies in the area of self-efficacy and achievement in math, note that this study differs from earlier studies in that it aims to derive estimates of causal effects by using both: (a) longitudinal data (as in a minority of related studies), from a large high-quality US dataset and multiple measurements at a 2.5-year interval and (b) Instrumental Variables (IV) methodology, which deals with potential endogeneity biases. This approach is better suited to derive causal effect estimates and identify gender differences in effects.

Earlier findings, such as those from the meta-analysis of longitudinal studies by Valentine et al. (2004), point to effect estimates which are consistent with a small favorable influence of positive self-beliefs on academic achievement (at about β = 0.1); this effect size meets or slightly exceeds Cohen's definition of small effect size. These studies used various measures of self-beliefs, i.e., self-concept, self-efficacy, and self-esteem (but generally not math-specific), different measurement delay, number of control variables, matching vs. non-matching domains, and estimation method (such as multivariate regression, path analysis, and structural equation modeling). No significant gender differences were identified in the metanalysis, despite theoretical considerations relating to academic and non-academic concerns suggesting gender as a possible moderator. In this study I found that a stronger effect of earlier self-efficacy on later achievement for males compared to females, while the point IV estimate from the baseline model (without teacher variables) for males is larger than the OLS estimate, far exceeding the benchmark for small effect size.

Comparing findings on presence of reciprocal effects and effect dominance to past evidence, the findings bear some similarity to those by Hwang et al. (2016), who found a reciprocal relationship, with the effect of past academic achievement on later self-efficacy beliefs being stronger than the effect of self-efficacy beliefs on academic achievement. However, no gender differences were established in these studies. I provide evidence of reciprocal effects for male students with the dominant effect from earlier math academic achievement on later math self-efficacy; evidence of reciprocal effects for female students is weaker, with the dominant effect from earlier math academic achievement on later math self-efficacy as well.

The main findings of this study relate to differences by estimation method (generally higher effect estimates from IV compared to using OLS estimation) and related heterogeneity of effects by gender (larger effect estimates for males compared to females). The IV estimator accounts for confounding due to unobserved attributes and measurement error, which are not accounted for by the OLS estimator. One possible contributor to the heterogeneity of effects by gender is misreporting error in perceived math self-efficacy, since self-efficacy may be subject to systematic self-report bias. Specifically, there may be gender differences in misreporting/accurate reporting of math self-efficacy. While the standard case of measurement error pertains to problems with the "measurement tool," here by systematic self-reporting bias I refer to non-random deviations between the self-reported and "true" values of the same measure (e.g., Bauhoff, 2011). The cognitive bias associated with over-estimating one's ability, is known as the Dunning–Kruger effect (Kruger and Dunning, 1999). If "true" self-efficacy values are those consistent with students' past mathematics achievements, reported self-efficacy values can be over/under-estimates of the "true" value. Using the standard non-classical "measurement error" literature,[6] which occurs when the error in reporting the covariate of interest is correlated with the true value of that variable or with the errors in measuring those values (e.g., Bound et al., 1994), biases in any direction may arise (see for example, O'Neil and Sweetman (2012).

6 In the case of "classical" measurement error (error term as mean zero and is uncorrelated with the "true" dependent and the independent variables), the estimated coefficient from a naïve estimator, such as the OLS estimator) will be biased toward the null (attenuation bias).

There are practical implications for the classroom from the findings in this study. Girls report lower mathematics self-efficacy than boys, while actual performance in mathematics does not differ significantly between genders. If the findings on gender differences in effects is associated with boys overestimating their past mathematics performance (positivity bias/positive illusions),[7] while girls underestimate them or rate them more accurately, the result could be girls not pursuing math intensive courses. Increasing girls' positivity bias in mathematics through feedback from teachers on their individual or group good past math performances toward bringing their perceptions into line with past achievements, is a promising intervention. Success of such an intervention requires teachers' feedback, boosting future strengthening of self-efficacy perceptions of girls. This can boost girls' willingness to pursue more advanced mathematics courses.

Limitations

The findings in this study can be generalized in the context of the relationship between mathematics-specific (academic) self-efficacy and mathematics performance in high school. One possible limitation to generalizing the findings, is that in the HSLS survey the mathematics domain on which assessment is based is somewhat narrowly defined (i.e., algebraic content domains and reasoning) so, one might hypothesize that with a different assessment domain, findings might have been somewhat different. Another limitation is that the derived IV estimates come with higher standard errors (as is usually the case); as a result, some of the differences in IV estimates (i.e., male vs. female) are suggestive, since differences cannot be demonstrated based on statistical significance at conventional levels of significance (for example, the difference in IV estimates by gender for the effect of earlier math self-efficacy on later math performance).

Conclusion

This study aimed to first, investigate potential endogeneities in the relationship between students' self-efficacy perceptions and their achievement in the domain of mathematics; second, after accounting for potential biases due to endogeneity, establish bidirectional causal relationships between self-efficacy and achievement; and third, identify differences by gender in effect estimates. Toward these aims, longitudinal data from the HSLS09 and first follow-up surveys on US high school students was used, along with Instruments Variables (IV) estimation. To uncover potential gender differences in effect estimates, models were estimated separately by gender.

The findings can be summarized as follows: (a) evidence for endogeneity of earlier self-efficacy as a predictor of later mathematics achievement and of earlier mathematics achievement as a predictor of later self-efficacy was found, but only for male students; (b) robust reciprocal effects from IV regressions were established only for male students, with the dominant effect from earlier achievement to later self-efficacy; and (c) while earlier correlational studies did not find significant gender differences in effects despite theoretical expectations for their existence, the findings in this study support higher effects for male students. Given that girls report lower mathematics self-efficacy than boys while their performance in mathematics does not differ significantly from that of boys, increasing girls' positivity bias in mathematics through feedback from teachers on their individual/group good past math performances, is a promising intervention.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://nces.ed.gov/surveys/hsls09/hsls09_data.asp.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

7 This is suggested as a possible explanation, rather than having been demonstrated.

References

Animasaun, I. L., and Abegunrin, O. A. (2018). Gender difference, self-efficacy, active learning strategies and academic achievement of undergraduate students in the department of mathematical sciences, federal university of technology, Akure, Nigeria. *Int. J. Teach. Case Stud.* 8, 255–280. doi: 10.1504/IJTCS.2017.08 8929

Bachman, J., O'Malley, P., Freedman-Doan, P., Trzesniewski, K., and Donnellan, M. B. (2011). Adolescent self-esteem: differences by race/ethnicity. gender, and age. *Self-Identity* 10, 445–473. doi: 10.1080/15298861003794538

Bandura, A. (1977). *Social Learning Theory*. Hoboken, NJ: Prentice-Hall.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *Am. Psychol.* 37, 122–147. doi: 10.1037/0003-066X.37.2.122

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Hoboken, NJ: Prentice Hall.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educ. Psychol.* 28, 117–148. doi: 10.1207/s15326985ep2802_3

Bauhoff, S. (2011). Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. *Health Services Outcomes Res. Methodol.* 11, 44–53. doi: 10.1007/s10742-011-0069-3

Bench, S., Lench, H., Liew, J., Miner, K., and Flores, S. (2015). Gender gaps in overestimation of math performance. *Sex Roles* 72, 536–546. doi: 10.1007/s11199-015-0486-9

Bound, J., Brown, C., Duncan, G., and Rodgers, W. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *J. Lab. Econ.* 12, 345–368. doi: 10.1086/298348

Cai, H., Brown, J. D., Deng, C., and Oakes, M. A. (2007). Self-esteem and culture: differences in cognitive self-evaluations or affective self-regard? *Asian J. Soc. Psychol.* 10, 162–170. doi: 10.1111/j.1467-839X.2007.00222.x

Chatard, A., Guimond, S., and Selimbegovic, L. (2007). How good are you in math?" the effect of gender stereotypes on students' recollection of their school marks. *J. Exp. Soc. Psychol.* 43, 1017–1024. doi: 10.1016/j.jesp.2006.10.024

Cimpian, J. R., Kim, T. H., and McDermott, Z. T. (2020). Understanding persistent gender gaps in STEM. *Science* 348, 1317–1319. doi: 10.1126/science.aba7377

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crocker, J., and Major, B. (1989). Social stigma and self-esteem: the self-protective properties of stigma. *Psychol. Rev.* 96, 608–630. doi: 10.1037/0033-295X.96.4.608

Dupeyrat, C., Escribe, C., Huet, N., and Regner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *Int. J. Educ. Res.* 50, 241–250. doi: 10.1016/j.ijer.2011.08.005

Gray-Little, B., and Hafdahl, A. R. (2000). Factors influencing racial comparisons of self-esteem: a quantitative review. *Psychol. Bull.* 126, 26–54. doi: 10.1037/0033-2909.126.1.26

Hannula, M., Bofah, E., Laura Tuohilampi, L., and Metsämuuronen, J. (2014). "A longitudinal analysis of the relationship between mathematics-related affect and achievement in Finland," in *Proceedings of the Joint Meeting of PME 38 and PME-NA 36*, eds S. Oesterle, P. Liljedahl, C. Nicol, and D. Allan (Vancouver, BC: PME).

Hughes, D., Rodriguez, J., Smith, E. P., Johnson, D. J., Stevenson, H. C., and Spicer, P. (2006). Parents' ethnic-racial socialization practices: a review of research and directions for future study. *Dev. Psychol.* 42, 747–770. doi: 10.1037/0012-1649.42.5.747

Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., and Hutchison, B. (2016). The relationship between self-efficacy and academic achievement: a 5-year panel analysis. *Asia-Pacific Educ. Res.* 25, 89–98. doi: 10.1007/s40299-015-0236-3

Ingels, S. J., and Dalton, B. (2013). *High School Longitudinal Study of 2009 (HSLS:09) - First Follow-up: A First Look at Fall 2009 Ninth Graders in 2012.* Washington, DC: U.S. Department of Education. NCES 2014-360.

Kiviet, J. (2020). Testing the impossible: identifying exclusion restrictions. *J. Economet*. 218, 294–316. doi: 10.1016/j.jeconom.2020.04.018

Kripfganz, S., and Kiviet, J. (2021). kinkyreg: instrument-free inference for linear regression models with endogenous regressors. *Stata J.* 21. doi: 10.1177/1536867X211045575

Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121

Louis, R. A., and Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *Int. J. Sci. Mathematics Educ.* 10, 163–1190. doi: 10.1007/s10763-011-9325-9

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: a multiwave, longitudinal panel analysis. *J. Educ. Psychol.* 82, 646–656. doi: 10.1037/0022-0663.82.4.646

Martin, A. J., and Debus, R. L. (1998). Self-reports of mathematics self-concept and educational outcomes: the roles of ego-dimensions and self-consciousness. *Br. J. Educ. Psychol.* 68, 517–535. doi: 10.1111/j.2044-8279.1998.tb01309.x

O'Neil, D., and Sweetman, O. (2012). *The Consequences of Measurement Error when Estimating the Impact of Obesity on Income*. Germany: Institute for the Study of Labour, Bohn. IZA DP No. 7008. doi: 10.1186/2193-8997-2-3

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Rev. Educ. Res.* 66, 543–578. doi: 10.3102/00346543066004543

Pajares, F., and Schunk, D. (2001). "Self-beliefs and school success: self-efficacy, self-concept, and school achievement," in *Perception*, eds R. Riding and S. Rayner (London: Ablex Publishing).

Pajares, F., and Schunk, D. (2002). "The development of academic self-efficacy," in *Development of Achievement Motivation: A Volume in Educational Psychology*, eds Allan Wigfield, S. Jacquelynne, and Eccles (Amsterdam: Elsevier).

Robins, R. W., and Beer, J. S. (2001). Positive illusions about the self: short-term benefits and long-term costs. *J. Pers. Soc. Psychol.* 80, 340–352. doi: 10.1037/0022-3514.80.2.340

Schöber, C., Schütte, K., Köllera, O., McElvany, N., and Gebaue, M. (2018). Reciprocal effects between self-efficacy and achievement in mathematics and reading. *Learn. Individual Differ.* 63, 1–11. doi: 10.1016/j.lindif.2018.01.008

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: a self-efficacy analysis. *J. Educ. Psychol.* 73, 93–105. doi: 10.1037/0022-0663.73.1.93

Schunk, D. H., Pintrich, P. R., and Meece, J. L. (2008). *Motivation in Education: Theory, Research, and Applications (3rd ed.).* Upper Saddle River, NJ: Pearson Education.

Stocking, M. L., and Lord, F. M. (1983). Developing a common metric in item response theory. *Appl. Psychol. Measurement* 7, 202–210. doi: 10.1177/014662168300700208

Taylor, S. E., and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* 103, 193–210. doi: 10.1037/0033-2909.103.2.193

U. S. Department of Education (2018). *High School Longitudinal Study of 2009 (HSLS:09). Base-year to Second Follow-up. Data File Documentation*. Washington, D.C.: National Centre for Education Statistics, Institute for Education Sciences.

Usher, E., and Pajares, F. (2008). Sources of self-efficacy in school: critical review of the literature and future directions. *Rev. Educ. Res.* 78, 751–796. doi: 10.3102/0034654308321456

Valentine, J. C., and DuBois, D. L. (2005). "Effects of on academic achievement and vice versa. separating the chicken from the egg," in *International Advances in Self-research: New Frontiers for Self-research*, eds H. W. Marsh, R. G. Craven, and D. M. McInerney (Greenwich, CT: Information Age).

Valentine, J. C., DuBois, D. L., and Cooper, H. (2004). The relation between self-beliefs and academic achievement: a meta-analytic review. *Educ. Psychol.* 39, 111–133. doi: 10.1207/s15326985ep3902_3

Williams, T., and Williams, K. (2010). Self-efficacy and performance in mathematics: reciprocal determinism in 33 nations. *J. Educ. Psychol.* 102, 453–466. doi: 10.1037/a0017271

Zakariya, Y. F. (2021). Self-Efficacy between previous and current mathematics performance of undergraduate students: an instrumental variable approach to exposing a causal relationship. *Front. Psychol.* 11:556607. doi: 10.3389/fpsyg.2020.556607

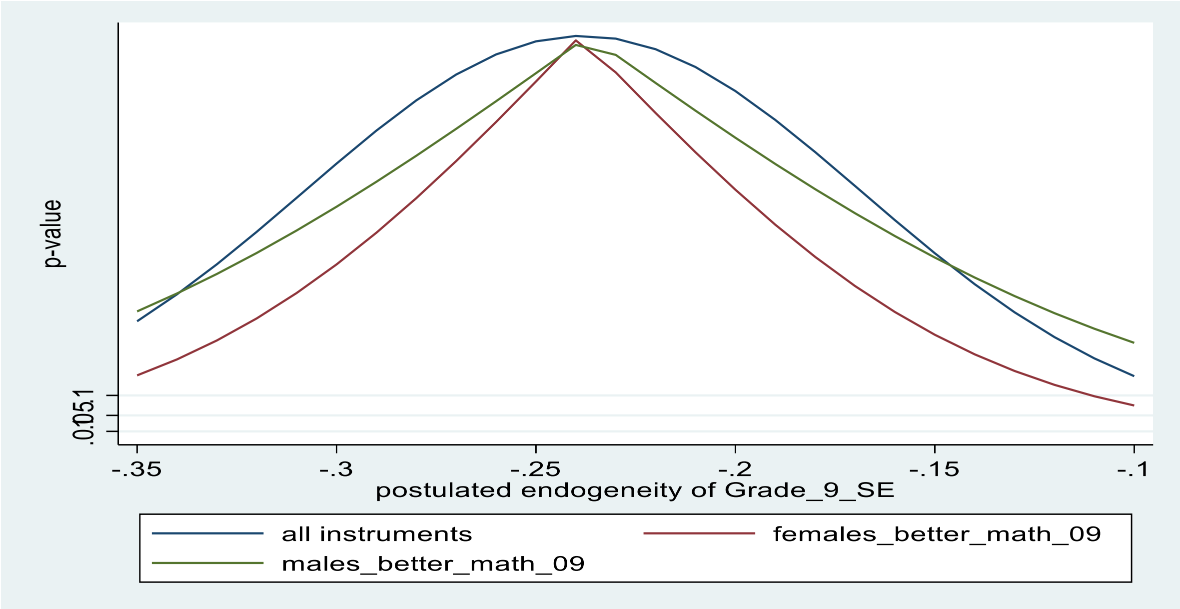


CHART 1
Tests of exclusion restrictions of instruments for Grade 9 math self-efficacy: MALES.

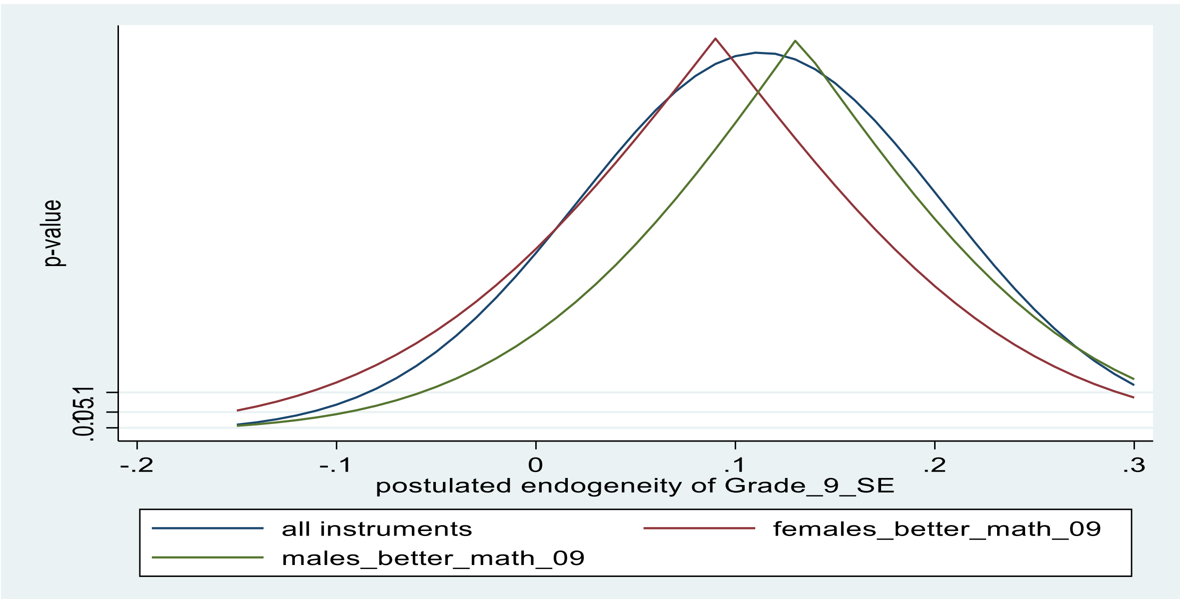


CHART 2
Tests of exclusion restrictions of instruments for Grade 9 math self-efficacy: FEMALES.

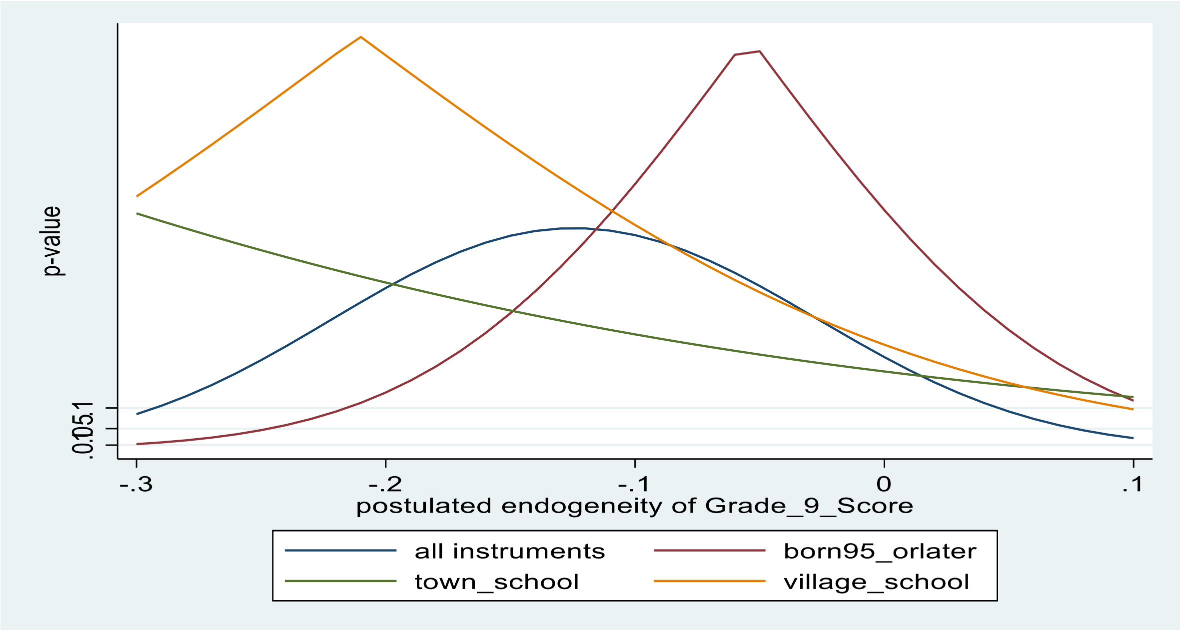


CHART 3
Tests of exclusion restrictions of instruments for Grade 9 math score: MALES.

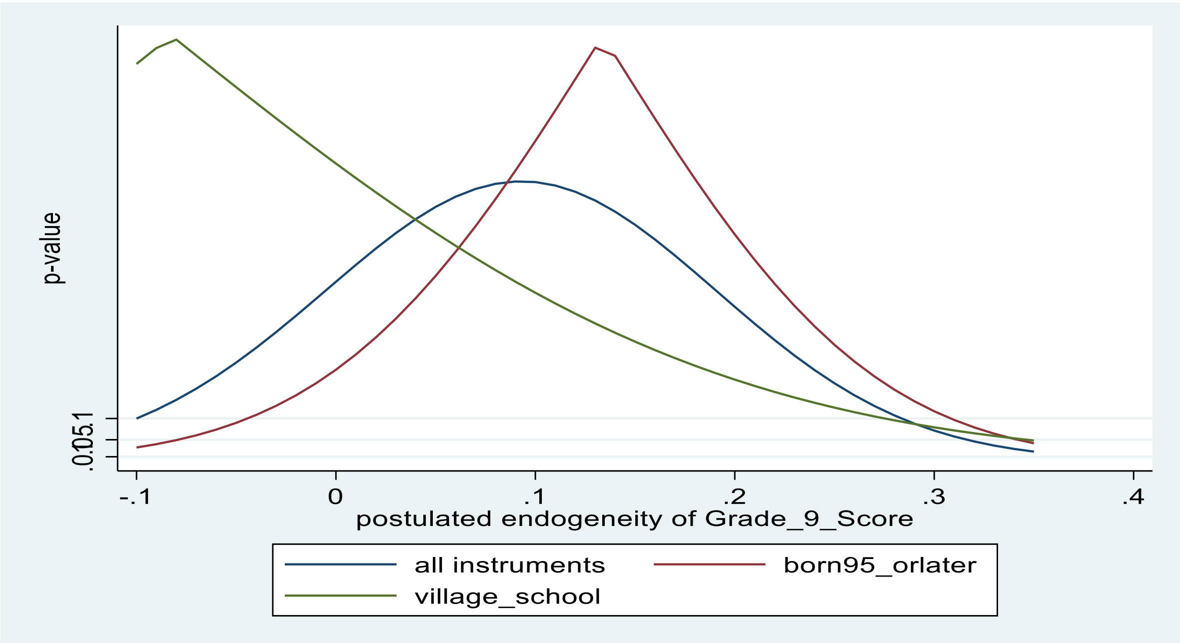


CHART 4
Tests of exclusion restrictions of instruments for Grade 9 math score: FEMALES.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fpsyg.2022.964967

OPEN ACCESS

EDITED BY
Yang Frank Gong,
University of Macau,
China

REVIEWED BY
Qiaoping Zhang,
The Education University of Hong Kong,
China
Sabina Valente,
University of Évora,
Portugal
Xiang Hu,
Renmin University of China, China

*CORRESPONDENCE
Jing Ma
majingmady@bnu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 10 June 2022
ACCEPTED 05 September 2022
PUBLISHED 23 September 2022

CITATION
Zhao W, Ma J and Cao Y (2022) What is effective classroom dialog? A comparative study of classroom dialog in Chinese expert and novice mathematics teachers' classrooms.
Front. Psychol. 13:964967.
doi: 10.3389/fpsyg.2022.964967

What is effective classroom dialog? A comparative study of classroom dialog in Chinese expert and novice mathematics teachers' classrooms

Wenjun Zhao[1], Jing Ma[2]* and Yiming Cao[3]

[1]School of Mathematical Sciences, Sichuan Normal Univeristy, Chengdu, China, [2]Research Center for Mathematics, Beijing Normal University at Zhuhai, Zhuhai, China, Laboratory of Mathematics and Complex Systems (Ministry of Education), School of Mathematical Sciences, Beijing Normal University, Beijing, China, [3]Faculty of Education, International Center for Research in Mathematics Education, Beijing Normal University, Beijing, China

Conducting effective classroom dialog is an important foundation for high-quality classrooms. This study investigates the characteristics of effective classroom dialog from the perspective of Chinese mathematics classrooms. Classroom videotapes from 40 expert and 33 novice teachers were selected from a national project and analyzed through a developed coding framework. Results showed that the dominant types of dialog in expert teachers' classrooms were related to Basic Knowledge, Construction, Analysis, and Personal Information. Compared to novice teachers, expert teachers' classrooms have a significantly lower proportion of dialog on Basic Knowledge and significantly higher proportions of dialog on Personal Information and Speculation. Based on expert teachers' classrooms, the characteristics of effective classroom dialog in the Chinese context were discussed. The analytical framework for classroom dialog developed in this study could be a powerful tool for subsequent research. Suggestions are provided on increasing the effectiveness of classroom dialog.

KEYWORDS

effective classroom dialogue, comparative study, Chinese mathematics classrooms, expert teachers, novice teachers

Introduction

It has been widely documented that effective classroom dialog can promote students' learning outcomes by facilitating their robust learning, deepening their understanding of knowledge, and developing their critical thinking skills (Mercer and Dawes, 2014; Resnick et al., 2015; van der Veen et al., 2017; Howe et al., 2019). Thus, the characteristics of effective classroom dialog and strategies to facilitate effective classroom dialog are of central interest to researchers (Howe and Abedin, 2013; Song et al., 2019).

While studies on classroom dialog have grown rapidly in the past 20 years, most have been conducted in Western contexts (Song et al., 2019). This study argues that viewing the characteristics of effective classroom dialog from the perspective of Chinese mathematics classrooms is necessary and valuable. Learning about teaching and learning from high-achieving education systems has become a trend (Li and Shimizu, 2009), and Chinese students constantly outperform their Western counterparts in mathematics in international comparative studies, such as the TIMSS (Trends in International Mathematics and Science Study) and PISA (Programme for International Student Assessment). It is widely agreed that such outstanding performances are related to the high-quality of instruction in Chinese classrooms (Stigler and Hiebert, 2009; Fan et al., 2015). By focusing on classroom dialog, an important foundation for high-quality instruction, this study investigates the wisdom of Chinese mathematics teaching.

The beginning of the 21st century has seen a wave of international comparative studies of classroom teaching and learning, with TIMSS video study and the LPS (The Learners' Perspective Study) being the main representatives. Many conclusions regarding the characteristics of Chinese mathematics classrooms have been drawn during this time (Leung, 2005; Huang, 2006). However, classroom teaching and learning have greatly changed in the last two decades as China's current mathematics curriculum reform has advanced (Cao and Leung, 2018). There is a need to examine what constitutes an effective classroom in the reform context. In general, China's current mathematics curriculum reform calls for a transition from examination-oriented to quality-oriented instruction to cultivate students as lifelong learners with problem-posing and solving abilities, improved communication and collaboration skills, and greater creativity (Cao and Leung, 2018). In terms of classroom instruction, it aims to shift from a teacher-centered approach (e.g., teacher control, lecturing, rote memorization, and extensive exercises and practices), to a more student-centered approach (e.g., self-regulated learning, exploratory and hands-on activities, group discussion, and project work) (Ministry of Education, 2001, 2011). Another important and recent trend in China's current curriculum reform is the emphasis on cultivating students' high-order thinking abilities (Ministry of Education, 2011, 2022). China's "Core Literacy for 21st Century Student Development" (Lin, 2016) and "5C Model for 21st Century Core Literacy" (Wei et al., 2020) consider high-order thinking abilities—including critical, creative, innovative, and problem-solving thinking—as key to core literacy development. These curriculum reform ideas and initiatives will be considered when constructing the analytic framework and discussing effective classroom dialog.

Furthermore, this study adopts an expert-novice comparison design, often used by cognitive psychologists to study knowledge in specialized domains (Berliner, 2001). Expert teachers are those with high attainments in classroom teaching, student achievement, and research (Goodwyn, 2016). Their classroom teaching practices are models and motivators for other teachers. A comparison with novice teachers can better reveal the unique characteristics of expert teachers' classrooms and provide suggestions for improving novice teachers' classroom practices.

The identification and selection of expert teachers vary from study to study (Brandt, 1986; Leinhardt, 1986), making comparison between studies difficult. There have been growing calls to identify expert teachers using a systematic and rigorous teacher credential mechanism (Berliner, 2001). China's nationwide professional title system for primary and secondary school teachers—which is based on teachers' educational background, teaching performance, student achievement, and teaching and research ability (Gao, 2016)—helps this study identify teachers who enjoy a good reputation in their teaching field and are recognized as experts in education and teaching.

Based on the above background and purpose, this study addresses the following questions:

1. What are the characteristics of Chinese expert mathematics teachers' classroom dialog?
2. What are the similarities and differences between Chinese expert and novice mathematics teachers' classroom dialog characteristics?

The first question explores what constitutes effective classroom dialog in Chinese expert teachers' classrooms. The second question explores the similarities and differences between expert and novice teachers to examine the former's classrooms' characteristics and provide suggestions for improving the latter's competencies for conducting effective classroom dialog.

Literature review

Notions and characteristics of effective classroom dialog

The existing literature includes various definitions for classroom dialog, all highlighting interaction as a key feature (Mercer et al., 2019). This study adopts Howe and Abedin's (2013) definition of classroom dialog, i.e., communication where "one individual addresses another individual or individuals and at least one addressed individual replies" (Howe and Abedin, 2013, p. 326), as it is broad enough to encapsulate other definitions' many commonalities (Song et al., 2019).

Much of the research on classroom dialog builds on Vygotsky (1978) socio-cultural theory, which discusses the relationships between thought, action, communication, and culture (Alexander, 2015). The central view of socio-cultural theory emphasizes that society can be seen as a network of shared activity systems whose interactions are mediated by language, rules, community, and division of labor (Lantolf, 2000). Learning is an activity in which the subject constructs meaning through dialogical interaction in a socio-cultural context. In learning activities, language is the medium, subjective meaning construction is the core, and cultural context is the foundation.

It has been widely documented that effective classroom dialog helps students exchange different ideas, develop critical thinking, and strengthen their understanding of knowledge, resulting in improved learning outcomes (Mercer and Dawes, 2014; van der

Veen et al., 2017; Howe et al., 2019). As such, researchers, policymakers, and educators are interested in learning how to facilitate effective classroom dialog (Zuiker and Anderson, 2021).

Researchers use "scaffolding "as a metaphor for effective classroom dialog and describe how teachers use dialog to build channels that guide students' independent inquiry, develop analytical and problem-solving skills, and promote critical thinking through discussion, questioning, and reflection, resulting in transferability and innovative capabilities (Bakker et al., 2015). Alexander (2008) proposed that dialogic teaching should be collective, reciprocal, supportive, cumulative, and purposeful. Liu (2013) echoed this, claiming that the ideal state of teacher-student dialog is one in which students express their mathematical ideas freely and openly. When teachers listen and respond effectively to their students' ideas, both parties can develop an understanding and create meaning in a discourse community. Howe et al. (2019) identified several productive forms of classroom dialog, including open questions, elaboration and reasoning, coordination across contributions, and metacognition.

Frameworks regarding the analysis of classroom dialog

Classroom dialog can be analyzed using a quantitative or qualitative approach. Each approach has its strengths and weaknesses, which makes them suitable for different purposes. A quantitative method often uses a coding scheme to analyze classroom videotapes/transcripts and look for patterns or relationships. For instance, many existing studies have adopted the FIAS to code classroom activities every 3 s to determine the teaching style based on the frequency distribution and interaction of each code (Martina et al., 2021; Zhao and Boonyaprakob, 2022). Qualitative methods often focus on certain classroom moments and seek the deeper meaning behind the dialog. A qualitative analysis framework should be open and sustainable to account for classroom dialog's social and cognitive nature and probe how language influences thinking and knowledge construction.

With both quantitative and qualitative methods, the unit of coding can vary from a single word to a sentence to several sentences, as long as it is well defined (Chin, 2006; Wells and Arauz, 2006). Howe et al. (2019) highlighted two frequently adopted macro-and micro-level classroom dialog analysis methods: turn-level and lesson-level, respectively. Turn-level analysis refers to the coding conducted at each turn (identified *via* speaker switch). A turn can be coded by more than one code if applicable. Lesson-level analysis can be applied if certain aspects of the lesson, such as the classroom atmosphere, are difficult to describe in a micro manner. In such a case, the classroom dialog can be rated comprehensively and holistically.

Several frameworks are available to analyze classroom dialog (Song et al., 2020), the most well-known being IRF (Sinclair and Coulthard, 1975), FIAS (Flanders, 1970), and CLASS (Pianta et al., 2008). However, these frameworks mainly focus on the form of classroom dialog. For instance, IRF categorizes classroom dialog by initiative, response, feedback, or evaluation, while FIAS lists 10 kinds of teacher or student behaviors. None can reflect the connotation and quality of the dialog. Therefore, some research teams have set out to develop frameworks that focus more on the effectiveness of classroom dialog (Howe et al., 2019). For instance, the Cambridge Educational Dialogue Research Group (see Howe et al., 2019) proposed a coding scheme to represent productive dialog that includes items like elaboration, reasoning, coordination, agreement, querying, reference back, and reference widely. Song et al. (2020) reviewed coding frameworks for classroom dialog over the past two decades, concluding that a dialogic framework should encapsulate six themes: prior knowledge, personal information, analysis, generalization, speculation, and uptakes.

However, these analytical frameworks in the literature have different indicators and standards, making it difficult to compare findings across studies. Given the rapid growth of research in this field, a comprehensive, general, and scientific analytical framework for classroom dialog is greatly needed to make cross-study comparisons more applicable and effective.

Research on classroom dialog in Chinese mathematics classroom

A systematic investigation of Chinese mathematics classrooms was conducted at the beginning of the 21st century through several international comparative studies of classroom teaching and learning, such as TIMSS video study and LPS.

The TIMSS 1999 video study revealed that classroom dialog in the classrooms of Hong Kong SAR is characterized by (1) whole-class interaction; (2) teacher-led lectures; (3) more reasoning and argumentation and more fully-developed expression; (4) a more coherent classroom; and (5) a higher likelihood of student engagement in mathematics learning, etc (Leung, 2005).

Four regions in China (Hong Kong SAR, Macau SAR, Shanghai, and Beijing) participated in the LPS Study, which recorded 10–15 consecutive lessons from at least three representative Grade 8 mathematics teachers in each region. The results showed that in Chinese classrooms, whole-class interaction was the main type of interaction behavior in Chinese mathematics classrooms; the teacher initiated most interactions, and the average volume of teacher discourse was 6.6 times that of students (Cao and He, 2009).

Huang (2006) synthesized the literature on research on teaching and learning in Chinese mathematics classrooms to summarize the characteristics of Chinese mathematics classrooms. Those related to classroom dialog included emphasizing on explanation and illustration, mathematical reasoning, development and construction of knowledge, and procedural problem practice and emphasizing on format, mathematical connections, stimulating questioning and teacher-student interaction, and a lack of realistic contextual connections.

An expert-novice comparison design was often adopted when discussing classroom teaching and learning effectiveness. Generally, expert teachers gave students more opportunities to express their ideas, be more sensitive to classroom tasks, and be more perceptive of social situations while problem-solving (Wang and Ye, 2020). In addition, expert teachers paid greater attention to developing students' mathematical and high-order thinking abilities (Huang and Li, 2012); gave higher-quality explanations that were more accurate and critical (Song et al., 2021a); shared better knowledge regarding eliciting and responding to student thinking; and encouraged students to generate their ideas and work (Gai et al., 2009). In contrast, novice teachers tended toward "teacher-centred interactive dialog," characterized by "teacher's question, student's answer, and teacher's evaluation." (Gai et al., 2009, p.38). Novice teachers often repeated students' answers and offered few opportunities for student participation (Wang and Ye, 2020).

The past two decades of curriculum reform have wrought many changes in classroom teaching and learning. However, few recent studies have investigated classrooms; the few local studies that were conducted mostly featured small sample size and yielded hard-to-generalize findings.

Methodology

Research context and participants

According to the latest standard for China's teachers' professional title promotion system released on August 28, 2015 (Ministry of Education, 2015), primary and secondary school teachers can be assessed as level 3, 2, 1, senior, or exceptional. This study defines novice teachers as those having fewer than 5 years of teaching experience and a level 3 or 2 professional title, and expert teachers as those with a senior or exceptional title.

According to the Ministry of Education (2015), senior and exceptional level teachers are required to (1) work and teach on the front line of education for a long time, act as guides and mentors to promote young students' healthy growth, excel at classroom teaching and counseling tasks, and achieve outstanding teaching and education results; (2) have an in-depth and systematic mastery of the curriculum, professional knowledge of the subjects they teach, outstanding educational and teaching performance, exquisite teaching art, and a unique teaching style; (3) lead and direct educational and teaching research, achieve creative results in educational thinking, curriculum reform, teaching methods, etc., apply them widely in their teaching practice, and play an exemplary and leading role in implementing quality education; and (4) have a bachelor's degree or higher.

Level 2 and 3 teachers are required to (1) be relatively proficient in the principles and methods of educating students and perform the work of classroom teachers and counsellors with good educational results; (2) show mastery of basic pedagogical and psychological theories and knowledge, have the necessary professional knowledge for the subjects they teach, independently master the syllabus and teaching materials of the subjects they teach, correctly impart knowledge and skills, and have good teaching effectiveness; (3) master educational and teaching research methods, actively conduct educational and teaching research, and practice innovation; and (4) either have a master's degree; or have a bachelor's degree, complete a 1-year apprenticeship, and pass the examination; or have a college degree and have taught in primary and junior high school for more than 2 years; or have a secondary teacher training school degree and have taught in elementary school for more than 3 years (Ministry of Education, 2015).

The lessons analyzed in this study were selected from a larger project that has systematically collected more than 500 junior secondary level mathematics lessons from over 10 provinces in Mainland China. Lesson selected for analysis in this study had to: (1) have been videotaped in the past 5 years (2017–2021); (2) be a regular lessons (not a public or competition lessons); and (3) cover as many provinces and districts as possible. Additionally, each school had to select one teacher and one lesson from that teacher; if a teacher had multiple lessons, their second or third was selected to reduce video disruptions. This process yielded lessons from 40 expert (averaging 22 years of teaching experience) and 33 novice teachers (averaging 3 years of teaching experience) in various domains (40 Algebra lessons; 20 Geometry lessons; three Statistics lesson; and 10 inquiry-based lessons). The project received ethical approval from Beijing Normal University in January 2017 and all participants, including principles, teachers, students, and guardians, signed informed consent forms before data collection.

Analytical framework

The analytical framework of this study was adapted from the coding instrument for productive classroom dialog developed by Song et al. (2021b), which contains nine categories: prior knowledge, personal information, analysis, coordination, speculation, construction, agreement, challenge, and instruction /guide.

Song et al. (2021b) framework is based on a systematic review of frameworks for coding toward classroom dialog over the past 30 years and many years of research experience. The framework is suitable for examining the effectiveness of classroom dialog as it not only focuses on dialog form but also reflects its function and quality.

In addition, this study developed sub-categories for the nine categories proposed by Song et al. (2019) to better reflect the characteristics of the mathematics subjects and align with the current curriculum reform in China, using the process described below.

First, the relevant literature was analyzed to look for potential categories. Using word frequency analysis, the research team analyzed recent policy documents regarding the development of future students (e.g., OECD future of education and skills 2030; PISA 2021/2018/2012 mathematics framework; develop students' core competency in China), the latest mathematics curriculum standards from China, the United States, the United Kingdom, Singapore, and Australia, and 180 papers related to classroom dialog in mathematics classes rooms to examine figure student development goals from local and international perspectives and capture the specific characteristics of classroom dialog in mathematics.

Second, we identified 30 high-frequency keywords to provide practical and valuable references for the development of the framework. Through extensive discussion, two experts in mathematics education further sorted these high-frequency keywords into 10 key student development objectives—basic knowledge, problem-solving, logical reasoning, practical exploration, cooperative communication, mathematical expression, transfer application, critical innovation, interest, and literacy—to help develop sub-categories for the study's analytical framework.

Third, referencing the above keywords, the research team developed two to four sub-categories for each category. In addition, the research team invited 20 experts in mathematics education to evaluate the framework and then revised it based on their feedback. After the revision, all 20 experts agreed that the framework had good content validity.

Lastly, three researchers majoring in mathematics education coded 15 lessons separately. The coding results were compared and any differences were resolved through extensive discussion. Descriptions of the coding framework were revised to reduce confusion. Using the final framework, three researchers coded 10 lessons separately, and the inter-rater reliability is 91% (number of dialog turns with consistent codes divided by the total number of dialog turns).

The final framework is presented in Appendix 1.

Analytical process

All the lessons were transcribed verbatim, and the transcripts were then analyzed. Classroom dialog was coded at the turn-level, i.e., a contribution to an exchange made by a participant in a single speaking turn or its constituents. Multiple codes could be applied to each turn of dialog, but multiple instances of a single code within one turn were marked only once. An example of the coding result is presented in Appendix 2.

Three researchers coded the lessons, and the coding results were double-checked. In the coding process, the researchers would constantly refer back to the classroom videotapes to ensure a good understanding of dialog context and improve the coding accuracy. The whole coding process was conducted using the Classroom Teaching Analysis Platform[1] developed by the research team.

In the end, the proportion of each code in expert and novice teachers' lessons was calculated and compared. We chose proportion rather than frequency because generally, dialog turns (the analytical unit) in expert teachers' classrooms were much fewer than those in novice teachers' classrooms. As such, the proportion could better reflect the distribution of different types of classroom dialog. In addition, a *t*-test (two-tail) was conducted (*Via* SPSS) to examine the significance level of the differences between expert and novice teachers.

Results

This study's results are reported in two parts. The first part presents the characteristics of classroom dialog by expert teachers to reveal the characteristics of effective classroom dialog in China. The second part compares and contrasts the characteristics of expert and novice teachers' classroom dialog.

Characteristics of classroom dialog in expert teachers' classrooms

From Figure 1, we can see that the proportion of various classroom dialog types from largest to smallest is Basic Knowledge (27.2%), Construction (20.2%), Analysis (17.6%), Personal Information (16.7%), Coordination (5.7%), Agree and challenge (4.9%), Guide and instruction (4.3%), and Speculation (3.4%).

Basic Knowledge contains mathematical concepts and symbols, relationships and operations, the history of mathematics, general knowledge, and other knowledge. Typically, classroom dialog related to basic knowledge accounts for the largest proportion, as learning basic knowledge is the basic task of mathematics classrooms. From Figure 2, we can see that in addition to a big proportion of dialog on newly learned knowledge (16%), there was also some dialog on prior-known knowledge (5.4%). By reviewing prior-known knowledge, teachers can deepen students' memory and lay the foundation for new knowledge. Leinhardt (1989) similarly reported that expert teachers often "used something familiar to teach something new" (p.66). In addition, expert teachers often repeated students' answers to emphasize relevant knowledge and deepen students' memory (5.8%).

The proportion of the code "Construction" (including the two sub-codes, Probing, and Extending) ranked second. From Figure 2, we can see that most of the dialog related to Construction was Probing (15.6% out of 20.2%), referring to the dialog in which the teacher/student builds on prior utterances to dig deeper. For instance, "How did you come up with this idea?" and "You said that the two triangles are similar; can you explain why?" The results showed that expert teachers often advanced their students' thinking through questioning and always required explanations and justifications. The large proportion of dialog related to Construction shows that expert teachers were adept at advancing their teaching based on students' thoughts/views. Such an observation was also reported by Even et al. (1993), who found that expert teachers could flexibly use students' responses to carry on their teaching rather than follow a fixed procedure.

1 https://159.75.97.242/

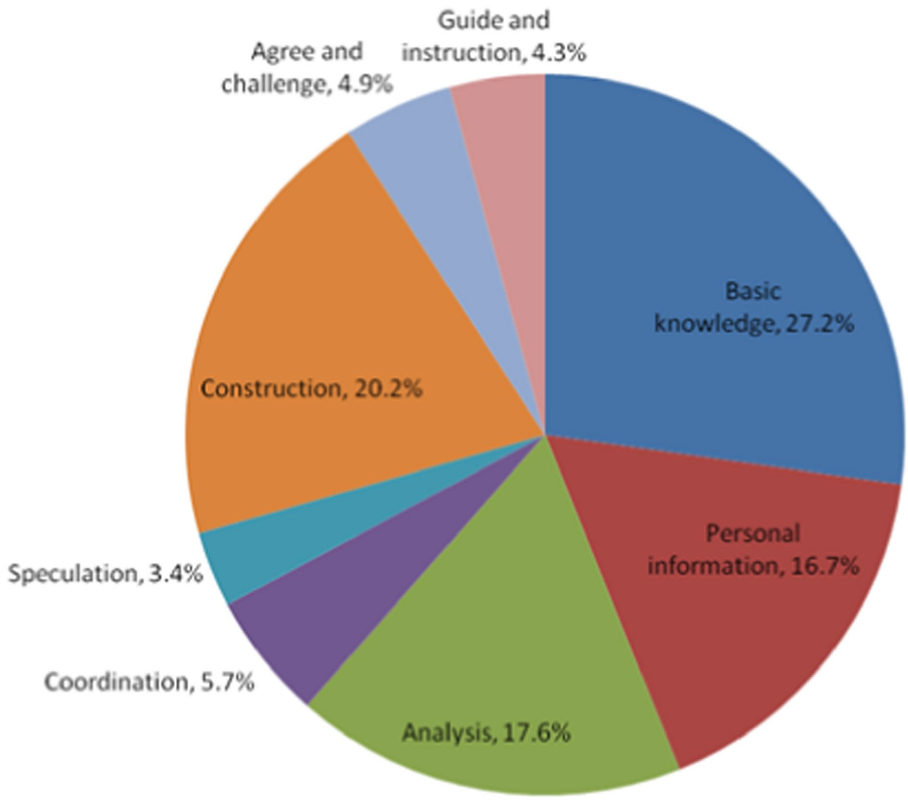


FIGURE 1
Distribution of dialog types in expert teachers' classrooms (first-level codes).

Dialog related to Analysis, e.g., extracting information from tasks, logical reasoning, analyzing and solving problems, explaining, arguing, and evaluating, ranked third. The teacher often guided students, helping them analyze the problem in depth and make a breakthrough to solve it, thereby improving their ability to analyze and solve problems, at the same time, in the process of argument and evaluation, students' logical reasoning ability. By giving examples and explanations, students developed the ability to transform complex and abstract mathematical knowledge into concrete, easily understandable knowledge and elaborate on it in their own words.

Personal Information code and its four sub-codes well reflected the ideas advocated by the current curriculum reform, which calls for connecting learning content to students' daily lives (Personal experience), encouraging students to freely express their ideas (Subjective opinion), encouraging divergent thinking (Imagination), and developing positive attitudes and values (Emotion and value). We can see that dialog related to personal information was considerable and consisted mainly of the sub-codes, Subjective opinion (13.1% out of 16.7%) and Personal experience (2.8% out of 16.7%). The expert teachers often encouraged their students to use their imagination and express their personal opinions, thereby developing the students' imagination, curiosity, and ability to ask questions.

The four other codes—Coordination, Speculation, Agreement and Challenge, and Guide and Instruction—occurred far less often. Several sub-codes reflecting the ideas of curriculum reform, such as Summarization (4%), Prediction and hypothesis (2.4%), Agree (3.3%), and Guide (3.4%), accounted for a relatively higher proportion, indicating that expert teachers consciously implemented curriculum reform ideas in their classrooms. However, some sub-codes important for developing students' higher-order thinking—such as Connection (1.3%), Modeling (0.4%), Migrating application (0.3%), and Challenge (1.5%)—comprised a very low proportion of classroom dialog, suggesting that Chinese teachers (even expert ones) paid insufficient attention to developing students' higher-order thinking in their classroom teaching.

Comparison of classroom dialog between expert and novice teachers' classrooms

Figure 3 and Table 1 show that novice and expert teachers' classrooms shared some characteristics. For instance, both emphasized Basic Knowledge, Construction, and Analysis. Results also showed that Speculation, which signifies high-order thinking, was comparatively rare in expert and novice teachers' classrooms.

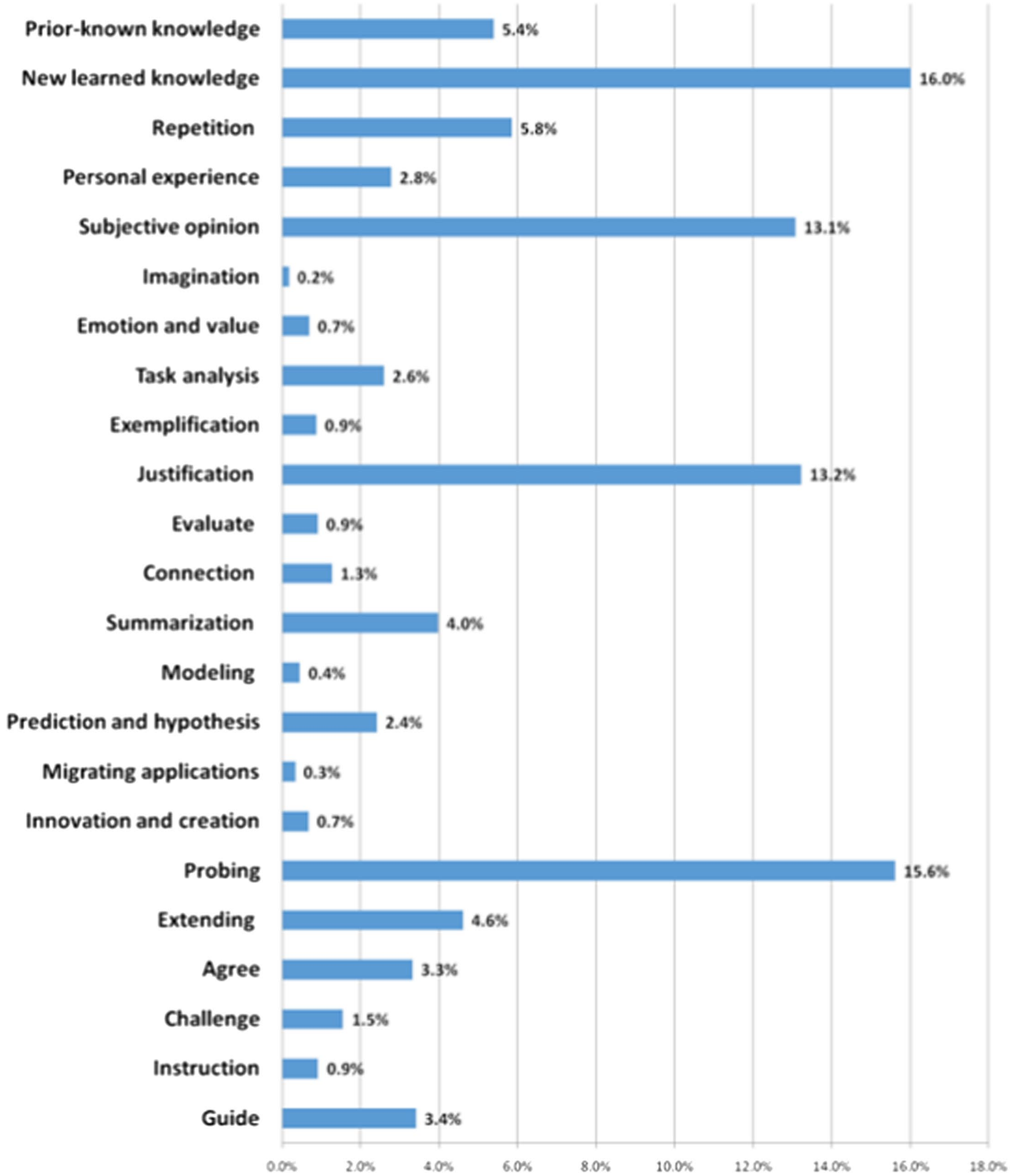


FIGURE 2
Distribution of dialog types in expert teachers' classrooms (second-level codes).

However, compared with novice teachers, expert teachers had (1) a significantly lower proportion of dialog related to Basic Knowledge (27.2% vs. 40.6%, $p=0.000$) and (2) a significantly higher proportion of dialog related to Personal Information (16.7% vs. 4.6%, $p=0.000$) and Speculation (3.4% vs. 0.4%, $p=0.000$).

The differences in dialog related to Basic Knowledge mainly lay in the sub-code "Newly learned knowledge." We can see that 40.6% of classroom dialog in novice teachers' classrooms was about basic knowledge (concepts, theorems, formulas, and operations) that was mostly new to students (26.5% out of 40.6%). As discussed earlier in the above section, expert teachers often introduced new knowledge *via* a fundamental review of prior-known knowledge, while novice teachers tended to teach new knowledge directly.

Regarding the Personal Information code, the differences mainly lay in the Subjective norm code (13.1% vs. 3.6%). Expert teachers' classrooms also featured slightly more dialog related to Personal experience (2.8% vs. 0.8%). A closer examination of the lesson transcripts revealed that expert teachers often started their lessons with a contextual problem closely related to students' personal experiences and asked more open questions, encouraging students to express their thoughts freely. Expert teachers tended to guide students to realize their mistakes, correct their answers,

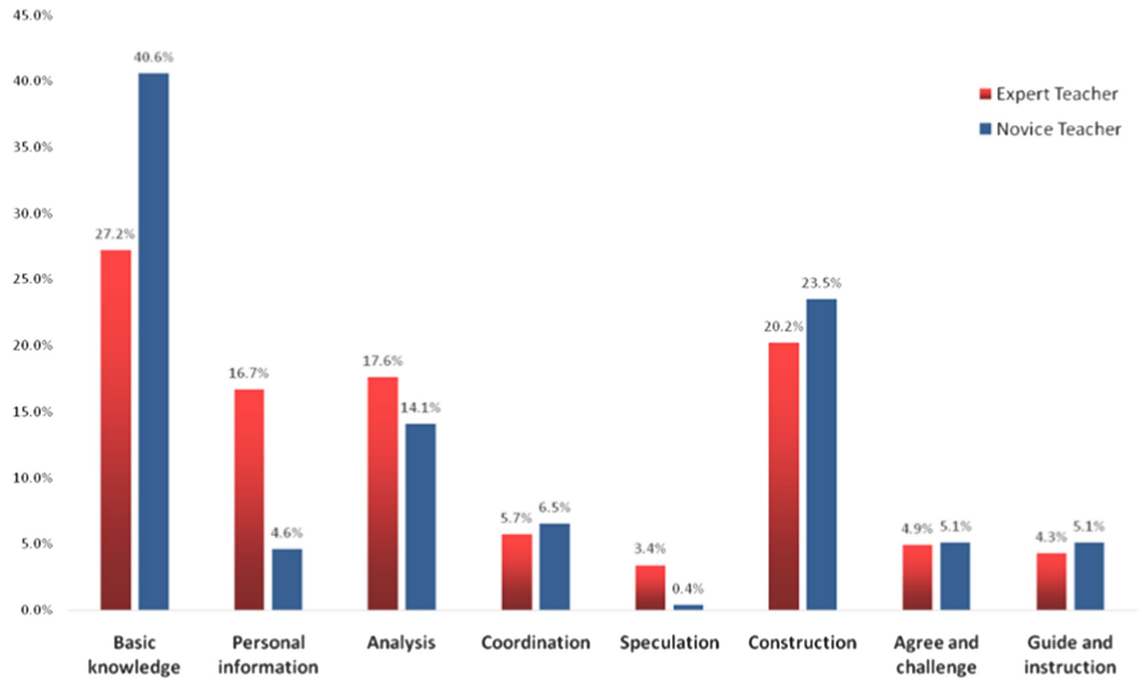


FIGURE 3
Distribution of dialogs types between expert and novice teachers' classrooms.

or invite other students to comment, whereas novice teachers often corrected the answer themselves. Connecting the learning content to students' daily lives and encouraging students to express their ideas are the key ideas of the current curriculum reform. Thus, the results showed that expert teachers' classrooms were generally more aligned with curriculum reform requirements.

Dialog related to Speculation refers to using existing knowledge and information to explore the unknown, inference, problem-solving, make hypotheses, predict the direction of things based on evidence, etc. It is a code signifying a high-order level of thinking. However, even though expert teachers had a significantly higher proportion of dialog on Speculation than novice teachers, both were quite low. It must be noted that the 33 lessons by novice teachers included no dialog on Innovation and Creation, which refers to ideas, thoughts, and opinions that are different from the norm or others. From the videotapes, we observed that novice teachers tended to adhere to the lesson plan rather than deviate based on students' reactions. They provided students fewer opportunities to express their thoughts to ensure the lesson ran smoothly.

Discussion and conclusion

Discussion

This study's results showed that expert and novice teachers paid attention to classroom dialog related to Basic Knowledge, Construction, and Analysis, the bases of Chinese mathematics lessons (Fan et al., 2015). Expert teachers facilitated significantly less dialog related to Basic Knowledge and significantly more dialog related to Personal Information and Speculation.

Based on the expert teachers' classrooms, we can say that characteristics of effective classroom dialog in the Chinese context include ensuring students' learning of basic knowledge, advancing students' responses through probing or extending questions, always requiring an explanation or justification, relating the lesson content to students' daily experiences, encouraging students to share their opinions, and facilitating inquiry-based activities.

Some of the characteristics mentioned above have been documented in the literature. For instance, that effective classroom dialog should be reciprocal and supportive (Alexander, 2008), emphasize open questions, elaboration, reasoning, and making connections (Howe et al., 2019), and make good use of students' information to carry on their teaching rather than follow a fixed procedure (Borko and Livingston, 1989). Questioning and guidance in effective classroom dialog were more effective, strategic, and adept at constructing cognitive networks (Song et al., 2021a).

The findings reflected some unique characteristics of Chinese classrooms, such as the high proportion of dialog related to basic knowledge, justification, and probing. In China, helping students develop a profound foundation of basic knowledge and skills are important instructional goals (Ministry of Education, 2022). Requiring rigorous justification and proof is also an important characteristic of Chinese mathematics classrooms (Huang and Wong, 2007). This study shows that expert and novice teachers frequently used "probing," an important and effective instructional strategy in

TABLE 1 *T*-test regarding the average percentages of each code between expert and novice teachers.

Codes	Sub-codes	Expert teachers (%)	Novice teacher (%)s	*T* statistic	*P* value
Basic knowledge*		27.2	40.6	4.784	0.000
	Prior-known knowledge	5.4	7.8	−1.482	0.143
	Newly learned knowledge*	16.0	26.5	−3.503	0.000
	Repetition	5.8	6.4	−0.545	0.588
Personal Information*		16.7	4.6	−7.116	0.000
	Personal Experience	2.8	0.8	2.699	0.01
	Subjective opinion*	13.1	3.6	7.386	0.000
	Imagination	0.2	0.1	0.387	0.700
	Emotion and value	0.7	0.1	2.939	0.005
Analysis		17.6	14.1	−1.657	0.102
	Task analysis	2.6	1.9	0.904	0.369
	Exemplification	0.9	0.7	0.427	0.671
	Justification	13.2	10.7	1.256	0.231
	Evaluate	0.9	0.7	0.461	0.646
Coordination		5.7	6.5	0.767	0.445
	Connection	1.3	2.3	−2.497	0.015
	Summarization	4.0	4.2	−0.301	0.764
	Modeling	0.4	0.0	2.428	0.02
Speculation*		3.4	0.4	−4.081	0.000
	Prediction and hypothesis*	2.4	0.4	3.517	0.000
	Migrating applications	0.3	0.0	1.841	0.077
	Innovation and creation	0.7	0.0	3.079	0.004
Construction		20.2	23.5	1.321	0.191
	Probing	15.6	19.7	−2.255	0.027
	Extending	4.6	3.8	1.062	0.292
Agree and challenge		4.9	5.1	0.309	0.758
	Agree	3.3	3.6	−0.495	0.622
	Challenge	1.5	1.5	0.220	0.827
Guide and instruction		4.3	5.1	0.763	0.448
	Instruction	0.9	1.8	−1.339	0.185
	Guide	3.4	3.3	0.132	0.895

*Significant difference was found ($p < 0.001$).

China (Fan et al., 2021). The results also showed that expert teachers' classrooms were more aligned with curriculum reform requirements, with considerable dialog related to personal information, prediction, and hypothesis, whereas novice teachers should loosen their control over the classroom and encourage more student participation.

The results also reveal that some codes related to higher-order thinking, like Evaluation, Connection, Migrating applications, Innovation, Creation, and Extending, only accounted for a low proportion of classroom dialog. It is suggested that both expert and novice teachers in China should pay attention to developing students' high-order thinking in classrooms.

Conclusion

This study finds that the dominant types of dialog in expert teachers' classrooms were Basic Knowledge, Construction, Analysis, and Personal Information. Compared to novice teachers, expert teachers' classrooms had a significantly lower proportion of dialog related to Basic Knowledge and significantly higher proportions of dialog related to Personal Information and Speculation. Expert teachers' classrooms to some extent reflected the characteristics of effective classroom dialog in the Chinese context, which emphasize basic knowledge, probing and extending students' responses, explanation, justification, real-life context, personal opinions, and inquiry-based activities. These characteristics not only reflect the traditional characteristics of Chinese classrooms but also the requirements of China's current curriculum reform.

Theoretically, this study broadens our understanding of effective classroom dialog from a Chinese perspective, which may have implications for improving the quality of classroom dialog in other contexts. The framework for analyzing classroom dialog in mathematics classrooms has good reliability and validity and can be adopted in future studies.

Practically, this study can help educators in China and other educational contexts to reflect on and improve their classroom practice. It is suggested that both expert and novice teachers in China should pay attention to cultivating students' higher-order thinking abilities through classroom dialog related to connection, modeling, migrating application and creation, etc. Novice teachers should reduce their control over the class, provide more opportunities for students to express their ideas, and increase students' actively participate in classroom activities. For teachers in other contexts, the Chinese expert teachers' experiences suggested that enhancing students' learning of basic knowledge, probing and extending students' answers through guiding questions, providing more opportunities for students to express their thoughts but always requiring a justification, facilitating inquiry-based activities, as well as flexibly adjust the instruction based on students' reactions, are effective strategies to increase the effectiveness of classroom dialog.

This study has several limitations. First, as its analysis was based on lesson transcripts, this study only examined the frequencies of different types of classroom dialog. Future studies could extend the analysis to include the duration of each code to yield more comprehensive results. Second, this study did not analyze teachers' and students' talk separately. As the current curriculum reforms call for students to play a central role in class, it would be worth examining students' different responses. Third, due to the relatively large number of teachers, only one lesson was selected from each teacher. Last, only quantitative data were reported. A qualitative analysis of classroom transcripts could enrich the results. These limitations will be considered in our subsequent studies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Beijing Normal University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YC, WZ, and JM: conceptualization. WZ and JM: methodology, formal analysis, and writing-original draft preparation. YC and WZ: data collection, writing-review, and editing. YC: supervision and project administration. All authors contributed to the article and approved the submitted version.

Funding

This study was fund by the China National Education Sciences Grant (2022): Evaluating classroom instruction through key classroom activities with the support of Artificial Intelligence (Grant No. CHA220300) and Faculty of Education, Beijing Normal University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.964967/full#supplementary-material

References

Alexander, R. J. (ed.) (2008). *Essays on pedagogy*. London: Routledge.

Alexander, R. J. (2015). Teaching and learning for all? The quality imperative revisited. *Int. J. Educ. Dev.* 40, 250–258. doi: 10.1016/j.ijedudev.2014.11.012

Bakker, A., Smit, J., and Wegerif, R. (2015). Scaffolding and dialogic teaching in mathematics education: introduction and review. *Zentbl. Didakt. Math.* 47, 1047–1065. doi: 10.1007/s11858-015-0738-8

Berliner, D. C. (2001). Learning about and learning from expert teachers. *Int. J. Educ. Res.* 35, 463–482. doi: 10.1016/S0883-0355(02)00004-6

Borko, H., and Livingston, C. (1989). Cognition and improvisation: differences in mathematics instruction by expert and novice teachers. *Am. Educ. Res. J.* 26, 473–498. doi: 10.3102/00028312026004473

Brandt, R. S. (1986). On the expert teacher: a conversation with David Berliner. *Educ. Leadersh.* 44, 4–9.

Cao, Y., and He, C. (2009). Research of the type of teacher-student interaction behavior subject in junior middle school mathematics classroom on the LPS video data. *J. Math. Educ.* 18, 38–41.

Cao, Y., and Leung, F. K.S. (eds.) (2018). *The 21st century mathematics education in China*. Berlin, Heidelberg: Springer, doi: 10.1007/978-3-662-55781-5.

Chin, C. (2006). Classroom interaction in science: teacher questioning and feedback to students' responses. *Int. J. Sci. Educ.* 28, 1315–1346. doi: 10.1080/09500690600621100

Even, R., Tirosh, D., and Robinson, N. (1993). Connectedness in teaching equivalent algebraic expressions: novice versus expert teachers. *Math. Educ. Res. J.* 5, 50–59. doi: 10.1007/BF03217254

Fan, H., Chen, X., and Zhang, J. (2021). Competency-oriented math classroom teaching evaluation based on the rigor/relevance framework ——analysis of national elementary mathematics model lessons (people's education press edition). *J. Math. Educ.* 30, 21–25.

Fan, L., Wong, N. Y., and Cai, J. (eds.) (2015). *How Chinese Teach Mathematics*. Singapore: World Scientific Publishing, 6.

Flanders, N. A. (ed.) (1970). *Analyzing teaching behavior*. New Jersey: Addison-Wesley Educational Publishers.

Gai, L., Zheng, C., and Wang, Y. (2009). A comparative study of the characteristics of chemistry classroom dialogue between expert and novice teachers. *Chem. Educ.* 5, 36–38. (In Chinese).

Gao, H. (2016). The present situation and analysis of primary and secondary school teachers' professional title system reform. *Teach. Educ. Res.* 28, 25–31.

Goodwyn, A. (ed.) (2016). *Expert teachers: An international perspective*. London: Routledge, doi: 10.4324/9780203069578.

Howe, C., and Abedin, M. (2013). Classroom dialogue: a systematic review across four decades of research. *Cambridge J. Educ.* 43, 325–356. doi: 10.1080/0305764X.2013.786024

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., and Wheatley, L. (2019). Teacher-student dialogue during classroom teaching: does it really impact upon student outcomes? *J. Learn. Sci.* 28, 462–512. doi: 10.1080/10508406.2019.1573730

Huang, R. (2006). Looking into Chinese mathematics classroom-an analysis based on the international comparative studies. *J. Math. Educ.* 15, 67–70.

Huang, R., and Li, Y. (2012). What matters most: a comparison of expert and novice teachers' noticing of mathematics classroom events. *Sch. Sci. Math.* 112, 420–432. doi: 10.1111/j.1949-8594.2012.00161.x

Huang, R., and Wong, I. (2007). Looking into mathematics classroom teaching in Hong Kong, Macau and Shanghai. *J. Math. Educ.* 16, 77–81.

Lantolf, J. P. (2000). "Introducing sociocultural theory," in *Sociocultural theory and second language learning*. ed. J. P. Lantolf (Oxford, UK: Oxford University Press), 1–26.

Leinhardt, G. (1986). Expertise in mathematics teaching. *Educ. Leadersh.* 3, 28–33.

Leinhardt, G. (1989). Math lessons: a contrast of novice and expert competence. *J. Res. Math. Educ.* 20, 52–75. doi: 10.5951/jresematheduc.20.1.0052

Leung, F. K. S. (2005). Implications of the third international mathematics and science study for mathematics curriculum reforms in Chinese communities. *J. Math. Educ.* 14, 7–11.

Li, Y., and Shimizu, Y. (2009). Exemplary mathematics instruction and its development in selected education systems in East Asia. *Zentbl. Didakt. Math.* 41, 257–262. doi: 10.1007/s11858-009-0179-3

Lin, C. (ed.) (2016). *Core literacy for 21st century student development*. Beijing: Beijing Normal University Press.

Liu, L. (2013). Construction of a content framework for the analysis of mathematics classroom teacher-student conversation. *Glob. Educ.* 42, 91–99.

Martina, F., Utari, I. R., and Riza, S. (2021). An analysis on teacher talk using Flanders interaction analysis categories (FIAC). *Int. J. Innov. Educ. Res.* 1, 31–52.

Mercer, N., and Dawes, L. (2014). The study of talk between teachers and students, from the 1970s until the 2010s. *Oxf. Rev. Educ.* 40, 430–445. doi: 10.1080/03054985.2014.934087

Mercer, N., Hennessy, S., and Warwick, P. (2019). Dialogue, thinking together and digital technology in the classroom: some educational implications of a continuing line of inquiry. *Int. J. Educ. Res.* 97, 187–199. doi: 10.1016/j.ijer.2017.08.007

Ministry of Education. (2001). *Curriculum standards for mathematics in nine-year compulsory education (experimental version)*. Beijing: Beijing Normal University Press.

Ministry of Education. (2011). *Curriculum standards for mathematics in nine-year compulsory education (2011 version)*. Beijing: Beijing Normal University Press.

Ministry of Education. (2015). *Guidance on deepening the reform of the title system for primary and secondary school teachers*. Ministry of Education of the people's Republic of China. Available at: http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1779/201509/t20150902_205165.html

Ministry of Education. (2022). *Curriculum standards for mathematics in nine-year compulsory education (2022 version)*. Beijing: Beijing Normal University Press.

Pianta, R. C., La Paro, K. M., and Hamre, B. K. (eds.) (2008). *Classroom assessment scoring system: Manual K-3*. Maryland: Paul H. Brookes Publishing.

Resnick, L. B., Asterhan, C. S., and Clarke, S. N. (eds.) (2015). *Socializing intelligence through academic talk and dialogue*. Washington, D.C.: American Educational Research Association, doi: 10.3102/978-0-935302-43-1.

Sinclair, J., and Coulthard, R. M. (eds.) (1975). *Towards an analysis of discourse: the English used by teachers and pupils*. New York: Oxford University Press.

Song, Y., Chen, X., Hao, T., Liu, Z., and Lan, Z. (2019). Exploring two decades of research on classroom dialogue by using bibliometric analysis. *Comput. Educ.* 137, 12–31. doi: 10.1016/j.compedu.2019.04.002

Song, Y., Ding, Y., Zhu, J., and Xu, C. (2021a). Research on interactive classroom teaching of multi-type teacher groups based on sequential pattern mining. *Mod. Educ. Technol.* 31, 40–48.

Song, Y., Hao, T., Liu, Z., and Lan, Z. (2020). "A systematic review of frameworks for coding towards classroom dialogue," in *Emerging Technologies for Education*. eds. E. Popescu, T. Hao, T. C. Hsu, H. Xie, M. Temperini and W. Chen (Cham: Springer), 226–236. doi: 10.1007/978-3-030-38778-5_25

Song, Y., Lei, S., Hao, T., Lan, Z., and Ding, Y. (2021b). Automatic classification of semantic content of classroom dialogue. *J. Educ. Comput.* 59, 496–521. doi: 10.1177/0735633120968554

Stigler, J. W., and Hiebert, J. (eds.) (2009). *The teaching gap: Best ideas from the World's teachers for improving education in the classroom*. New York: Free Press.

van der Veen, C., de Mey, J. R. P. B., van Kruistum, C. J., and van Oers, B. (2017). The effect of productive classroom talk and metacommunication on young children's oral communicative competence and subject matter knowledge: an intervention study in early childhood education. *Learn. Instr.* 48, 14–22. doi: 10.1016/j.learninstruc.2016.06.001

Vygotsky, L. (1978). "Interaction between learning and development," in *Mind in society: The development of higher psychological processes*. eds. M. Cole, V. John-Steiner, S. Scribner and E. Souberman (Cambridge, Mass: Harvard University Press), 79–91.

Wang, S. K., and Ye, L. J. (2020). A comparative study of the teaching language of a novice teacher and an expert teacher in algebra instruction. *Res. Math. Educ.* 23, 13–22. doi: 10.7468/jksmed.2020.23.1.13

Wei, R., Liu, J., Bai, X., Ma, X., Liu, Y., Ma, L., et al. (2020). The research design of the 5C framework for twenty-first century key competences. *J. East China Norm. Univ.* 2, 20–28.

Wells, G., and Arauz, R. M. (2006). Dialogue in the classroom. *J. Learn. Sci.* 15, 379–428. doi: 10.1207/s15327809jls1503_3

Zhao, K., and Boonyaprakob, K. (2022). A case study of Chinese students learning Thai as a foreign language: Flanders interaction analysis category system. *Int. J. TESOL Educ.* 2, 145–169. doi: 10.54855/ijte.222110

Zuiker, S. J., and Anderson, K. T. (2021). Fostering peer dialogic engagement in science classrooms with an educational videogame. *Res. Sci. Educ.* 51, 865–889. doi: 10.1007/s11165-019-9842-z

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 October 2022
DOI 10.3389/fpsyg.2022.992625

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University,
China

REVIEWED BY
Na Li,
Central China Normal University, China
Yuri Sato,
The University of Tokyo, Japan

*CORRESPONDENCE
Hiroaki Ayabe
ayabe@academion.com

SPECIALTY SECTION
This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 12 July 2022
ACCEPTED 12 September 2022
PUBLISHED 03 October 2022

CITATION
Ayabe H, Manalo E and de Vries E (2022)
Problem-appropriate diagram instruction
for improving mathematical word problem
solving.
Front. Psychol. 13:992625.
doi: 10.3389/fpsyg.2022.992625

Problem-appropriate diagram instruction for improving mathematical word problem solving

Hiroaki Ayabe[1,2]*, Emmanuel Manalo[1] and Erica de Vries[3]

[1]Graduate School of Education, Kyoto University, Kyoto, Japan, [2]Department of System Neuroscience, National Institute for Physiological Sciences, Okazaki, Japan, [3]LaRAC, Univ. Grenoble Alpes, Grenoble, France

The use of diagrams can be effective in solving mathematical word problems solving. However, students worldwide do not construct diagrams unprompted or have trouble using them. In the present study, the effects of problem-appropriate diagram use instruction were investigated with an adaptation of the multiple baseline design method. The instruction for using line diagrams, tables, and graphs was provided to 67 junior high school students in a staggered manner and the effects on problem solving of three different types of problems was examined. The results showed that use of problem-appropriate diagrams increased and persisted over time. More importantly, the instruction led to increases in problem solving performance and to decreases in perceived cognitive load. These findings support the argument that effective diagram use depends on the acquisition not only of declarative knowledge, but also sufficient procedural and conditional knowledge.

KEYWORDS

self-constructed diagrams, instructional methods, mathematical word problem solving, cognitive load, representational effect, multiple baseline design, Japanese students, visual representation

Introduction

In mathematics education, teachers draw on mathematical word problem solving to facilitate application of acquired knowledge and skills to real and hypothetical problems and situations (Schoenfeld, 1985; Reed, 1999). Students, however, experience difficulties in solving word problems (Mayer et al., 1992; Reed, 1999; Jitendra et al., 2007; Boonen et al., 2014) since it requires more than simple retention and recall of facts and procedural steps. An effective heuristic to alleviate these difficulties is the use of diagrams (Hembree, 1992; Stylianou and Silver, 2004; Jitendra et al., 2007; Boonen et al., 2014). Diagrams facilitate self-explaining which in turn leads to deeper understanding (Ainsworth and Th Loizou, 2003), promote the construction of mental models for drawing inferences, and provide guidance towards appropriate learning behaviour (Butcher, 2006; van der Meij et al., 2017). More specifically in mathematics, diagrams enable the construction of accurate solutions by enhancing information and knowledge access (Chu et al., 2017; Cooper et al., 2018). However, although these studies contribute to understanding the role of diagrams in learning, they provide only limited insights about constructing effective diagrams for oneself. In many situations, students need to construct their own diagrams for solving word problems in classroom exercises, homework, or tests. Previous research shows several obstacles when students are required to construct their own diagram, instead of just inspect and manipulate a given diagram. Indeed, students may omit constructing a diagram, fail to construct an appropriate diagram, or still draw incorrect inferences from their diagram (Hegarty and Kozhevnikov, 1999; Uesaka and Manalo, 2006; Corter and Zahner, 2007; Uesaka et al., 2007; van Garderen et al., 2012). In the current study, we set out to study instruction as a way of improving students' construction of appropriate diagrams in mathematical word problem solving.

The representational effect or the appropriateness of a diagram for a specific problem

Diagrams enhance understanding of a problem through the representation of its elements and their interrelations (Hembree, 1992). In other words, they facilitate the construction of a schema or mental model of the problem text (Zahner and Corter, 2010). Solving a mathematical word problem involves two steps: generation of a problem representation from the text and implementation or computation of the solution (Kintsch and Greeno, 1985; Lewis and Mayer, 1987; Hegarty et al., 1995; Duval, 2006). The first step can be assimilated to a translation from one type of representation to another (Ainsworth, 2006), also termed conversion from one semiotic register to another (Duval, 2006). In Duval's terminology, natural language, equations, and Cartesian graphs constitute different semiotic registers for representing abstract mathematical objects not directly available to the senses. For example, a problem stated in natural language can be converted into a line graph, but neither text nor graph can be equated with the mathematical object (e.g., a linear function) underlying the problem. Research shows the importance of this first step: when students construct an accurate visual-schematic representation of a problem situation, they are more likely to produce the correct answer (Hegarty and Kozhevnikov, 1999; Boonen et al., 2014).

Diagrams may also facilitate the second step of implementing the solution to the problem. Different isomorphic representations of the same abstract structure or mathematical object differ in their potential for solving a problem, termed "the representational effect" (Zhang and Norman, 1994; Zhang, 1997). Any problem can have multiple alternative forms of external representations (Schnotz and Kürschner, 2008). Thus, different types of diagrams attract attention to different features and may give "representational guidance" (Suthers, 2003). For example, tables attract attention to empty cells and may reveal patterns in a series of quantities in a problem. In mathematics, Duval spoke of "operational significance" (Duval, 2006): a representation in a particular semiotic register is meaningful because of the operations that it affords. For example, tables and graphs do not give a visual-schematic representation of the problem situation, but instead provide a schema for how the problem can be solved (Novick and Hurley, 2001; Zahner and Corter, 2010; Uesaka and Manalo, 2012). In effect, they facilitate what Duval (2006) described as transformations within the same semiotic register. For example, graphs allow visual inspection, which helps in identifying points of intersection of two or more trajectories.

There are a number of studies that have experimentally demonstrated the importance of matching problem requirements with representational affordances of diagrams. Hurley and Novick (2010), for example, asked participants to solve problems using diagrams that did or did not match the problem requirements. Predictably, they found poorer performance (i.e., longer time to solve, inaccurate inferences) in mismatched cases. It is clear therefore that in solving mathematical word problems, not just any diagram will be efficient: the kind of diagram selected and constructed must match the requirements of the problem at hand. We will call this *problem-appropriateness* of a diagram. Ideally, students need to acquire the whole repertoire of diagrams because problem solving in mathematics requires "representational flexibility" (Nistal et al., 2009) or "meta-representational competence" (diSessa, 2004; Verschaffel et al., 2020). In order to achieve meta-representational competence, students first need to acquire knowledge of the different types of representations. Grawemeyer and Cox (2008) demonstrated that such knowledge is crucial when solving "representationally specific tasks" (those that can only be carried out effectively with the use of a very limited range of representations).

Three kinds of representational knowledge are a prerequisite for effective diagram construction and use: declarative (knowing that), procedural (knowing how), and conditional (knowing when; cf. Paris et al., 1983; Garner, 1990). In sum, students need to know that certain kinds of diagrams are helpful for solving certain kinds of problems (declarative knowledge). They need to know how to correctly construct the appropriate diagram based on relevant information in the problem description (procedural knowledge). Finally, they need to know when to use a diagram as well as when to use a specific kind of diagram (conditional knowledge). The question arises whether instruction about representational knowledge of different types of diagrams would increase unprompted diagram use *per se*, and problem-appropriate diagram use in particular. In investigating different types of diagram instruction, the current study addresses this question and thus goes one step further than previous studies on the interplay between different types of diagrams and different types of problems.

Cognitive load associated with constructing diagrams

One reason for the observed difficulties in constructing diagrams may lie in insufficient cognitive resources. From the perspective of cognitive load theory, problem solving tasks can only be successfully undertaken if the required or resulting cognitive load does not exceed the capacity of working memory (Sweller, 1994; Sweller et al., 1998). The effort for visually representing concrete details explicitly described in a word problem is low (e.g., illustrating details). In contrast, the construction of an abstract diagram that does not visually resemble the represented entities, such as a table or a graph, requires more transformational steps. Thus, such a construction is more difficult and demands higher amounts of cognitive effort (Uesaka and Manalo, 2012). Problem-appropriate diagram instruction may reduce cognitive load through schema construction (Sweller et al., 1998; Schnotz and Kürschner, 2007). Schemas cluster elements of a problem and its solution together making them more manageable. Problem-appropriate instruction may draw attention to specific problem features that provide clues for selecting the most appropriate diagram (Duval, 1999, 2006), as well as the relevant declarative, procedural, and conditional knowledge for actually constructing and using that diagram (cf. Paris et al., 1983).

The present study

For developing knowledge about diagrams, appropriate instruction appears to be necessary (van Meter and Garner, 2005; Jitendra et al., 2007; van Garderen, 2007; Uesaka et al., 2010; Manalo et al., 2019). Although instruction appears to promote spontaneity in diagram use, the role of cognitive load and the effect on the correctness in problem solving, particularly where more complex problems are involved, has not been established. Our main purpose therefore was to investigate whether diagram instruction results in increases in *unprompted* diagram construction. Moreover, we expect an increase of *problem-appropriate* diagrams following corresponding diagram-specific instruction. As a result, correctness in solving corresponding word problems should increase and persist over time. Finally, we expect to see corresponding decreases in levels of perceived cognitive load when working on mathematical word problems.

Three diagram-specific instructions for line diagrams, tables, and graphs were designed and tested on three corresponding types of problems "Compare quantities," "Predict patterns," and "Compare trajectories" respectively. These types of problems and diagrams for solving them are a very important part of the Japanese school curriculum (Ayabe et al., 2021). An adaptation of the multiple baseline design method (Baer et al., 1968; Morgan and Morgan, 2009) was used in order to compare the use of different kinds of diagrams and performance on different types of problems across time following different types of instruction. This design involves giving the three types of instruction in a staggered manner and observing the effect on all types of problems for the same participants. For example, an increase in the use of line diagrams specifically and corresponding improvement in problem solving performance should only occur *after* the line diagram instruction and exclusively for the targeted Compare quantities problems, *not* the Predict patterns and Compare trajectories problems. Thus, this design is more appropriate than a "no instruction" control group because it allows comparisons of (1) the same students (within-participant design) and (2) several kinds of instruction. Our specific hypotheses were as follows:

> *H1*: Diagram instruction leads to an overall increase in unprompted use of diagrams.
>
> *H2*: Diagram instruction leads to an increase in the use of problem-appropriate diagrams persisting in time.
>
> *H3*: Diagram instruction increases problem-solving performance (correct answer rates).
>
> *H4*: Diagram instruction reduces perceived cognitive load.

Materials and methods

A faculty ethics committee of Kyoto University approved the study. Participation was voluntary, and prior to the study, participants received verbal and written explanations. Informed consent was obtained from all participants and their parents.

Participants

Seventy junior high school students (aged approximately 14 years, all Japanese) from three regular classes of a junior high school in a small city in Japan participated in the study (ability grouping is not usually practiced in schools in Japan). Students in Japan perform well in mathematics by world standards (Japan ranked 5th in mathematics in PISA 2018; OECD, 2019). We used *G*Power* (Faul et al., 2007) to estimate the minimum sample size for our within-participant design. This estimated that 46 participants would be required to detect a statistically significant difference for the assumed small to medium size effect ($f = 0.25$, α-level $p = 0.05$, power = 0.80). Considering class sizes in the school (≤25 students), and allowing for dropout, three classes were included to ensure minimal sample size. The experimental sessions were conducted during regular class sessions. All the students participated but three missed some sessions and their data were excluded. Data from 67 students (female = 36) were used in the analyses.

Problem-appropriate diagram instruction

Three dedicated instruction sessions covered the use of line, table, and graph diagrams. Instruction and practice sessions were held during regular class sessions (45 min duration). The instructions were given by the first author, assisted by a school mathematics teacher.

To ensure fidelity to plan and equivalence of the three instruction sessions, the authors discussed all contents and the instructional steps were determined in advance. PowerPoint slides were prepared and used to guide instruction. The instruction covered (1) the characteristics and functions of each kind of diagram (declarative knowledge), (2) the types and features of mathematics word problems that each diagram is useful for (conditional knowledge), and (3) the ways of constructing and the reasoning behind each diagram (procedural knowledge). During practice, the students solved example problems and constructed diagrams individually.

Line diagram instruction

Line diagrams, also known as "line numbers" or "tape diagrams" (Murata, 2008), visually express quantities as line segments. Line diagrams allow inferring relationships between sums, differences, multiples, and proportions (declarative knowledge). Constructing a line diagram involves converting quantities to lines to enable easier visual comparisons of the lengths of the lines. Conditional knowledge included that line diagrams are helpful for solving complex problems about relationships between quantities. For developing procedural knowledge, students were asked to construct line diagrams in solving three word problems (isomorphic but different from those used in the tests).

Table instruction

The instructor explained and demonstrated how tables are effective for organizing numbers or quantities of two variables of interest. The students were told that creating an array for one variable and then arranging the second variable in a corresponding array would clarify the relationship between the two variables. Thus, a table makes it easier to find the rule that determines how the two variables change (declarative knowledge). The conditional knowledge conveyed was that tables are helpful for identifying a consistent pattern or rule of change in quantities to predict a future amount. For developing procedural knowledge, the students practiced constructing tables for use in solving three isomorphic word problems.

Graph instruction

The instructor explained that graphs (more specifically, cartesian graphs) are useful for visually representing complex variations or changes of quantities and gave a demonstration on how to represent two variables of a word problem as points with connecting lines on the x-and y-axes. The declarative knowledge included that graphs enable visual awareness of the change in quantities as they increase, decrease, or remain the same across space and time. It also included knowledge about how graphs can be used, such as extending two lines on graphs to find their intersection. The conditional knowledge conveyed was that a graph should be used for complicated processes of change that require projections of future events. Again, for developing procedural knowledge, the students practiced constructing graphs in solving three isomorphic problems.

Mathematical word problems

Five isomorphic problems (same problem structure but with different cover stories) for each of three problem types (Compare quantities, Predict patterns, and Compare trajectories) were used for three types of problem-appropriate diagrams (line, table, and graph diagrams respectively).

Compare quantities problems contained information about the magnitudes of lengths or distances. Solving these problems involved comparing these quantities. Line diagrams are appropriate because constructing a correct visual representation of the lengths not only provides a schematic layout of the problem situation, but also supports identification and working out of missing or unknown lengths (see example problems in Table 1).

Predict patterns problems contained information about quantities at multiple times or stages. Students were not informed of the rule-based character of the changes. Solving these problems required students to infer the rule and predict future quantities. Tables are appropriate because their structure makes patterns of changes visible, which leads to apprehending the underlying rule.

Compare trajectories problems contained information about actions from two or more entities (usually people). Solving these problems required students to compare trajectories of the different entities. Graphs are appropriate because they enable plotting distances (relative to a point of reference) across time, which in turn enables comparing trajectories.

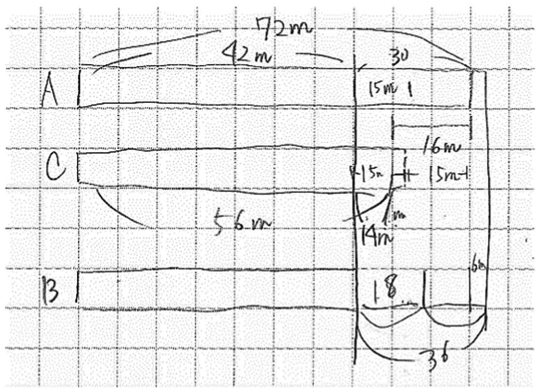
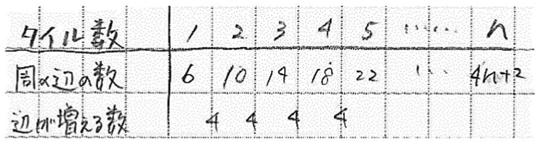
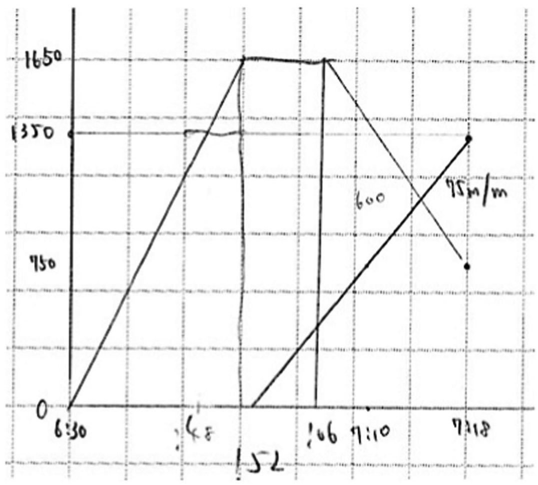
Prior to the study, the 15 problems (5*3) were given to five mathematics teachers (female = 1; mean teaching experience = 9.2 years, *SD* = 2.8 years) to check whether they were comparable and suitable for the intended grade level (14-year-olds at junior high school). Minor adjustments were made based on the teacher feedback. The revised problems were administered to 29 students from another school (female = 12; mean age = 13.2 years). Multiple comparisons using paired t-tests of the correct answer rates revealed no significant differences between the five problems of each type. Thus, they were considered equivalent and randomly used in the five test phases: Pre-test, Post-test after each of the three instruction sessions, and Delayed post-test.

Dependent measures

Unprompted and problem-appropriate diagram use

An analysis grid was constructed for scoring the kind of diagram (Table 2). Numbers, equations, formulas, or computations in columns were not considered as diagrams.

TABLE 1 Example problems (translated from Japanese) and student-constructed problem-appropriate diagrams.

Compare quantities problem There are three counters A, B, and C, at a concert venue for customers with A, B, and C-type tickets. Upon opening, queue length at counter A is 72 m, unknown at counter B, and 56 m at counter C. Fifteen minutes after opening, the queue length at A is 6 m shorter than twice the length at B, and the length of C is 1 m shorter than half the length at A. How much did queue length shorten in the first 15 min? How long was the queue length at B when the gate was opened? (counters become shorter at constant and identical speed).	**Line** 
Predict patterns problem You have to arrange regular hexagonal tiles of 3 cm sides one at a time. Each new tile has to touch only one side of the tiles that are already placed. However, once placed, a tile can have more than one of its sides touching other tiles. When the number of sides (sides not in contact with other sides) around the figure becomes 86, how many tiles will you have arranged? When 26 tiles are placed, what is the length (in cm) around the figure?	**Table** 
Compare trajectories problem Manny leaves home at 6:30 am and walks 1,650 m to school. At school, he discovers that he forgot his lunch box and goes back. His mother discovers the lunch box and decides to bring it to him. At 7:18 am, Manny calls his mother's mobile phone from a convenience store 900 m away from his school. She tells him that she already passed the convenience store at 7:10 am. They meet at the convenience store and Manny gets his lunch box. What time did his mother leave home? How long did Manny stay at school before returning home? (Manny and his mother walk at the same speed. The house, convenience store, and school are on the same route.)	**Graph** 

Two teachers, with no vested interest in the study, rated all 1,005 answer sheets (5*3*67) in random order, blind to both test phase (5 phases) and type of problem (3 types). The teachers first rated 20% of the answer sheets, and compared and discussed their ratings with the first author. The teachers then independently scored the remaining answer sheets. Overall interrater agreement was high (Cohen's kappa = 0.918). Unprompted diagram use was calculated as the presence of any kind of diagram. Problem-appropriate diagram use was calculated as the use of a specific kind of diagram for a specific type of problem (line diagram for Compare quantities, table for Predict patterns, and graph for Compare trajectories problems).

Correctness in problem solving

Correctness in problem solving was scored independently of diagram use. For each question (answer sheet), two answers were required for 0.5 points each. Correctness was scored 1 if both answers were correct, 0.5 if only one of them was correct, and 0 if both were incorrect or answers were missing.

Cognitive load

Cognitive load was measured using a short questionnaire for intrinsic cognitive load (Leppink et al., 2014) translated to Japanese, with some minor adjustments. The questionnaire comprised four items, for example "I invested a very high mental effort in the complexity of this activity," to be answered on a

TABLE 2 Analysis grid for scoring constructed diagrams.

Line	A line diagram consists of line segments or rectangular forms (tapes, bars) representing quantities. Two or more segments should be aligned so that their lengths can be compared. Segments without units or scales such as in geometric shapes or pictures should not be counted as a line diagram.
Table	A table contains at least two arrays of numbers resulting in a matrix of at least two by three (2 × 3) cells. A table need not have a legend, labels, or borders. A 1×2 or 2 × 2 table or an incomplete table (< 6 cells) should not be counted as a table.
Graph	A graph is a Cartesian coordinate system for plotting at least two functions in which a quantity (cost, distance) varies in time. The points and lines do not need to be correct. An empty x-y plane without points or lines is not counted as a graph.
Illustration	Any other graphical or pictorial visual expression or depiction.

TABLE 3 Summary of the multiple baseline design. All five test phases contained a Compare quantities, a Predict patterns, and a Compare trajectories problem.

Session	Day	Instruction	Test Phase
1	1		Pre-test
2	6	Line	Post line test
3	9	Table	Post table test
4	13	Graph	Post graph test
5	22		Delayed test

10-point Likert-type scale (0 = "not at all the case" to 9 = "completely the case"). The reliability of the scale was confirmed on the 15 problems in the preliminary study (Cronbach's alpha ranged from 0.67 to 0.93).

Design and procedure

Following the multiple baseline method, instruction in the use of line diagrams, tables, and graphs was provided in a staggered manner in three sessions, respectively, (see Table 3).

The procedures used in administering the tests were identical across the five phases. Each test contained the three types of word problems in random order. Students were given 8 min to solve each problem. Students filled out the cognitive load questionnaire after solving each problem. All answer sheets were collected at the end of each session. No marks, grades, or feedback on the tests were given in between sessions.

Analyses

Unprompted and problem-appropriate diagram use were dichotomous dependent variables (0 or 1). Therefore, Cochran's Q, a non-parametric test, was used for analysis of main phase effects and McNemar's test was used for pairwise comparisons. Correctness in problem solving had three possible scores (0, 0.5, 1) and perceived cognitive load ranged from 0 to 36. A repeated-measures analysis of variance was run on these variables as it is robust against violations of normal distribution assumptions (Schmider et al., 2010). The Greenhouse–Geisser correction was used when the sphericity assumption was not met. We performed confirmatory analysis with the non-parametric Friedman test.

Results

Did diagram instruction lead to an overall increase in unprompted use of diagrams?

Figure 1 shows diagram use (top row) as a function of problem type and test phase and allows comparing the percentage of answer sheets that included a diagram of any of the four kinds (cumulated shaded parts of the bars) against those that did not include any diagrams at all (white part of the bars). As expected, the unprompted use of any diagram seems to increase as a result of the instructions (white portion decreases over time) but only for the Predict patterns and Compare trajectories problems.

The analysis showed no significant phase effect for Compare quantities problems [Cochran's $Q_{(4)} = 7.26$, $p = 0.12$]. Moreover, no significant difference was found in diagram use between the tests immediately before and after the line instruction [Pre-test versus Post line, McNemar's $\chi^2_{(1)} = 2.88$, $p = 0.90$, p values were multiplied by 10 as a Bonferroni correction]. Thus, the overall level of unprompted diagram use of any kind of diagram was high for Compare quantities problems (> 70%) from the beginning and stayed at such a high level throughout the five phases.

In the Predict patterns problems, a significant phase effect was found (Cochran's $Q_{(4)} = 48.35$, $p < 0.001$). However, no significant difference was found in unprompted use of any diagram between the tests immediately before and after the table instruction [Post line versus Post table McNemar's $\chi^2_{(1)} = 4.84$, $p = 0.28$]. A significant increase was observed only after graph instruction [Post line versus Post graph McNemar's $\chi^2_{(1)} = 22.09$, $p < 0.001$]. Moreover, unprompted use of any diagram for Predict patterns problems did not increase nor decrease from the Post graph to the Delayed test [Post graph versus Delayed McNemar's $\chi^2_{(1)} = 0.11$, $p = 1.00$]. Thus, unprompted use of any diagram increased following table instruction in the corresponding Predict patterns problems, but only towards the end of the procedure.

Finally, a significant phase effect was also found for Compare trajectories problems [$Q_{(4)} = 64.55$, $p < 0.001$]. The significant increase in unprompted use of any diagram followed graph instruction [Post table versus Post graph McNemar's $\chi^2_{(1)} = 13.44$, $p < 0.01$]. It seemed to still increase from the Post graph to the Delayed test, but this was not significant [Post graph versus Delayed McNemar's $\chi^2_{(1)} = 3.00$, $p = 0.83$]. Hence, following graph instruction, unprompted diagram use of any diagram for the

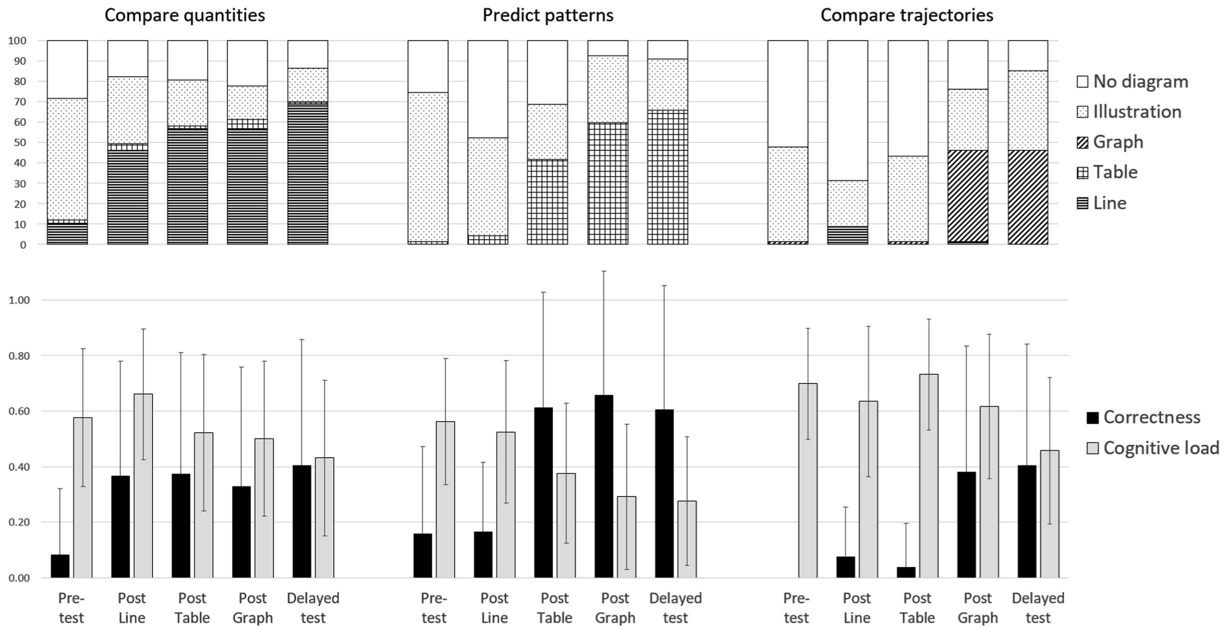


FIGURE 1
Diagram use **(top)**, correctness in problem solving and perceived cognitive load **(bottom)** as a function of problem type and test phase. Cognitive load was normalized to range from 0 to 1.

corresponding Compare trajectories problems significantly increased and sustained.

Did diagram instruction lead to a persisting increase in problem-appropriate diagrams?

We expected an increase of problem-appropriate diagrams specifically. In other words, we expected increases in the use of line diagrams for comparing quantities, tables for predicting patterns, and graphs for comparing trajectories. Such problem-appropriate diagram use should occur directly following the corresponding instruction and persist in time even after alternative diagram instruction. Figure 1 does show this expected pattern of results. The use of each of the three types of diagrams increases after the corresponding instruction but only for the expected type of problem. Following line instruction, although some line diagrams were also used for comparing trajectories, the use of line diagrams increased for comparing quantities [phase effect in line diagram use, Cochran's $Q_{(4)} = 69.30$, $p < 0.001$], but not for the other two types of problems. Following table instruction, the use of tables increased for predicting patterns [phase effect in table use, $Q_{(4)} = 121.39$, $p < 0.001$], not for the other two types of problems. And finally, following the graph instruction, the use of graphs increased for comparing trajectories [phase effect in graph use, $Q_{(4)} = 105.86$, $p < 0.001$], again not for the other two types of problems.

Individual comparisons (again with Bonferroni corrections) confirmed the pattern of results. In the Compare quantities problems, the increase in line diagram use took place directly after the line instruction [between Pre-test and Post line test, McNemar's $\chi^2_{(1)} = 19.20$, $p < 0.001$]. It seemed to still increase in the Delayed test. Indeed, there was a significant difference between the Post line and Delayed test [$\chi^2_{(1)} = 10.71$, $p < 0.01$]. Thus, as Figure 1 shows, appropriate use of line diagrams for Compare quantities problems directly followed line instruction and still increased even after alternative table and graph instructions.

Similar results were obtained for the Predict patterns problems. Following table instruction, table use increased significantly for solving these problems [between Pre-test to Post table test, $\chi^2_{(1)} = 27.00$, $p < 0.001$] and increased still further from Post table to the Delayed test [$\chi^2_{(1)} = 16.00$, $p < 0.001$]. Thus, appropriate use of tables for Predict pattern problems started directly after instruction, and even continued to increase, rather than decrease, after alternative diagram instruction.

Finally, in the Compare trajectories problems, graph use increased significantly just after the graph instruction [between Pre-test to Post graph test, $\chi^2_{(1)} = 29.00$, $p < 0.001$]. However, unlike above, appropriate diagram use did not further increase in the Delayed test.

In all cases, the increase in problem-appropriate diagrams was observed directly after instruction but exclusively for the corresponding type of problem. Moreover, problem-appropriate diagram use tended to intensify, even despite the instruction on alternative diagrams. These results provide full support for the second hypothesis.

Did diagram instruction increase correctness in problem solving?

We expected an increase in correctness for each problem type directly following the corresponding diagram instruction and remaining stable over time. Such an improvement in problem solving can indeed be seen in Figure 1 (black bars in bottom bar graphs). Since the diagram-appropriate instruction took place in a staged fashion, the number of baseline data points differs for the three problem types, Comparing quantities, Predicting patterns, and Comparing trajectories (one, two, and three baseline data points, respectively). We therefore tested the pattern of results with three separate repeated-measurements analysis of variance, one for each problem type.

ANOVA revealed a significant phase effect for Compare quantities problems, $F_{(4, 264)} = 9.24$, $p < 0.001$, $\eta_G^2 = 0.08$. Comparing adjacent phases showed that only the first contrast, between Pre-test and Post line test, reached significance, $t_{(66)} = 5.11$, $p < 0.001$, Cohen's $d = 0.88$, all p values were Bonferroni adjusted. Thus, correctness for Compare quantities problems significantly increased following line diagram instruction and the higher level of performance in problem solving persisted throughout the four subsequent test phases.

We also found a significant test phase effect for Predict patterns problems [$F_{(3.34, 220.12)} = 40.06$, $p < 0.001$, $\eta_G^2 = 0.26$]. For this type of problem, Figure 1 shows that correctness increased directly after the appropriate table instruction. Indeed, in comparing adjacent test phases, the contrast for the comparison between Post line and Post table tests was significant, $t_{(66)} = 8.95$, $p < 0.001$, $d = 1.55$. Thus, correctness for Predict pattern problems augmented after the table instruction and the higher level maintained throughout the subsequent tests.

Finally, the test phase effect was significant for Compare trajectories problems, $F_{(2.39, 157.56)} = 33.03$, $p < 0.001$, $\eta_G^2 = 0.26$. Figure 1 clearly shows improved problem solving directly after the problem-appropriate graph instruction. This expected distinct increase in correctness after graph instruction was significant, $t_{(66)} = 6.31$, $p < 0.001$, $d = 1.09$. Improved correctness for Compare trajectories problems sustained at the obtained higher level in the Delayed test.

Friedman test results provided confirmation of these significant results in the Compare quantities problems [$\chi^2_{(4)} = 34.15$, $p < 0.001$], the Predict patterns problems [$\chi^2_{(4)} = 99.04$, $p < 0.001$], and the Compare trajectories problems [$\chi^2_{(4)} = 90.93$, $p < 0.001$]

Finally, we examined the relation between the use of problem-appropriate diagrams and correctness in problem solving in the Delayed test. Chi-square tests for contingency tables showed that the students who produced an appropriate diagram also obtained higher correctness in problem solving [Compare quantities, $\chi^2_{(2)} = 7.16$, $p < 0.05$; Predict patterns, $\chi^2_{(2)} = 19.30$, $p < 0.001$; Compare trajectories, $\chi^2_{(2)} = 12.83$, $p < 0.01$]. These results show that the use of problem-appropriate diagrams is indeed concurrent with correctness in problem solving, providing full support for the third hypothesis.

Did diagram instruction reduce perceived cognitive load?

Finally, we expected that diagram instruction would decrease perceived cognitive load. Figure 1 shows that while the perceived cognitive load seems slightly decreasing over time (gray bars in bottom bar chart), the relation to diagram instruction is less marked. Again, we ran a separate analysis for each of the three problem types for the same reason given above.

The ANOVA showed a significant phase effect in the Compare quantities problems [$F_{(3.52, 232.01)} = 14.51$, $p < 0.001$, $\eta_G^2 = 0.08$]. Unexpectedly, perceived cognitive load actually increased significantly following line diagram instruction [Pre-test versus Post line test, $t_{(66)} = 2.81$, $p < 0.05$, $d = 0.49$]. Subsequently, a significant decrease took place from the Post line to the Post table test, $t_{(66)} = 4.14$, $p < 0.001$, $d = 0.72$. Perceived cognitive load was lowest at Delayed test (significantly lower than at Pre-test, $t_{(66)} = 4.13$, $p < 0.001$, $d = 0.71$). In other words, line diagram instruction did not immediately lead to cognitive load reduction in solving the Compare quantities problems, but a delayed reduction could be observed.

A significant phase effect was also found for the Predict patterns problems, $F_{(3.58, 236.09)} = 35.78$, $p < 0.001$, $\eta_G^2 = 0.19$. In the Predict patterns problems, the pattern of perceived cognitive load variations fully supported the fourth hypothesis. No change in reported cognitive load was found prior to table instruction [Pre-test versus Post line test, $t_{(66)} = 1.21$, $p = 0.46$ (ns), $d = 0.21$], but a significant decrease followed table instruction [$t_{(66)} = 4.37$, $p < 0.001$, $d = 0.76$, as well as a further decrease observed in the next Post graph test, $t_{(66)} = 2.95$, $p < 0.05$, $d = 0.51$. Cognitive load did not further decline in the Delayed test, $t_{(66)} = 0.64$, $p = 0.52$ (ns), $d = 0.11$]. Thus, evidence was found that table instruction reduced perceived cognitive load in solving the corresponding Predict patterns problems.

Finally, the analysis of perceived cognitive load showed a main effect of test phase for the Compare trajectories problems, $F_{(3.41, 224.80)} = 22.77$, $p < 0.001$, $\eta_G^2 = 0.14$. The contrasts showed that, while there was no change in perceived cognitive load between Pre-test and Post line test, there was an unexpected increase at Post table test [i.e., Post line test versus Post table test, $t_{(66)} = 3.70$, $p < 0.01$, $d = 0.64$]. Following graph instruction, the reported load then significantly decreased [i.e., Post table test versus Post graph test, $t_{(66)} = 3.38$, $p < 0.01$, $d = 0.58$]. A further decrease in perceived cognitive load was found at the Delayed test [Post graph test versus Delayed test, $t_{(66)} = 5.50$, $p < 0.001$, $d = 0.95$]. It is possible to interpret the decline in cognitive load from Post table test through to Delayed test as possibly stemming from practice effects. However, given that no decrease in perceived cognitive load actually occurred until *after graph instruction was provided,* we believe that on the whole these results can be taken as supporting the fourth hypothesis.

Discussion

The results of the present study provide support for the hypotheses that we tested. Diagram instruction increased unprompted use of diagrams and, more importantly, it increased the use of problem-appropriate diagrams. These increases in use persisted in time. Furthermore, the instruction led to increases in student problem solving performance and to decreases in their perception of cognitive load associated with that problem solving. In this section, we consider the reasons for and meaning of these results, and discuss their theoretical, research, and practical implication.

Promoting unprompted and problem-appropriate diagram use

Two previously identified key challenges are that students generally lack spontaneity in diagram use and that, even when they construct diagrams, these are often not appropriate for the problem (Hegarty and Kozhevnikov, 1999; Uesaka and Manalo, 2006; Corter and Zahner, 2007; Uesaka et al., 2007; van Garderen et al., 2012). The findings of the present study demonstrate that with instruction focusing on the correspondence between different types of problems and different kinds of diagrams both of these challenges can be resolved.

Previous research revealed that deficiencies in declarative knowledge is one important reason why students do not use diagrams when they should. Previous research also showed that instruction promotes greater spontaneity in the use of diagrams in mathematical word problem solving (Uesaka et al., 2010). This was confirmed in the present study: there were significant increases in unprompted use of any diagrams in both the Predict patterns and Compare trajectories problems following diagram instruction. In the Compare quantities problems, increases in the unprompted use of any diagrams also followed instruction, but these were not significant. The most likely reason was that the level of diagram use in attempts at solving the Compare quantities problems was already high at the first baseline (Pre-test), and so the increases that followed were proportionally small. Note that in all three problem types, most of the diagrams that participants constructed prior to instruction were illustrations (see Figure 1), which would not have been helpful toward obtaining the correct solutions.

Instruction specifically should promote the construction of effective diagrams. Thus, in the present case, the goal of instruction was not for students to construct any diagram because not all diagrams are equal in helping toward generating the required answers. Different representations, even when they are isomorphic, vary in their potential for solving a problem (Zhang and Norman, 1994; Zhang, 1997; Duval, 2006; Schnotz and Kürschner, 2008). Therefore, a crucial purpose of instruction is to enable students to determine and construct the most appropriate diagram to match the requirements of a problem. In the present study, for all three problem types, significant increases in problem-appropriate diagrams were evidenced following instruction, and those increases maintained. In the test phases following each instruction session, all three problem types were administered (in a random order) but, in each case, a significant increase was observed only in the problem type corresponding to the kind of diagram for which instruction had just been provided. This result suggests that, when given instruction, students are able to distinguish pertinent features of a problem and consequently select the most appropriate kind of diagram for solving it. They are able to develop both the necessary conditional and procedural knowledge.

Reducing cognitive load and improving word problem solving

A third important point is that, if students construct a problem-appropriate diagram, it should lead to a better problem solving performance. Again, this was demonstrated in the present study: the increases in appropriate diagram use coincided with significant improvements in problem solving performance. This outcome is understandable when we consider the representational effect mentioned earlier (Zhang and Norman, 1994; Zhang, 1997) and the specific operations that representations can enable in mathematical problem solving (Duval, 2006). More specifically, while some diagrams give an accurate visual-schematic representation for understanding a problem (Hegarty and Kozhevnikov, 1999; Boonen et al., 2014), they may not help in actually solving it. Problem-appropriate diagrams, especially for more complex mathematical word problems, are not just visual or topographical representations. They are of a more abstract nature that enables drawing inferences or executing necessary operations. The execution of such operations is quite specific and systematic, requiring the connections between pertinent details in the problem text, the choice and construction of the diagram, and the derivation of the solution, to be explicitly explained – and practiced – in instruction sessions provided.

This brings up a fourth important point: that problem-appropriate instruction likely reduces the cognitive load experienced during problem solving, thereby facilitating the unprompted and appropriate use of diagrams, as well as freeing up cognitive resources that can be used in working out the answers. Evidence suggesting this was obtained in the current study: in all three problem types, instruction led to immediate or subsequent reductions in reported cognitive load, which coincided with increases in both appropriate diagram use and correct answer rates. According to cognitive load theory, the acquisition of knowledge and understanding relevant to a task leads to schema construction, which in turn leads to a reduction in intrinsic cognitive load and to freeing up of resources in working memory (Sweller et al., 1998). In the case of problem solving and diagram use, prior to instruction the experience of cognitive load would likely be high, especially if the student is unsure about what to do. However, when problem-appropriate instruction is provided, the student would learn what to do and possess a schema to use for solving the problem. This means that the student's experience of cognitive load would likely decrease (Fuchs et al., 2020). Such decreases could have arisen because of practice effects (Wesnes and Pincock, 2002), so it would be useful in future studies to obtain direct measurements of cognitive load (e.g., brain activity). In the present study, there is also evidence from the multiple baseline design that no significant decreases in cognitive load occurred prior to instruction, even in the Compare trajectories problems with three baseline points.

Theoretical implications

The findings of this research provide useful insights about the use of self-constructed diagrams in problem solving. They emphasize the importance of paying sufficient attention to the cultivation of procedural and conditional knowledge. In most Japanese mathematics classrooms, for example, teachers only demonstrate the use of diagrams, without any explicit explanation of how to select, construct, and use them (Uesaka et al., 2007). Despite being familiar with the types of word problems and diagrams, students did not spontaneously use diagrams and failed to solve the problems at baseline. Thus, without proper explanations students have gaps in their procedural and conditional knowledge for diagram use, which not only explain the lack of spontaneous use, but also of inappropriate use and inability to draw the necessary inferences (Hegarty and Kozhevnikov, 1999; Uesaka and Manalo, 2006; Corter and Zahner, 2007; Uesaka et al., 2007). The cultivation of the necessary (and presumably incomplete) procedural knowledge and conditional knowledge was addressed in this study through explicit instruction in problem-appropriate diagram use – which proved effective in improving problem solving behaviours.

The findings also indicate that an important consequence of such instruction is the reduction of cognitive load, specific to the problem type dealt with in the instruction. Our results suggest that cognitive load reduction is instrumental not only in promoting spontaneity in diagram use, but also in allowing sufficient cognitive resources to bear on the problem and hence to solve it successfully.

Furthermore, the findings draw attention to the distinction between two important functions that diagrams can serve in mathematical word problem solving: providing an accurate visual-schematic representation for understanding the problem and providing a schema or operational tool for solving it. Most students are aware of the first of these functions, which is why even prior to instruction many of the participants in the present study produced illustrations. However, such illustrations even when they portray an accurate schema of the problem situation, may not help in working out the solution to the problem. More complex problems often require the use of more abstract diagrams (tables, graphs) that do not visually portray the problem situation but instead directly facilitate obtaining the required solutions. In the research area of diagram use in mathematical word problem solving, little work has been undertaken on this second function (Verschaffel et al., 2020). We believe it deserves more attention as, among other things, the transformational steps involved in their construction need to be better understood.

Research implications

In this research, the multiple baseline design allowed within-participant comparisons without requiring a control group. Multiple testing phases showed increases in performance only for the expected types of problems directly following the corresponding instruction. This design is more commonly used for evaluating individual behavioural change in response to an intervention, particularly when there is an expectation that the change would be irreversible (Baer et al., 1968; Morgan and Morgan, 2009). Apart from across individuals (participants or clients), variations of the multiple baseline design include across settings, behaviours (Morgan and Morgan, 2009), and populations (Hawkins et al., 2007). The design has previously been used to evaluate the effect of providing instruction on mathematics skills to students. However, usually, instruction in a single mathematical operations using a particular teaching approach is evaluated across individual students (Rivera and Smith, 1988). In the present study, we used the design to evaluate the effect of instruction on multiple aspects of participant responding (behaviour, performance, perception) across variations in types of problems, with the aim of demonstrating the need for problem-appropriate diagram instruction.

Like in previous studies, we expected resulting changes to be irreversible, and thus to maintain in post-instruction test phases. But we also expected the effects to be problem type-specific, with limited or no transfer across the problem types. Our results confirmed these expectations. In fact, the multiple baseline design has proven crucial in demonstrating not only the problem type-specific effects of the instructions, but also the co-occurrence of pertinent changes in behaviour, performance, and perception (increases in appropriate diagram use and correct answer rates, along with decreases in cognitive load). Therefore, from a research design perspective, we have been able to demonstrate a useful variation of the multiple baseline design that may have potential further applications in classroom educational research.

Practical implications

The results of the present study indicate that teachers need to explicitly provide instruction on diagram use if their students are to use them effectively in mathematical word problem solving. Many students will not likely construct a diagram if they lack adequate knowledge and skills: it may seem too demanding, and any effort in constructing a diagram may not pay off. Necessary problem-appropriate diagram instruction largely depends on teachers possessing the corresponding knowledge and skills. However, some teachers may be proficient in using diagrams in mathematical word problem solving, but may not have considered how to articulate such knowledge to convey it effectively to their students. It is therefore important to incorporate training in this area both for pre-service teachers in mathematics education, as well as for in-service teachers who may need upskilling through professional development courses.

Limitations and directions for future research

In the present study, we tested our hypotheses on the use of only three kinds of diagrams to solve three kinds of mathematical word problems. This is an important limitation to note as there are other kinds of diagrams that can be used to solve other types of problems, and it would be imperative to examine those in future research. Furthermore, our student participants all came from the same grade level in one school. We acknowledge that student capabilities in both mathematical word problem solving and diagram use would vary according to their age and grade level, as well as other aspects of their educational experiences. Thus it would also be useful to evaluate the effectiveness of problem-appropriate instruction on diagram use on students at other grade levels and from different educational backgrounds.

The instructions were also provided by the first author and an assisting teacher rather than the students' real classroom teachers. An important step to take in future research would be to develop and evaluate instruction that real classroom teachers could use in cultivating diagram use capabilities in their own students.

Conclusion

The results of this study indicate that instruction on diagram use enables the construction and use of appropriate diagrams, improves ability to correctly solve problems, and reduces perception of the cognitive load associated with mathematical word problem solving. The instruction needs to be problem-appropriate, meaning that students need to learn specific details about the construction and use of different kinds of diagrams relevant to solving specific types of problems. As mathematical word problem solving is one crucial means by which understanding of the relevance of mathematics in the real world is cultivated, and diagram use is arguably one of the most effective heuristics for solving them, the effect of instruction indicated by our findings warrants serious consideration – especially as the extent to which such instruction is currently provided in most classrooms may be too general and thus inadequate.

Data availability statement

The datasets presented in this article are not readily available because the datasets generated during and/or analyzed during the current study are not publicly available because we did not obtain consent for secondary use from the participants but have a possibility to be available from the corresponding author on reasonable request. Requests to access the datasets should be directed to ayabe@academion.com.

Ethics statement

The studies involving human participants were reviewed and approved by Psychology Research Ethics Review Board, Graduate School of Education, Kyoto University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

HA and EM conceived the idea of the study. HA conducted experiments (including preliminary experiments) and drafted the original manuscript. EV developed a data scoring plan, and HA scored. EV and EM oversaw the rigor of the scoring process and its results. All authors developed the statistical analysis plan. HA conducted the analysis and EV additionally checked the results. All authors contributed to the interpretation of the results. EM supervised the conduct of this study. All authors reviewed the manuscript draft and revised it critically on intellectual content. All authors contributed to the article and approved the submitted version.

Acknowledgments

HA's work on this study was supported by research grants received from the Future Education Research Institute, and the Japan Society for the Promotion of Science (JSPS: Grant for Young Scientists, 20J23507). EM's work on this research was supported by a grant-in-aid (20K20516) received from JSPS. The authors express their gratitude to the students and teachers of Gifu Shotoku Gakuen University Junior High School, Japan (especially Yukio Fuwa, the Head Teacher) and the staff of Achieve Academy, Japan (especially Noriko Hanaki, the School Manager) for their assistance in the collection and analysis of the data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learn. Instr.* 16, 183–198. doi: 10.1016/j.learninstruc.2006.03.001

Ainsworth, S., and Th Loizou, A. (2003). The effects of self-explaining when learning with text or diagrams. *Cogn. Sci.* 27, 669–681. doi: 10.1016/S0364-0213(03)00033-8

Ayabe, H., Manalo, E., Fukuda, M., and Sadato, N. (2021). "What diagrams are considered useful for solving mathematical word problems in Japan?" in *Diagrammatic representation and inference. Diagrams 2021. Lecture notes in artificial intelligence*. eds. A. Basu, G. Stapleton, S. Linker, C. Legg, E. Manalo and P. Viana (Cham: Springer), 12909, 79–12983.

Baer, D. M., Wolf, M. M., and Risley, T. R. (1968). Some current dimensions of applied behaviour analysis. *J. Appl. Behav. Anal.* 1, 91–97. doi: 10.1901/jaba.1968.1-91

Boonen, A. J. H., van Wesel, F., and Jolles, J., & van der Schoot (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: an item-level analysis in elementary school children. *Int. J. Educ. Res.*, 68, 15–26. doi: 10.1016/j.ijer.2014.08.001

Butcher, K. R. (2006). Learning from text with diagrams: promoting mental model development and inference generation. *J. Educ. Psychol.* 98, 182–197. doi: 10.1037/0022-0663.98.1.182

Chu, J., Rittle-Johnson, B., and Fyfe, E. R. (2017). Diagrams benefit symbolic problem-solving. *Br. J. Educ. Psychol.* 87, 273–287. doi: 10.1111/bjep.12149

Cooper, J. L., Sidney, P. G., and Alibali, M. W. (2018). Who benefits from diagrams and illustrations in math problems? Ability and attitudes matter. *Appl. Cogn. Psychol.* 32, 24–38. doi: 10.1002/acp.3371

Corter, J. E., and Zahner, D. C. (2007). Use of external visual representations in probability problem solving. *Stat. Educ. Res. J.* 6, 22–50. doi: 10.52041/serj.v6i1.492

DiSessa, A. A. (2004). Metarepresentation: native competence and targets for instruction. *Cogn. Instr.* 22, 293–331. doi: 10.1207/s1532690xci2203_2

Duval, R. (1999). "Representation, vision and visualization: cognitive functions in mathematical thinking. Basic issues for learning." in *Twenty First Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, 1, 3–26*. Available at: http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED466379%5Cnhttp://informahealthcare.com/doi/abs/10.1076/noph.25.1.3.7140

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educ. Stud. Math.* 61, 103–131. doi: 10.1007/s10649-006-0400-z

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fuchs, L., Fuchs, D., Seethaler, P. M., and Barnes, M. A. (2020). Addressing the role of working memory in mathematical word-problem solving when designing intervention for struggling learners. *ZDM-Math. Edu.* 52, 87–96. doi: 10.1007/s11858-019-01070-8

Garner, R. (1990). When children and adults do not use learning strategies: toward a theory of settings. *Rev. Educ. Res.* 60, 517–529. doi: 10.3102/00346543060004517

Grawemeyer, B., and Cox, R. (2008). "The effects of users' background diagram knowledge and task characteristics upon information display selection," in *Diagrams 2008, lecture notes in artificial intelligence*. eds. G. Stapleton, J. Howse and J. Lee, vol. *5223* (Berlin: Springer), 321–334.

Hawkins, N. G., Sanson-Fisher, R. W., Shakeshaft, A., D'Este, C., and Green, L. W. (2007). The multiple baseline design for evaluating population-based research. *Am. J. Prev. Med.* 33, 162–168. doi: 10.1016/j.amepre.2007.03.020

Hegarty, M., and Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *J. Educ. Psychol.* 91, 684–689. doi: 10.1037/0022-0663.91.4.684

Hegarty, M., Mayer, R. E., and Monk, C. A. (1995). Comprehension of arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *J. Educ. Psychol.* 87, 18–32. doi: 10.1037/0022-0663.87.1.18

Hembree, R. (1992). Experiments and relational studies in problem solving: a meta-analysis. *J. Res. Math. Educ.* 23, 242–273. doi: 10.2307/749120

Hurley, S. M., and Novick, L. R. (2010). Solving problems using matrix, network, and hierarchy diagrams: the consequences of violating construction conventions. *Q. J. Exp. Psychol.* 63, 275–290. doi: 10.1080/17470210902888908

Jitendra, A. K., Griffin, C. C., Haria, P., Leh, J., Adams, A., and Kaduvettoor, A. (2007). A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *J. Educ. Psychol.* 99, 115–127. doi: 10.1037/0022-0663.99.1.115

Kintsch, W., and Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychol. Rev.* 92, 109–129. doi: 10.1037/0033-295X.92.1.109

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., and van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learn. Instr.* 30, 32–42. doi: 10.1016/j.learninstruc.2013.12.001

Lewis, A. B., and Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *J. Educ. Psychol.* 79, 363–371. doi: 10.1037/0022-0663.79.4.363

Manalo, E., Uesaka, Y., Chen, O., and Ayabe, H. (2019). "Showing what it looks like: teaching students how to use diagrams in problem solving, communication, and thinking," in *Deeper learning, dialogic learning, and critical thinking: Research-based strategies for the classroom*. ed. E. Manalo (London: Routledge), 231–246.

Mayer, R. E., Lewis, A. B., and Hegarty, M. (1992). "Mathematical misunderstandings: qualitative reasoning about quantitative problems," in *The nature and origins of mathematical skills*. ed. J. I. D. Campbell (Amsterdam: Elsevier), 137–154.

Morgan, D. L., and Morgan, R. K. (2009). *Single-case research methods for the behavioral and health sciences.* Thousand Oaks, CA: Sage.

Murata, A. (2008). Mathematics teaching and learning as a mediating process: the case of tape diagrams. *Math. Think. Learn.* 10, 374–406. doi: 10.1080/10986060802291642

Nistal, A. A., van Dooren, W., Clarebout, G., Elen, J., and Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: a critical review. *ZDM - Int. J. Math. Educ.* 41, 627–636. doi: 10.1007/s11858-009-0189-1

Novick, L. R., and Hurley, S. M. (2001). To matrix, network, or hierarchy: that is the question. *Cogn. Psychol.* 42, 158–216. doi: 10.1006/cogp.2000.0746

OECD (2019). Programme for international student assessment (PISA): results from PISA 2018, country note: Japan. Avalailbe at: https://www.oecd.org/pisa/publications/PISA2018_CN_JPN.pdf (Accessed February 12, 2021).

Paris, S. G., Lipson, M. Y., and Wixson, K. K. (1983). Becoming a strategic reader. *Cont. Edu. Psychol.* 8, 293–316. doi: 10.1016/0361-476X(83)90018-8

Reed, S. K. (1999). *Word problems: Research and curriculum reform.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Rivera, D., and Smith, D. D. (1988). Using a demonstration strategy to teach midschool students with learning disabilities how to compute long division. *J. Learn. Disabil.* 21, 77–81. doi: 10.1177/002221948802100203

Schmider, E., Ziegler, M., Danay, E., Beyer, L., and Bühner, M. (2010). Is it really robust?: reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology* 6, 147–151. doi: 10.1027/1614-2241/a000016

Schnotz, W., and Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educ. Psychol. Rev.* 19, 469–508. doi: 10.1007/s10648-007-9053-4

Schnotz, W., and Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: visualization effects on mental model construction. *Instr. Sci.* 36, 175–190. doi: 10.1007/s11251-007-9029-2

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Cambridge, MA: Academic Press.

Stylianou, D. A., and Silver, E. A. (2004). The role of visual representations in advanced mathematical problem solving: an examination of expert-novice similarities and differences. *Math. Think. Learn.* 6, 353–387. doi: 10.1207/s15327833mtl0604_1

Suthers, D. D. (2003). Representational guidance for collaborative inquiry. *Arg. Learn* 1994, 27–46. doi: 10.1007/978-94-017-0781-7_2

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J., van Merrienboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Uesaka, Y., and Manalo, E. (2006). "Active comparison as a means of promoting the development of abstract conditional knowledge and appropriate choice of diagrams in math word problem solving" in *Diagrams 2006. Lecture notes in artificial intelligence 4045*. eds. D. Barker-Plummer, R. Cox and N. Swoboda (Berlin: Springer), 181–195.

Uesaka, Y., and Manalo, E. (2012). Task-related factors that influence the spontaneous use of diagrams in math word problems. *Appl. Cogn. Psychol.* 26, 251–260. doi: 10.1002/acp.1816

Uesaka, Y., Manalo, E., and Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learn. Instr.* 17, 322–335. doi: 10.1016/j.learninstruc. 2007.02

Uesaka, Y., Manalo, E., and Ichikawa, S. (2010). "The effects of perception of efficacy and diagram construction skills on students' spontaneous use of diagrams when solving math word problems" in *Diagrams 2010. Lecture notes in artificial intelligence 6170*. eds. A. K. Goel, M. Jamnik and N. H. Narayanan (Berlin Springer), 197–211.

van der Meij, J., van Amelsvoort, M., and Anjewierden, A. (2017). How design guides learning from matrix diagrams. *Instr. Sci.* 45, 751–767. doi: 10.1007/s11251-017-9425-1

van Garderen, D. (2007). Teaching students with LD to use diagrams to solve mathematical word problems. *J. Learn. Disabil.* 40, 540–553. doi: 10.1177/00222194070400060501

van Garderen, D., Scheuermann, A., and Jackson, C. (2012). Examining how students with diverse abilities use diagrams to solve mathematics word problems. *Learn. Disabil. Q.* 36, 145–160. doi: 10.1177/0731948712438558

van Meter, P., and Garner, J. K. (2005). The promise and practice of learner-generated drawing: literature review and synthesis. *Educ. Psychol. Rev.* 17, 285–325. doi: 10.1007/s10648-005-8136-3

Verschaffel, L., Schukajlow, S., Star, J., and Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM-Math. Edu.* 52, 1–16. doi: 10.1007/s11858-020-01130-4

Wesnes, K., and Pincock, C. (2002). Practice effects on cognitive tasks: a major problem? *Lancet Neurol.* 1:473. doi: 10.1016/S1474-4422(02)00236-3

Zahner, D., and Corter, J. E. (2010). The process of probability problem solving: use of external visual representations. *Math. Think. Learn.* 12, 177–204. doi: 10.1080/10986061003654240

Zhang, J. (1997). The nature of external representations in problem solving. *Cogn. Sci.* 21, 179–217. doi: 10.1016/S0364-0213(99)80022-6

Zhang, J., and Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cogn. Sci.* 18, 87–122. doi: 10.1207/s15516709cog1801_3

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpsyg.2022.1028145

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University,
China

REVIEWED BY
Haode Zuo,
Yangzhou University,
China
Lakshmi Narayan Mishra,
VIT University,
India

*CORRESPONDENCE
Mohd Effendi Ewan Mohd Matore
effendi@ukm.edu.my

SPECIALTY SECTION
This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 25 August 2022
ACCEPTED 29 September 2022
PUBLISHED 20 October 2022

CITATION
Sim SH and Mohd Matore MEE (2022) The relationship of Grasha–Riechmann Teaching Styles with teaching experience of National-Type Chinese Primary Schools Mathematics Teacher.
Front. Psychol. 13:1028145.
doi: 10.3389/fpsyg.2022.1028145

The relationship of Grasha–Riechmann Teaching Styles with teaching experience of National-Type Chinese Primary Schools Mathematics Teacher

Sze Hui Sim[1] and Mohd Effendi Ewan Mohd Matore[2]*

[1]Sekolah Jenis Kebangsaan Cina Kepong 2, Selangor, Malaysia, [2]Faculty of Education, Research Centre of Education Leadership and Policy, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia

Grasha–Riechmann Teaching Styles have a high potential to be applied in Mathematics especially to help increase teacher educators' knowledge. However, very little attention has been paid to the study of identifying the teaching style patterns of Mathematics teachers at the primary school National-Type Chinese Primary Schools or *Sekolah Jenis Kebangsaan Cina* SJKC. There is increasing concern about how this teaching style related to the teaching experience. This study aims to identify the patterns of Grasha–Riechmann Teaching Styles among primary school Mathematics teachers and the relationship between Grasha–Riechmann Teaching Styles with teaching experience. The quantitative approach through a survey was applied to 97 Mathematics teachers of SJKC Kepong, Kuala Lumpur using the simple random sampling method. The instrument was adapted from the Grasha–Riechmann Teaching Styles Questionnaire (1996), which measures five teaching styles such as Personal Model Teaching Style, Expert Teaching Style, Formal Authority Teaching Style, Delegator Teaching Style, and Facilitator Teaching Style. The patterns showed that the Personal Model Teaching Style is the most dominant, and the Facilitator Teaching as the least dominant style. The Spearman's Rho Correlation also reported a very weak significant correlation between Grasha–Riechmann Teaching Styles with the teachers' Mathematics teaching experience, specifically for Expert, Formal Authority, and Facilitator Teaching Styles. The study provides practical implications for educators' professional development to diversify the training of teachers by experience and adapt them to the needs of student learning in primary school. These findings trigger ideas to get a better understanding by other demographic variables such as gender, age, and complexity of Mathematics subject.

KEYWORDS

Grasha–Riechmann, teaching style, teaching experience, mathematics, teacher, relationship, National-Type Chinese Primary Schools

10.3389/fpsyg.2022.1028145

Introduction

Statistics of student enrolment in Science, Technology, Engineering, and Mathematics (STEM) in Malaysia reported an increasingly alarming downward pattern. Salhani (2019) found a decline in student involvement in STEM fields from a 49% increase in 2012 to a 44% drop in 2018. This scenario clearly shows a decline in the involvement of nearly 6,000 students a year in STEM fields. This indirectly contributes to the significant difference in student involvement with only 334,742 students in STEM fields in higher learning institutions compared to 570,858 students in non-STEM fields (Wan Faizal, 2019). The phenomenon of student decline in STEM fields is not in line with the target of the Malaysia Education Development Plan 2013–2025 to attract the interest and awareness of the community in STEM fields in the second wave of PPPM in 2016–2020. Although many initiatives have been undertaken by the ministry to increase the interest of students and the community in STEM fields, the pattern of decline remains.

Therefore, the role of Mathematics teachers in primary schools is very important in inculcating and nurturing students' interest in STEM fields since the early schooling stage. Ainonmadiah et al. (2016) explained that teachers and students have close interpersonal relationships, where the teaching styles of teachers in the classroom are a component or element that ensures a sense of learning among students. In addition, the methods and practices of learning and facilitation (PdPc) are said to have a direct impact on students' perceptions and concerns in mastering Science and Mathematics (Anis Humaira et al., 2019). This shows that the role of teachers in the presentation of PdPc to some extent influences students' decisions and interests in the field of Mathematics.

Teacher quality gives the highest contribution to the success of a student. Based on a study by Nur Farhah and Fatimah Wati (2018), what students obtain is not dependent on the school attended but on the teachers at the school. Creative and innovative teaching practices can attract students (Rumainah and Faridah, 2017). Grasha (1996) also stated that the teaching styles of teachers are crucial for creating an ideal learning climate in the classroom. A good teaching style pattern among Mathematics teachers in primary schools is important so that the presentation of PdPc by the teachers can provide a positive perception and influence among students on the field of Mathematics. The literature also found a lack of research on the patterns of Grasha–Riechmann Teaching Styles among Mathematics teachers at the primary school level, particularly for National-Type Chinese Primary Schools (SJKC) compared to mainstream national schools.

Meanwhile, Normiati and Abdul (2019) opined that the teaching strategy of senior teachers is more traditional whereby the teaching experience factor that is honed by the interaction of the teachers' work environment promises maturity and expertise, especially in the development of effective teaching. The analysis also found that teachers with 6 to 10 years of teaching experience complied with each instruction and implemented teaching in accordance with the guidelines of the education system. Meanwhile, the group of teachers with low experience, specifically 1 to 5 years of teaching experience, scarcely implemented such teaching and was not proficient in carrying out the teaching concept of 21st Century Learning (PAK21). Dewaele et al. (2018) also proved that teaching experience has a statistically significant influence on the creativity, classroom management, and pedagogical skills of teachers. More experienced teachers in the profession are also more creative in the classroom, are better at managing classroom activities, and report stronger pedagogical skills.

Nevertheless, there is a lack of empirical evidence on the link between teaching experience and the Grasha–Riechmann Teaching Styles among primary school Mathematics teachers, which has not been widely tested in the Malaysian educational context specifically for National-Type Chinese Primary Schools (SJKC). Most of the studies only explained in depth on teaching experiences that are just emphasizing on the certain group of teachers. However, it was very limited discussion on Mathematics primary school teacher. The lacking information will affect the effort on matching the teacher training quality with the teaching style in Chinese schools. In this case, the preference of teaching style by experience will help teachers to adapt creatively their teaching skill in Mathematics.

Accordingly, this study was conducted to investigate the relationship between Grasha–Riechmann Teaching Styles and the teachers' Mathematics teaching experience.

Therefore, this study has two objectives:

1. Determine the patterns of Grasha–Riechmann Teaching Styles practiced by primary school Mathematics teachers.
2. Determine whether there is a significant relationship between Grasha–Riechmann Teaching Styles of Mathematics teachers with their Mathematics teaching experience.

Hence, the following hypotheses are constructed:

> *Ho1:* There is no significant relationship between the overall Grasha–Riechmann Teaching Styles of Mathematics teachers with their Mathematics teaching experience.
>
> *Ho2:* There is no significant relationship between the Expert Teaching Style of Mathematics teachers with their Mathematics teaching experience.
>
> *Ho3:* There is no significant relationship between the Formal Authority Teaching Style of Mathematics teachers with their Mathematics teaching experience.
>
> *Ho4:* There is no significant relationship between the Personal Model Teaching Style of Mathematics teachers with their Mathematics teaching experience.

Ho5: There is no significant relationship between the Facilitator Teaching Style of Mathematics teachers with their Mathematics teaching experience.

Ho6: There is no significant relationship between the Delegator Teaching Style of Mathematics teachers with their Mathematics teaching experience.

Literature review

Teaching styles

Mazaheri and Ayatollahi (2019) defined teaching styles as teachers' preferred ways to solve problems, perform tasks, and make decisions in the teaching process. Heydarnejad et al. (2017) defined teaching style as teachers' personal qualities and attitudes in teaching, which are reflected through the use of teaching techniques, activities, and approaches in teaching specific subjects in the classroom. In other words, the teaching style is a combination of motivation, personality, attitude, belief, and strategies in teaching (Karimnia and Mohammdi, 2019). Therefore, the teaching styles of teachers represent their behavior while teaching in the classroom and are one of the main determining factors for the success of student learning (Baradaran, 2016; Rosalia, 2017).

The teaching styles outlined by Grasha (1996) refer to the beliefs, behaviors, and needs of teachers that emerge in an educational context. Grasha believed that a teacher's teaching style reflects the personal qualities of the teacher in terms of how to teach, guide, and direct the teaching process, thus impacting students and their ability to learn. In general, the success or failure of students is associated with a teacher's teaching style, which is directly related to the teaching methods used during teaching. Indirectly, the teaching style turns into one of the components of a comprehensive transfer of teaching content. The teaching style may also be influenced by factors such as educational background, teaching experience, cultural background, and individual personal interests (Nouraey and Karimnia, 2016; Tavakoli and Karimnia, 2017). These factors can be identified by observing and studying teacher behavior.

Literature related to teaching styles displays various theories, models, and categorization of teaching styles through the use of different terminologies. For example, this includes the categorization of teaching styles to a didactic direct style and a student-centered indirect style (Flanders, 1970), Formal-Informal (Bennett et al., 1976), Open-Traditional (Solomon and Kendall, 1979), Intellectual Excitement-Interpersonal Rapport (Lowman, 1995), and Expert, Formal Authority, Personal Model, Facilitator, and Delegator Teaching Styles (Grasha, 1996) are among the few terminologies used to better explain these constructs. In the present study, only the Grasha–Riechmann Teaching Style model is discussed because it has potential in the context of Mathematics teaching.

Grasha–Riechmann Teaching Styles (1996)

Grasha (Heydarnejad et al., 2017) argued that teaching styles involve constant teacher behavior in interaction with students during the teaching-learning process. Grasha described teaching styles as a criterion for personal qualities and behaviors that govern the way teachers manage classes. Hence, it can be said that teaching styles consist of all techniques, activities, and teaching approaches used by a teacher in the teaching process. The five dimensions of the Grasha–Riechmann Teaching Styles are Expert Teaching Style, Formal Authority Teaching Style, Personal Model Teaching Style, Facilitator Teaching Style, and Delegator Teaching Style. The dimensions or attributes of the Grasha–Riechmann Teaching Styles in terms of teacher roles, student characteristics, and the aspect of the advantages and disadvantages of each teaching style are explained in the following sub-sections.

Expert Teaching Style

Grasha (1996) argued that teachers with the Expert Teaching Style have knowledge and expertise on what students want to learn. The Expert Teaching Style makes teachers maintain their status as experts among their students by displaying accurate and comprehensive knowledge. In this regard, teachers with the Expert Teaching Style encourage students to face challenging situations to develop competencies in learning. As an expert, teachers also play a role in conveying information and expecting students to learn what they receive and take advantage of the information presented. Teachers are also careful in communicating information and will ensure that students are always ready.

In the Expert Teaching Style of Bergil and Erçevik (2019) can be seen as an advantage where teachers have accurate and comprehensive knowledge, skills, and information on the scope of targets to be taught to students. This comprehensive consolidation of knowledge, information, and skills can benefit experienced students. However, it should be emphasized that excessive use of Expert Teaching Styles may scare and curb the learning of students who are inexperienced or do not have a basic knowledge of the expected target topic. Furthermore, the presentation of knowledge or information conveyed by the teachers may not be of interest and motivation to the students at all. Additionally, the display of teacher knowledge and skills may not always show students the implicit thought process that produces an answer.

Formal Authority Teaching Style

Grasha (1996) reported that Formal Authority Teaching Style requires teachers to have status or position among students. This is because teachers are considered members of schools or faculties who contribute to the teaching and learning process by providing positive and negative feedback to the students. In this context, the teachers create concrete learning situations by setting learning objectives, rules, expectations, and principles of learning for students. Accordingly, teachers with the Formal Authority Teaching Style focus on preparing students with the necessary thinking structure in learning. The teachers also care about the correct, accepted, and standard way of doing things. Thus, students can be motivated through quality, effective, and meaningful learning methods.

The main advantages of the Formal Authority Teaching Style are emphasized on teacher expectations, methods, and standard ways to do things during the teaching and learning process. However, the use of the Formal Authority Teaching Style can also result in limited, permanent, and inflexible student engagement in the learning process. Hence, a strong attachment to the Formal Authority Teaching Style can contribute to rigid, standard, and less flexible ways of managing students.

Personal Model Teaching Style

Grasha (1996) explained that Personal Model Teaching Style refers to teachers who teach based on their own example. They will directly guide and encourage students to emulate it. Teachers with Personal Model Teaching Style set the prototypes for thinking and behaving. In this regard, the teachers constantly supervise, guide, and instruct students by showing them how to do things. In doing so, the teachers motivate students to observe, imitate, or reflect on the methods and approaches provided by them. The need for direct observation and imitation by students is the main strength of the Personal Model Teaching Style (Bergil and Erçevik, 2019). Teachers with Personal Model Teaching Style encourage students to observe and then imitate the teachers' approach that is considered appropriate. However, some teachers may believe that their approach is the best and this consequently makes some students feel that they have low capacities if they cannot meet those expectations and standards. As a result, the students will feel less confident and demotivated in learning that exceeds their ability.

Facilitator Teaching Style

Grasha (1996) stated that the Facilitator Teaching Style emphasizes teacher and student interaction. Therefore, teachers with this teaching style act as facilitators in the classroom. They guide students by asking questions, exploring options, suggesting alternatives, and encouraging students to make informed decisions. The main goal of teaching is to nurture students who are independent and have high self-efficacy, where teachers encourage students to initiate and carry out their own responsibilities in learning. The choices, questions, and opportunities provided by the teachers serve as a guide and lead the students in learning situations. In the Facilitator Teaching Style, students can develop their own learning criteria. This style also shows that teachers are more likely to guide students to carry out project-based activities and provide optimal motivation to students. In this regard, the teachers work with the students on project assignments on a consultative basis by providing support and encouragement to the students. The main strength of the Facilitator Teaching Style is that the personal flexibility given by teachers is focused on the needs and objectives of student learning (Bergil and Erçevik, 2019). This will enable students to explore alternative options and methods of action. Nonetheless, the main drawback of the Facilitator Teaching Style is that it is time-consuming. Teachers and students may need more time, especially in the implementation of practical activities or project assignments. In addition, the Facilitator Teaching Style may also become ineffective when a more direct approach is needed. In fact, students may feel uncomfortable if this mechanism is not used positively and in a motivational manner.

Delegator Teaching Style

Delegator Teaching Style refers to teachers who emphasize the development of a student's self-capacity. Students will be encouraged to conduct self-learning such as projects and teachers will act as a source of reference. The Delegator Teaching Style aims to develop students' competencies by giving them autonomous characteristics. In this style, students are expected to work on projects independently and function as members with autonomous powers within their group. When the students need help, they can refer to teachers as a source of information to meet their needs (Grasha, 1996).

In the use of the Delegator Teaching Style, students consider themselves independent, capable, and autonomous. As a result, each student has the opportunity to become initiative and self-reflect by evaluating himself or herself. Nevertheless, teachers are sometimes confused about students' willingness to take responsibility and face the need for autonomy. The students among youth also need to self-assess their ability to face adversity in life (Mohd Effendi Ewan et al., 2021). This situation can indirectly cause the students to feel worried and anxious in their efforts to carry out the tasks given by the teachers. Therefore, as a weakness, it should be borne in mind that students may not have the desired ability to fulfill their autonomous obligations. Besides, students may need rigorous supervision and intensive encouragement to overcome a sense of anxiety and reform themselves in learning norms.

Present study

Studies on the patterns of Grasha–Riechmann Teaching Styles were found to be very limited both locally and internationally. Over the most recent 5 years, fewer research reports were found to identify the relationship of teaching experience with the Grasha–Riechmann Teaching Styles among teachers. In fact, international research on the Grasha–Riechmann Teaching Styles of teachers is barely focusing on teaching experience but is more likely to focus on teacher creativity and burnout (Ghanizadeh and Jahedizadeh, 2016), teacher self-efficacy (Baleghizadeh and Shakouri, 2017), student academic achievements (Khalid et al., 2017; Martin, 2019), student motivation (Massaada, 2016;

Rosalia, 2017), as well as teachers' behavior management and instructional management (Kazemi and Soleimani, 2016).

In Malaysia, one study had conducted to identify the Grasha–Riechmann Teaching Styles practiced among science lecturers of pre-university colleges in Penang (Anis Humaira et al., 2019). The findings found that the dominant teaching style of the Science lecturers was the Expert Teaching Style, followed by the Personal Model Teaching Style, Facilitator Teaching Style, and Delegator Teaching Style. Meanwhile, the Formal Authority Teaching Style was least practiced by lecturers. The Spearman's Rho analysis found that it has no significant relationship between teaching styles and teaching experience ($r = 0.089$) with professional qualifications ($r = 0.193$). Ainonmadiah et al. (2016) also reviewed the relationship between Grasha–Riechmann Teaching Styles and the level of skipping secondary school in Bachok District, Kelantan. The findings showed that the most dominant category of teaching style practiced by the teachers was the Personal Model Teaching Style, followed by the Delegator Teaching Style, Expert Teaching Style, Formal Authority Teaching Style, and Facilitator Teaching Style. The analysis proved that there was a significant relationship related to the teaching styles of teachers based on the flow of subjects. However, Pearson's Correlation Test showed no significant link between teaching styles and the level of skipping school ($r = 0.062$).

Another study by Mazeni and Hasmadi (2017) has identified the Grasha–Riechmann Teaching Styles practiced by Kemas Kindergarten teachers in Kelantan. A total of 50 kindergarten teachers were selected through strata sampling. The findings showed Delegator Teaching Style as the main teaching style, Facilitator as the second teaching style, followed by Personal Model, Expert Teaching Styles, and finally Formal Authority Teaching Style. The mean score achievement proved that the majority of the teachers used Delegator Teaching Style, while the least used teaching style was the Formal Authority Teaching Style. Teachers at Kemas Kindergarten used the Delegator Teaching Style because they had no formal educational background in early childhood education. In a different study, Nur Liyana and Zakiah (2017) findings also suggested that the Personal Model Teaching Style was most dominant.

As for the global context, Alami and Ivaturi (2016) in Oman reviewed a comparison of the Grasha–Riechmann Teaching Styles used by English lecturers at the Salalah College of Technology (SCT). The findings showed that the lecturers used different Grasha–Riechmann Teaching Styles, where the Expert Teaching Style was the dominant teaching style. It proved that the lecturers preferred to act as experts by displaying knowledge, disseminating information, and encouraging students to apply the information provided to them. Furthermore, this study found that lecturers with less than 5 years of teaching experience preferred to use the Facilitator Teaching Styles compared to other teaching styles. The Delegator Teaching Style was also used by lecturers with 5–10 years of teaching experience. However, overall, the Chi-Square Test proved no great connection between Grasha–Riechmann Teaching Styles and the teaching experience of English lecturers.

In Iran, Mazaheri and Ayatollahi (2019) conducted a study to explore the relationships among brain dominance, teaching experience, and the Grasha–Riechmann Teaching Styles of English teachers. This study involved 100 Iranian English teachers who taught at Shiraz high school. The results showed significant links among brain dominance, teaching experience, and the Grasha–Riechmann Teaching Styles of teachers. The results of Pearson's correlation showed a significant moderate relationship between teaching experience and the Grasha–Riechmann Teaching Styles of teachers ($r = 0.324$, $p < 0.05$). It was also found that the teaching experience can predict the teaching style of a teacher with a beta value of 0.265. Meanwhile, in Turkey, Bergil and Erçevik (2019) concluded that the Personal Model and Facilitator Teaching Styles with 32.4% and 35.3%, respectively, were the choices of the English teachers. On the other hand, the Delegator and Expert Teaching Styles with 5.9% and 11.8%, respectively, were less applied by the teachers.

Another study in Spain (Fernandez-Rivas and Espada-Mateos, 2019) has analyzed the use of Grasha–Riechmann Teaching Styles among Physical Education teachers with reference to teaching experience and age of the teachers. The study used a sample of 455 Physical Education teachers covering a wide range of age groups and teaching experiences. The findings showed a significant relationship between the use of Grasha–Riechmann Teaching Styles among teachers and their teaching experience. The study also showed that teachers who have worked between 1 and 20 years used Facilitator and Delegator Teaching Styles more often than the teachers with more than 21 years of experience at school. In addition, the results proved that younger and less experienced teachers regularly used traditional teaching styles.

In addition, Martin (Martin, 2019) determined the dominant teaching style and the relationship between the teachers' grade levels and their teaching styles. Based on the findings of the study, the Personal Model Teaching Style was the most dominant teaching style among respondents. The Expert and Formal Authority Teaching Styles were often used by Level Two primary school teachers, while the Personal Model and Facilitator Teaching Styles were applied by Level One primary school teachers. Furthermore, the Facilitator and Personal Model Teaching Styles were also seen to produce the highest academic achievement in Mathematics.

In another study, Massaada (2016) analyzed the Grasha–Riechmann Teaching Styles applied by English teachers and their influence on student motivation at Majene State High School 2. The results showed that English teachers in Majene State High School 2 applied four teaching styles, namely Expert Style, Formal Authority, Personal Model, and Facilitator. Meanwhile, the dominant teaching styles in this study were Expert, Personal Model, and Formal Authority Teaching Styles.

Rosalia (2017) conducted a study to identify the Grasha–Riechmann Teaching Styles applied by English teachers as well as the most influential teaching style toward the motivation of students at SMK Negeri 5 Makassar. The results showed that English teachers at SMK Negeri 5 Makassar applied three teaching styles, namely Expert Teaching Style, Formal Authority Teaching Style, and Facilitator Teaching Style. The researchers found that the most influential teaching style toward student motivation was the Facilitator Teaching Style, which focuses on teacher-student interaction. The results also showed that the teachers tried to change their teaching style by applying fun activities.

Previous research on Grasha–Riechmann Teaching Styles has also been conducted by Kazemi and Soleimani (2016). The researchers randomly selected 103 English teachers working in private language learning centers. The results showed that Iranian English teachers followed the intervention class management approach and most of them used the Formal Authority Teaching Style. The Grasha–Riechmann Teaching Styles of teachers were also found to correlate significantly with their behavior management and instructor management.

Based on the study in Malaysia and worldwide, it can be concluded that there were clear gaps in past studies on the exploration of the Grasha–Riechmann teaching style patterns among primary school Mathematics teachers to be filled on the context of SJKC. Hence, this study is relevant to be carried out to increase the empirical evidence of existing knowledge in the teaching styles for Chinese-based type of school.

Materials and methods

The study used a survey with quantitative approach through questionnaires. Questionnaires were given to obtain respondents' feedback on the Grasha–Riechmann Teaching Styles among primary school Mathematics teachers. In this study, the independent variables include the teaching experience, while the dependent variable is the Grasha–Riechmann Teaching Styles of Mathematics teachers.

Samples

The population for this study involves Mathematics teachers from National-Type Chinese Primary Schools (SJKC) Kepong 1, Kepong 2, and Kepong 3 Sentul Zone, Kuala Lumpur. The total population of Mathematics teachers from three primary schools in Kepong is 130. Based on the sample calculation *via* Raosoft version 3.1.9.4 statistical software, the alpha value was recorded at 0.05. The calculated results further estimated a minimum sample size of 130 respondents; however, after considering a 50% possible loss rate (respondents refusing to participate or withdrawing), the recommended sample size for this study entails 97–98 respondents. In this study, random sampling was used easily for selecting the respondents of the study as it involves the selection of respondents with important knowledge or information that is consistent with the purpose of the study (Donkor, 2019). In the context of this study, the selected teachers only involve Mathematics teachers at the SJKC level in Kepong, Kuala Lumpur.

Instrumentation

The study used an instrument adapted from the Grasha–Riechmann Teaching Styles Questionnaire (1996; Grasha, 1996). Donkor (Donkor, 2019) explained that the use of questionnaires is appropriate for obtaining data and information quickly, especially when a large number of respondents is required. The instrument outlines five constructs that comprise five different types of teaching styles (Expert Teaching Style, Formal Authority Teaching Style, Personal Model Teaching Style, Facilitator Teaching Style, and Delegator Teaching Style). The instrument with a seven-point Likert scale was modified to a five-point Likert scale to avoid confusion and facilitate the interpretation or preference of the respondents. In addition, the researchers translated the questionnaire from English to Bahasa Malaysia through a direct translation. The instrument has also undergone face validity and content validity processes by experts in Pedagogy and Mathematics. Besides, the reliability aspect of the instrument was tested using Cronbach's Alpha analysis to test for internal consistency. Mailizar (2018) stated that internal consistency refers to consistency in instruments. All the 40 items in instrument had good internal consistency (Cronbach's Alpha from 0.64 to 0.83) and high reliability (coefficient from 0.65 to 0.86). The respondents' demographics such as gender and teaching experience are included in the questionnaire. Overall, the instrument comprises 40 items, specifically eight items in each construct. The distribution of the items by construct is depicted in Table 1.

The questionnaire consists of two sections, namely Section A (teacher information) and Section B (Grasha–Riechmann Teaching Styles of teachers). The score for each item is determined by the respondents' responses on a 5-point Likert scale: 1 = Strongly Disagree; 2 = Disagree; 3 = Neither Agree nor Disagree; 4 = Agree; and 5 = Strongly Agree. The questionnaire used the Likert scale because this scale is easy to administer and takes a while to receive feedback from the respondents (Seng et al., 2016).

TABLE 1 Distribution of items by construct.

Teaching style	Item number	Number of items
Expert Teaching Style	1–8	8
Formal Authority Teaching Style	9–16	8
Personal Model Teaching Style	17–24	8
Facilitator Teaching Style	25–32	8
Delegator Teaching Style	33–40	8
Total		40

Administration

For the administration of data collection, permission and approval were officially obtained from the Department of Education Office (JPP) Sentul Division, Kuala Lumpur. Subsequently, a pilot study was carried out to test the validity and reliability of each item in the questionnaire. A total of 35 Mathematics teachers from the SJKC in Keramat Zone, who have the same profile characteristics as the actual respondents, were selected as respondents in the pilot study. The Grasha–Riechmann Teaching Styles Questionnaire was prepared in Google Forms and the questionnaire link was given to the mathematics teachers involved. On average, the pilot study respondents completed the questionnaire within 10 to 15 min. Through this pilot study, flaws and confusion found in the questionnaire such as spelling errors, the ambiguity of the meaning of statements, and fewer clear instructions were improved based on the comments given. The respondents indicated that the items presented in the questionnaire were easy to understand and this proves that the face validity has been met. Subsequently, content validity through experts and the actual study were conducted.

Data analysis

Data preparation

The quantitative data were processed using IBM Statistical Package for Social Science (SPSS) version 26.0 software. Descriptive analysis and inference analysis were used to obtain the results. Data analysis based on the research question is shown in Table 2.

TABLE 2 Research objectives and data analysis.

Research objective	Data analysis
1. Determine the pattern of Grasha–Riechmann Teaching Styles practiced by primary school Mathematics teachers.	Frequency and descriptive
2. Determine whether there is a significant relationship between the Grasha–Riechmann Teaching Styles of Mathematics teachers with their Mathematics teaching experience.	Spearman rank

For the first objective, descriptive frequencies and mean scores of the five constructs of Grasha–Riechmann Teaching Styles were obtained and compared to identify the patterns of Grasha–Riechmann Teaching Styles practiced by the respondents. For the second objectives, some considerations were made prior to the use of parametric tests such as Pearson's Correlation. One of the main considerations is the normality of data. In the context of this study, non-parametric tests such as Spearman Rank were used because the Kolmogorov–Smirnov Test showed that the data did not meet the normal distribution requirement due to a significance level of $p = 0.000$ ($p < 0.05$). Accordingly, the Spearman Rank Tests were selected as an alternative to parametric testing.

The correlational strength was determined based on Chua's (Chua, 2021) interpretation. According to Chua (2021), correlation coefficient (r) refers to the measurement value of the relationship between two variables and this r value ranges between +1.00 and −1.00. Since perfect correlation is rare in research, correlation coefficients are reported in two decimal points. Table 3 shows the strength levels of correlation coefficient values (r).

TABLE 3 Strength level of correlation coefficient size.

Correlation coefficient size (r)	Correlation strength
• 0.91 to 1.00 or −0.91 to −1.00	Very strong
• 0.71 to 0.90 or −0.71 or −0.90	Strong
• 0.51 to 0.70 or −0.51 to −0.70	Moderate
• 0.31 to 0.50 or −0.31 to −0.50	Weak
• 0.01 to 0.30 or −0.01 to −0.30	Very weak
0.00	No correlation

Results

The profile of the respondents is shown in Table 4. The number of female respondents is 42 (43.3%), while the number of male respondents is 55 (56.7%). Most of the respondents were 21–30 years old, which comprise 40 respondents with 41.2%. Meanwhile, respondents aged 31–40 years and 51–60 years constituted 23.8% and 20.6%, respectively, and those aged 41–50 years only constituted 14 respondents with 14.4%. As for the respondents' Mathematics teaching experience, a total of 53 respondents (54.6%) had 1–10 years of teaching experience in the mathematics subjects. This is followed by 18 respondents (18.6%) with 11–20 years of teaching experience and 15 respondents (15.5%) with 21–30 years of Mathematics teaching experience. Meanwhile, only 11 respondents (11.3%) had more than 31 years of experience in teaching Mathematics subjects.

For the first objective, the scores of frequencies, mode, mean average, and standard deviation for the five constructs of the Grasha–Riechmann Teaching Styles of teachers were obtained and compared to identify the patterns of the Grasha–Riechmann Teaching Styles practiced by primary school Mathematics teachers. The respondents' responses to the five Grasha–Riechmann Teaching Style constructs were explained separately. Table 5 shows that the majority of respondents agreed and strongly agreed with almost all items, alongside the mode value of 4 representing the scale of an agreement to all items, except Item 7 for the Expert Teaching Style construct. The item "I give students negative feedback if the performance is unsatisfactory" recorded a mode value of 2, which suggests that the majority of the

TABLE 4 Demographic profile.

Demographic profile	N	%
Gender		
Female	42	43.3
Male	55	56.7
Age		
21–30 years old	40	41.2
31–40 years old	23	23.8
41–50 years old	14	14.4
51–60 years old	20	20.6
Mathematics teaching experience		
1–10 years	53	54.6
11–20 years	18	18.6
21–30 years	15	15.5
31 years and above	11	11.3
Total	97	100

respondents were not in agreement with this statement. The mean scores also indicated that the dominant teaching style adopted by the primary school Mathematics teachers is Personal Model Teaching Style (Mean = 4.12, Standard Deviation = 0.62), followed by the Expert Teaching Style (Mean = 3.87, Standard Deviation = 0.69), Formal Authority Teaching Style (Mean = 3.86, Standard Deviation = 0.65), Delegator Teaching Style (Mean = 3.79, Standard Deviation = 0.69), and finally the Facilitator Teaching Style (Mean = 3.77, Standard Deviation = 0.76). This proves that the mathematics teachers were more inclined to the Personal Model Teaching Style than other teaching styles. However, these teachers used the Facilitator Teaching Style the least.

To answer the second research question, Table 6 indicates that H_{O1}, H_{O2}, H_{O3}, H_{O4}, H_{O5}, and H_{O6} have been rejected. The Spearman Rank test showed a very weak and significant positive correlation between the overall Grasha–Riechmann Teaching Styles and Mathematics teaching experience [$r_s(95) = 0.274$, $p = 0.007$]. This also indicates the likelihood that the more teaching experience that the teachers have in mathematics, the higher application of Grasha–Riechmann Teaching Styles among them.

Looking at each construct of teaching style, interesting findings that can be noted are the significant positive relationships for the Expert Teaching Style [r_s (95) = 0.266, $p = 0.009$], Formal Authority Teaching Style [r_s (95) = 0.217, $p = 0.033$], and Facilitator Teaching Style [r_s (95) = 0.345, $p = 0.001$]. The highest correlation was recorded by the Facilitator Teaching Style, followed by the Expert Teaching Style and the Formal Authority Teaching Style. The two more hypotheses that failed to be rejected were the Personal Model Teaching Style and the Delegator Teaching Style. Thus, there was no relationship between the Grasha–Riechmann Teaching Styles for both constructs and Mathematics teaching experience, as reported in Table 6. Figure 1 shows the scatter plots for the correlation between Mathematics teaching experience and the Grasha–Riechmann Teaching Styles. Scattered data plots with fit lines indicated poor relationships of the variables involved.

Discussion

This study focuses on identifying the patterns of Grasha–Riechmann Teaching Styles of primary school Mathematics teachers and the relationship between Grasha–Riechmann Teaching Styles with the teachers' Mathematics teaching experience.

The findings for the first research objective have shown that the Personal Model Teaching Style was most practiced by primary school Mathematics teachers. On the other hand, the Facilitator Teaching Style was practiced the least. Literature on the patterns of the Grasha–Riechmann Teaching Styles among teachers is very limited both locally and abroad. The findings are consistent with past findings in which the Personal Model Teaching Style precedes other teaching styles and is the top choice for educators (Abdull Sukor et al., 2014; Ainonmadiah et al., 2016; Beers, 2016; Martin, 2019). Studies by Nur Liyana and Zakiah (2017) and Ghanizadeh and Jahedizadeh (2016) also reported similar findings by which the Personal Model Teaching Style is dominant compared to other teaching styles. The contribution of this study is seen to reinforce the findings on the Grasha–Riechmann Teaching Style patterns practiced by primary school SJKC teachers in Malaysia.

The findings clearly demonstrated the trend of using the Personal Model Teaching Style where the primary school Mathematics teachers would teach based on their own examples, such as setting a prototype for students to think and behave by giving direct guidance and subsequently encouraging the students to emulate them (Grasha, 1996). This style also encourages the environment to be student-centered teaching and learning process that involves an inquiry process in problem-solving questions that can increase mathematics performance (Suik Fern and Mohd Effendi Ewan, 2022). In addition, teachers with a high Personal Model Teaching Style were found to be very concerned with the students' mastery of the lesson contents. They often demonstrated how to master the contents of lessons, concepts, and principles and relate them by giving examples based on their own experiences. In the context of Mathematics, the teachers preferred this teaching style as they can supervise, guide, direct, and motivate students to make observations and subsequently emulate the approach that the teachers have shown.

From a theoretical point of view, one of the logical explanations for the trend of using Personal Model Teaching Style among primary school Mathematics teachers in this study can be explained through Bandura's Social Learning Theory (Bandura, 1962), which explains that humans naturally learn through the process of observation and imitation. This theory emphasizes the need for direct observation and imitation by students in the use of the Personal Model Teaching Style. In this study, the mathematics subjects require individual attention and regular training to help improve students' academic achievements (Beers, 2016). Due to the conceptualized nature of Mathematics, primary school students who are in the early stages of learning are encouraged to observe and refer to the work steps or procedures shown by the teachers. The information they obtained is gradually stored in

TABLE 5 Frequency and descriptive analysis for Grasha–Riechmann Teaching Styles.

	Item code	Strongly disagree		Disagree		Neither agree nor disagree		Agree		Strongly agree		Mode	Mean	SD
		f	%	f	%	f	%	f	%	f	%			
Expert	S1	0	0	6	6.2	3	3.1	51	52.6	37	38.1	4	3.87	0.69
Teaching	S2	0	0	3	3.1	32	33	46	47.4	16	16.5	4		
Style	S3	0	0	0	0	12	12.4	67	69.1	18	18.6	4		
	S4	0	0	6	6.2	5	5.2	72	72.4	14	14.4	4		
	S5	0	0	1	1	16	16.5	76	78.4	4	4.1	4		
	S6	0	0	0	0	9	9.3	55	56.7	33	34.0	4		
	S7	12	12.4	35	36.1	23	23.7	18	18.6	9	9.3	2		
	S8	0	0	0	0	7	7.2	77	79.4	13	13.4	4		
Formal	M1	0	0	6	6.2	16	16.5	65	67	10	10.3	4	3.86	0.65
Authority	M2	0	0	0	0	9	9.3	74	76.3	14	14.4	4		
Teaching	M3	0	0	0	0	17	17.5	70	72.2	10	10.3	4		
Style	M4	0	0	3	3.1	15	15.5	71	73.2	8	8.2	4		
	M5	0	0	1	1.0	7	7.2	79	81.4	10	10.3	4		
	M6	0	0	6	6.2	13	13.4	65	67.0	13	13.4	4		
	M7	6	6.2	7	7.2	16	16.5	55	56.7	13	13.4	4		
	M8	0	0	1	1.0	37	38.1	47	48.5	12	12.4	4		
Personal	S1	0	0	8	8.2	24	24.7	45	46.4	20	20.6	4	4.12	0.62
Model	S2	0	0	1	1.0	10	10.3	59	60.8	27	27.8	4		
Teaching	S3	0	0	0	0	6	6.2	65	67.0	26	26.8	4		
Style	S4	0	0	0	0	7	7.2	54	55.7	36	37.1	4		
	S5	0	0	10	10.3	15	15.5	54	55.7	18	18.6	4		
	S6	0	0	0	0	1	1.0	74	76.3	22	22.7	4		
	S7	0	0	0	0	4	4.1	64	66.0	29	29.9	4		
	S8	0	0	0	0	3	3.1	73	75.3	21	21.6	4		
Facilitator	M1	6	6.2	11	11.3	19	19.6	42	43.3	19	19.6	4	3.77	0.76
Teaching	M2	6	6.2	0	0	11	11.3	73	75.3	7	7.2	4		
Style	M3	0	0	0	0	19	19.6	69	71.1	9	9.3	4		
	M4	0	0	7	7.2	9	9.3	71	73.2	10	10.3	4		
	M5	0	0	14	14.4	7	7.2	71	73.2	5	5.2	4		
	M6	0	0	31	32.0	13	13.4	43	44.3	10	10.3	4		
	M7	0	0	1	1.0	18	18.6	71	73.2	7	7.2	4		
	M8	0	0	0	0	11	11.3	65	67.0	21	21.6	4		
Delegator	S1	0	0	6	6.2	31	32.0	47	48.5	13	13.4	4	3.79	0.67
Teaching	S2	0	0	18	18.6	19	19.6	53	54.6	7	7.2	4		
Style	S3	0	0	7	7.2	36	37.1	48	49.5	6	6.2	4		
	S4	0	0	2	2.1	37	38.1	51	52.6	7	7.2	4		
	S5	0	0	0	0	19	19.6	59	60.8	19	19.6	4		
	S6	0	0	0	0	22	22.7	59	60.8	16	16.5	4		
	S7	0	0	0	0	17	17.5	59	60.8	21	21.6	4		
	S8	0	0	1	1.0	16	16.5	69	71.1	11	11.3	4		

memory and released when they responded to the mathematical questions given to them.

Preliminary research on Grasha (1994) also found that the Personal Model Teaching Style was used dominantly by all instructors at all levels of academic education, be it professors, associate professors, tutors, or teachers. Besides, the Personal Model Teaching Style was found to have been dominantly used in teaching courses at all higher education levels compared to other teaching styles. This clearly shows that teachers tend to use the Personal Model Teaching Style characterized by a hands-on approach and encourage students to observe and replicate it (Abdull Sukor et al., 2014; Martin, 2019). In short, the Personal

TABLE 6 Spearman's rho correlation for Grasha–Riechmann Teaching Styles.

Grasha–Riechmann teaching style	Mathematics teaching experience		Decision
	Correlation coefficient	Sig. (2-tailed)	
Overall	0.274**	0.007**	H_1 rejected
Expert Teaching Style	0.266**	0.009**	H_2 rejected
Formal Authority Teaching Style	0.217*	0.033**	H_3 rejected
Personal Model Teaching Style	0.062	0.548	H_4 failed to reject
Facilitator Teaching Style	0.345**	0.001**	H_5 rejected
Delegator Teaching Style	0.192	0.060	H_6 failed to reject

**Correlation is significant at the 0.01 level (2-tailed).

Model Teaching Style emphasizes teaching through observation and guidance where the teacher serves as a real model in controlled learning activities.

More interestingly, the frequent use of Personal Model Teaching Style among Mathematics teachers is also likely due to the potential of this teaching style toward the academic growth of students in Mathematics. Generally, the teaching style of teachers has a great impact on the motivation and achievement of students in a particular subject (Heydarnejad et al., 2017). To explain the relationship between Grasha–Riechmann Teaching Styles and the academic growth of students in Mathematics, Martin (2019) conducted a comprehensive quantitative correlation study on students' academic achievements in 37 U.S. international schools. This study also showed that the Personal Model Teaching Style produced the highest growth in academic achievement in Mathematics. Meanwhile, in Malaysia, Abdull Sukor et al. (2014) conducted a study to identify the relationship between the teaching styles of lecturers and university students' academic involvement. The study found that most students were more likely to engage in learning when the lecturers used the Personal Model Teaching Style to deliver their teaching.

The findings also showed the Facilitator Teaching Style as the least dominant teaching style among respondents. The findings are in line with Ainonmadiah et al. (2016) findings on teachers from five schools in Bachok District, Kelantan. The findings of this study can be explained through Piaget's Theory of Cognitive Development (Piaget, 1936), which classifies the cognitive development of children at the age of seven to 12 years old as a level of concrete operation marked by the use of clear and logical rules. Children of this age are seen applying logical thinking over concrete objects, but neither in the abstract nor hypotheses. Hence, the children's way of thinking is still limited as the focus is more on concrete objects and can only solve problems encountered directly (Astuti, 2018).

Therefore, the limited cognitive development of children at the primary school level has limited the use of the Facilitator Teaching Style that emphasizes students' self-learning through the implementation of practical activities or project assignments. This is because children at this stage are less capable of solving hypothetical problems and abstract tasks (Lutz et al., 2018). If the Facilitator Teaching Style is applied by teachers regardless of the needs of the students' level of learning, especially in primary schools, then it is very likely that this will result in the negative view of students of Mathematics due to the "Mathematical stress" factor experienced. Furthermore, the implementation of the teaching process to the physically and mentally unprepared group of students may also cause difficulties and waste of time. Hence, these factors are likely to contribute to the infrequent use of Facilitator Teaching Style among primary school Mathematics teachers. However, this explanation requires further research to be done in the future in order to empirically make stronger confirmation.

The current study also showed the implicit finding that primary school Mathematics teachers are more likely to apply teacher-centered teaching orientation (Personal Model Teaching Style, Expert Teaching Style, and Formal Authority Teaching Style) compared to student-centered teaching orientation (Delegator Teaching Style and Facilitator Teaching Style). However, this scenario is inconsistent with the demands of 21st Century Learning (PAK21) where both teacher-centered and student-centered teaching orientations should be balanced to achieve maximum student learning outcomes. The previous study also shows that the level of readiness of trainee teachers is high in terms of interest, knowledge, and skills in integrating PAK-21 in teaching and learning mathematics (Mazlini et al., 2021).

In a teacher-centered classroom, students become passive without having control over self-learning. In this context, teachers make all decisions regarding the curriculum, teaching methods, and assessments, thus hindering the development of student competencies and learning achievements (Dole et al., 2016; Goff, 2016; Lak et al., 2017). On the other hand, in a student-centered classroom, students are given more attention and responsibility for self-learning (Upadhya and Lynch, 2019). Student-centered strategies include techniques such as active learning, problem-solving through critical and creative thinking, role play, and group learning such as cooperative learning. Through this teaching orientation, students are given the opportunity to build deep knowledge and understanding of the learning contents, thus inculcating a positive attitude in the learning process (Alam, 2016). Hence, it is evident that student-centered teaching styles should not be ignored but require serious attention and the proactive steps of teachers to balance them with student learning, especially in today's era of rapid and challenging development of educational transformation.

The findings for the second research objective showed a very weak and significant positive correlation between the Grasha–Riechmann Teaching Styles of Mathematics teachers with their Mathematics experience teaching. These findings confirm the relationship between the two variables presented in the conceptual framework of the study.

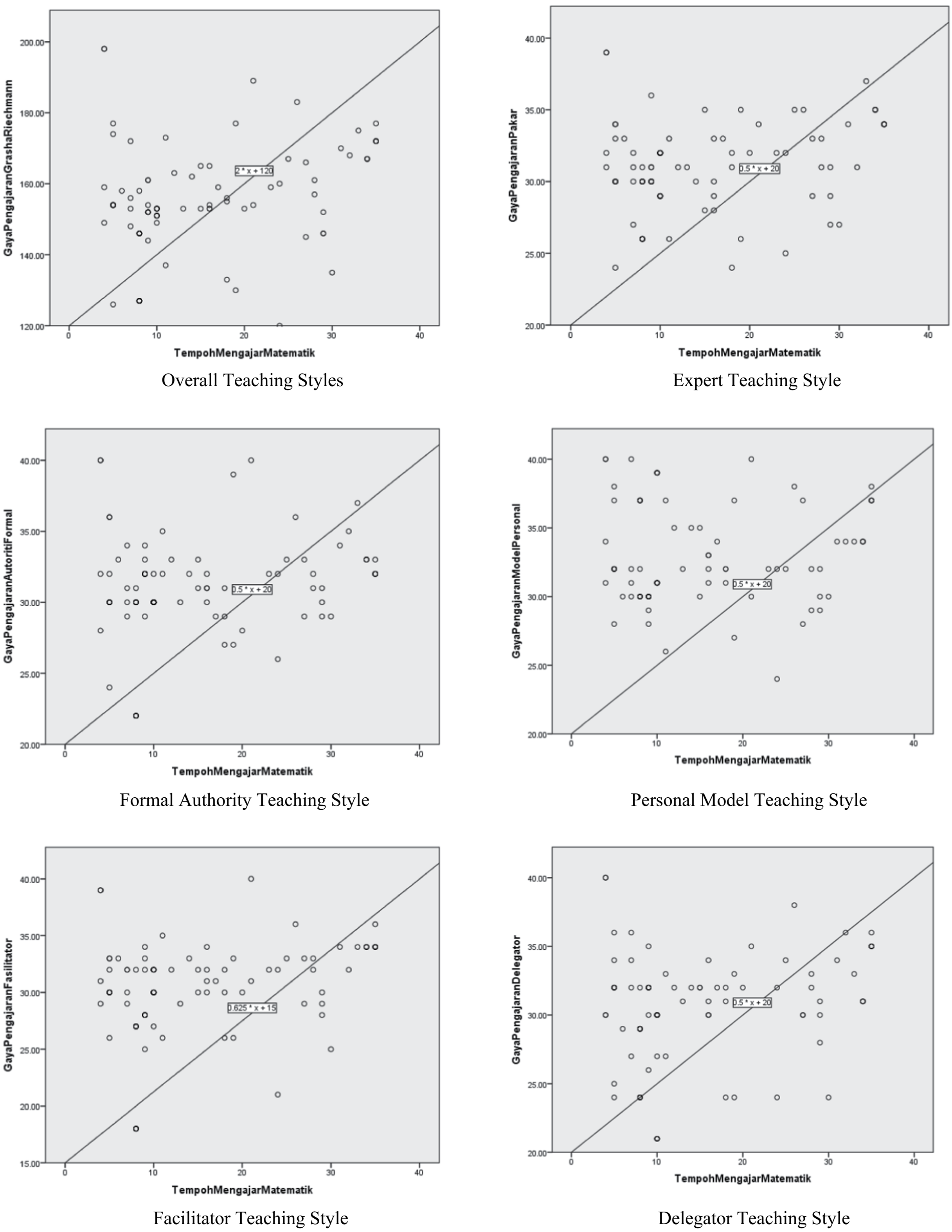


FIGURE 1
Scatter plots for the correlation between Mathematics teaching experience and Grasha–Riechmann teaching styles.

Based on the context of the five constructs of the Grasha–Riechmann Teaching Styles, the findings showed significant relationships among the Expert Teaching Style, Formal Authority Teaching Style, and Facilitator Teaching Style of teachers with their Mathematics teaching experience. However, there was no significant relationship involving the Personal Model Teaching Style and Delegator Teaching Style with the teachers' Mathematics teaching experience. The findings directly provide an interpretation that the Grasha–Riechmann Teaching Styles of Mathematics teachers have a significant relationship with their Mathematics teaching experience, specifically for the Expert Teaching Style, Formal Authority Teaching Style, and Facilitator Teaching Style.

The findings of this study are consistent with past studies that examined the relationship between the Grasha–Riechmann Teaching Styles of teachers and their teaching experience. More interestingly, the context of this study that involves Mathematics teachers has similar results to past studies involving teachers in different courses, which have shown a significant relationship between the Grasha–Riechmann Teaching Styles of English teachers in public and private sectors in Iran with their teaching experience (Piaget, 1936; Hosseini Fatemi and Raoufi, 2014; Mazaheri and Ayatollahi, 2019). In different courses, a study by Kothari and Pingle (2015) on administrative instructors from various schools throughout India as well as a study by Sabado and Allan (2019) on Technical Vocational Education (TVE) teachers in the Philippines also found similar findings to the current study, in which the Grasha–Riechmann Teaching Styles of teachers had a significant relationship with their teaching experience. This proves that the Grasha–Riechmann Teaching Styles do not only have a significant relationship with the teaching experience of Mathematics teachers but also those specializing in other areas.

Theoretically, teachers' teaching experience indeed had a statistically significant influence on the creativity, classroom management, and pedagogical skills of the teachers (Dewaele et al., 2018). More experienced teachers in the profession are more creative in the classroom, more skilled at managing classroom activities, and have stronger pedagogical skills. Hosseini Fatemi and Raoufi (2014) confirmed significant differences in the teaching styles between less experienced teachers (1–10 years of teaching experience) and more experienced teachers (over 15 years of teaching experience), where more experienced teachers had higher mean scores through the application of the Expert Teaching Style, Formal Authority Teaching Style, and Facilitator Teaching Style. More experienced teachers were also found to have a higher level of knowledge, expertise, and mastery of course materials in order to be able to act as experts who display comprehensive knowledge and deliver information effectively (Sabado and Allan, 2019).

In other words, more experienced teachers have accurate and comprehensive knowledge, skills, and information on the target scope to be taught to students, in line with the requirements to apply the Expert Teaching Style. Hence, this phenomenon is seen as a contributor that may lead to a significant positive correlation between the teaching styles of Mathematics teachers and their Mathematics teaching experience.

Based on a study by Bruno et al. (2019) that used administrative data for 10 years in Los Angeles, teachers with low teaching experience tended to face problems, especially in the aspect of teacher pedagogical skills. Compared to experienced teachers, inexperienced teachers faced difficulties in classroom management and in controlling student behavior during the Learning and Facilitation (PdPc) process. On the other hand, experienced teachers were more likely to engage in standard classroom management efforts and were able to create concrete learning situations by setting learning objectives for students (Faruji, 2012). Besides, more experienced teachers were also more concerned with accepted, accurate, and standardized ways of doing things as prioritized in the Formal Authority Teaching Style.

Past research has also shown that experienced teachers have a better understanding of student needs and are able to explore options to meet these needs of students (Sabado and Allan, 2019). Evidently, a deep understanding of the various needs of students helps experienced teachers build close interpersonal relationships with students (Rahimi and Asadollahi, 2021). Indirectly, experienced teachers can interact well with their students as well as guide and direct the students by asking questions, exploring options, and suggesting alternatives, in line with the characteristics of the Facilitator Teaching Style outlined by Grasha (2002). Accordingly, experienced teachers can become good facilitators compared to inexperienced teachers.

This could explain the findings in the context of this study, which illustrates a significant positive correlation relationship between the Facilitator Teaching Style of Mathematics teachers and their Mathematics teaching experience. However, a gap exists when the relationship between teachers' teaching experience and the Grasha–Riechmann Teaching Styles of primary school Mathematics teachers has not been widely tested, regardless of the Malaysian educational context or abroad. Therefore, the findings require further studies in the future for a more robust confirmation through empirical evidence.

Implication

The findings of this study provide theoretical implications by strengthening and expanding the concept of the Grasha–Riechmann Teaching Styles by proving Personal Model Teaching Style as the dominant teaching style compared to other teaching styles in the context of primary school Mathematics teachers, which is in line with Grasha (1996) initial study. In addition, implications for the body of knowledge also exist through new findings in terms of the differences in Grasha–Riechmann Teaching Styles based on gender as well as the relationship between teaching styles and teaching experience in the context of primary school Mathematics teachers. The added value is clearly obtained through a specific analysis of each Grasha–Riechmann Teaching Style. These findings have indirectly added value to existing knowledge and a deeper understanding of the Grasha–Riechmann Teaching Styles practiced by teachers.

Practical implications also exist as the findings of this study provide early information to educators in diversifying teaching styles and improving the learning needs of different students in the classroom. In facing the challenges of Industrial Revolution 4.0, the rapid transformation of education has demanded teacher initiation in transforming the traditional teacher-centric teaching method into a more student-centric teaching approach. In student-centered teaching orientation, students are given more attention and responsibility for self-learning. The main responsibility of a teacher is to build and maintain a conducive learning environment where students are encouraged to build their own knowledge while the teacher acts as a facilitator and guide. In this regard, teachers are encouraged to adopt student-centered strategies that include techniques such as active learning, problem-solving by engaging in critical and creative thinking, role play, and group or cooperative learning. Indirectly, students are able to build meaningful relationships among existing knowledge, new knowledge, and the processes involved in learning, which are in line with the demands of 21st Century Learning (PAK21).

Likewise, implications exist in the pedagogical practices of different teachers. In the context of this study, the Ministry of Education can enhance the pedagogical skills of novice teachers by formulating a more comprehensive, appropriate, and practical strategy to improve professionalism among teachers in Malaysia. Seminars on knowledge, skills, and professional practices such as Content and Methods of Pedagogical Subjects, Creativity, and Pedagogical Innovation as well as Instructional Leadership courses can be implemented to improve the quality of teaching styles among teachers.

Conclusion

This study has revealed that primary school Mathematics teachers preferred the Personal Model Teaching Style, whereas the Facilitator Teaching Style was less popular among primary school Mathematics teachers. The findings also showed a very weak and significant positive correlation between Grasha–Riechmann Teaching Styles and teaching experience, which encompass the Expert Teaching Style, Formal Authority Teaching Style, and Facilitator Teaching Style. Besides, based on the results of this study, some things need to be taken seriously because the current study only covers a factor as the independent variable, namely teaching experience, which fall into the demographic category. However, past studies have highlighted that the Grasha–Riechmann Teaching Styles of teachers are not influenced by one aspect or factor alone. Therefore, more in-depth studies are needed because there is a possibility of other influential factors that have yet to be explored.

In this regard, further studies can be developed by examining other demographic factors such as school type, age, subject flow, professional qualifications, and socioeconomic status of teachers. The Grasha–Riechmann Teaching Styles can also be attributed to other factors such as emotion, self-efficacy, personality traits, creativity, attitudes, thinking style, and autonomy of teachers as well as student-related factors such as the level of reasoning, learning style, academic involvement, student interest or motivation, and the number of students in the classroom. Besides, the study also has delimitations that can be improved in the future. As the respondents in this study only involve Mathematics teachers from SJKC Kepong 1, Kepong 2, and Kepong 3 Sentul Zone, Kuala Lumpur. For future research, it was suggested to use a larger sample involving more types of schools. It may also consider a comparison between rural and urban schools as well as between regular and high-performance schools to look at the differences in Grasha–Riechmann Teaching Styles. A wider and larger population can provide a more comprehensive and detailed picture of the Grasha–Riechmann Teaching Styles of teachers. This study is also limited to primary school Mathematics teachers only; hence, future studies on the Grasha–Riechmann Teaching Styles should involve various subjects and other levels of study. In terms of methodology, the data were based on questionnaires only; thus, it is better to combine both qualitative and quantitative methods to obtain more comprehensive and meaningful findings to improve the teaching styles of teachers.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Education Planning and Research Division, Ministry of Education Malaysia and Faculty of Education, UKM. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

SS and MM: conceptualization, validation, resources, and data curation. SS: methodology, software, formal analysis, investigation, writing—original draft preparation, and project administration. MM: writing—review and editing, visualization, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Faculty of Education, Universiti Kebangsaan Malaysia (UKM) through the grant of Ganjaran Penyelidikan (GP-2021-K021854).

Acknowledgments

Special thanks and appreciation of gratitude to my supervisor, MM, the Faculty of Education, UKM, and Ministry of Education (MOE) for giving us the golden opportunity to conduct this wonderful project and guiding us from the beginning to the end, including each of the participants for their support during the data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdull Sukor, S., Nurahimah, M. Y., Izam, M. G., Rafisah, H. O., and Nur Fatirah, M. D. (2014). The relationship between lecturers' teaching style and students' academic engagement. *Procedia. Soc. Behav. Sci.* 118, 10–20. doi: 10.1016/j.sbspro.2014.02.002

Ainonmadiah, M. N., Awanis, M., Norhana, M. S., Juliana, B., and Siti Noor, I. (2016). Hubungan antara gaya pengajaran guru dengan tahap ponteng sekolah menengah di Daerah Bachok, Kelantan. In *Proceeding of International Consortium of Education and Culture Research Studies (ICECRS)*. Kedah, Malaysia: Universiti Utara Malaysia. 1; pp. 341–358. Available at http://dx.doi.org/10.21070/picecrs.v1i1.503

Alam, M. (2016). Constructivism: paradigm shift from teacher centered to student centered approach. *Int. J. Indian Psychol.* 4, 51–59. doi: 10.1051/shsconf/20184200038

Alami, M., and Ivaturi, P. (2016). Typical teaching styles among ELC lecturers at Salalah college of technology, Oman. *J. Appl. Linguist. Lang. Res.* 3, 251–261.

Anis Humaira, M. M., Nurhazwani, M. S., and Hafsah, H. (2019). Gaya pengajaran Grasha dalam kalangan pensyarah sains di Kolej Pra-Universiti di Pulau Pinang. *Jurnal Pendidikan Sains Dan Matematik Malaysia* 9, 16–24. doi: 10.37134/jpsmm.vol9.2.3.2019

Astuti, N. P. E. (2018). Teacher's instructional behaviour in instructional management at elementary school reviewed from Piaget's cognitive development theory. *J. Math. Behav.* 52, 138–153.

Baleghizadeh, S., and Shakouri, M. (2017). Investigating the relationship between teaching styles and teacher self-efficacy among some Iranian ESP university instructors. *Innov. Educ. Teach. Int.* 54, 394–402. doi: 10.1080/14703297.2015.1087329

Bandura, A. (1962). *"Social learning through imitation," Nebraska Symposium on Motivation, ed. M. R. Jones.* Lincoln: University of Nebraska Press, 211–274.

Baradaran, A. (2016). The relationship between teaching styles and autonomy among Iranian female EFL teachers teaching at advanced levels. *Engl. Lang. Teach.* 9:223. doi: 10.5539/elt.v9n3p223

Beers, C. (2016). *Early Childhood Preservice Teachers' Knowledge of Children's Cognitive Development and Developmentally Appropriate Pedagogical Practices: Understanding the Role of Clinical Experiences*. Florida Atlantic University Boca Raton, Florida.

Bennett, N., Jordan, J., Long, G., and Wade, B. (1976). *Teaching Styles and Pupil Progress*. Open Books Publishing Limited: London.

Bergil, A. S., and Erçevik, A. (2019). The prospective EFL teachers' impressions towards teaching styles: foresights for their professions. *J. Lang. Linguist. Stud.* 15, 1236–1251. doi: 10.17263/jlls.668394

Bruno, P., Rabovsky, S. J., and Strunk, K. O. (2019). Taking their first steps: the distribution of new teachers in school and classroom contexts and implications for teacher effectiveness. *Am. Educ. Res. J.* 15, 75–80. doi: 10.3102/00028312198820

Chua, Y. P. (2021). *Kaedah Dan Statistik Penyelidikan Buku 2: Asas Statistik Penyelidikan, 4th Edn*. McGraw-Hill Education (Malaysia) Sdn Bhd: Selangor.

Dewaele, J. M., Gkonou, C., and Mercer, S. (2018). Do ESL/EFL teachers' emotional intelligence, teaching experience, proficiency and gender affect their classroom practice? *Emot. Sec. Lang. Teach.* 18, 125–141. doi: 10.1007/978-3-319-75438-3_8

Dole, S., Bloom, L., and Kowalske, K. (2016). Transforming pedagogy: changing perspectives from teacher-centered to learner-centered. *Interdisciplin. J. Probl. Based Learn.* 10, 1–15. doi: 10.7771/1541-5015.1538

Donkor, A. (2019). In-service teachers' use of ICT in teaching mathematics in Ghana. A case study in the Cape Coast Metropolis. Doctorate Thesis. University of Cape Coast.

Faruji, L. F. (2012). Teachers' teaching styles at English language institutes in Iran. *Int. J. Soc. Sci. Educ.* 2, 364–373.

Fernandez-Rivas, M., and Espada-Mateos, M. (2019). The knowledge, continuing education and use of teaching styles in physical education teachers. *J. Hum. Sport Exerc.* 14, 99–111. doi: 10.14198/jhse.2019.141.08

Flanders, N. A. (1970). *Analyzing Teacher Behavior*. Addison Wesley: Reading, MA.

Ghanizadeh, A., and Jahedizadeh, S. (2016). EFL teachers' teaching style, creativity, and burnout: a path analysis approach. *Cogent Educ.* 3, 115–127. doi: 10.1080/2331186X.2016.1151997

Goff, K. G. (2016). Investigation of learner-centered instruction and teacher-centered instruction in a high school wind band class. Doctorate Thesis. Florida: University of Florida State.

Grasha, A. F. (1994). A matter of style: the teacher as an expert, formal authority, personal model, facilitator and delegator. *Coll. Teach.* 42, 142–149. doi: 10.1080/87567555.1994.9926845

Grasha, A. F. (1996). *Teaching with styles: A practical guide to enhance learning by understanding learning and teaching styles*; Alliance Publisher: New York.

Grasha, A. F. (2002). *Teaching with style: A practical guide to enhancing learning by understanding teaching and learning styles*; Alliance Publisher: San Bernadino, CA.

Heydarnejad, T., Hosseini Fatemi, A., and Ghonsooly, B. (2017). An exploration of EFL teachers' teaching styles and emotions. *J. Appl. Linguist. Lang. Res.* 4, 26–46.

Hosseini Fatemi, A., and Raoufi, R. (2014). Burnout and teaching style among Iranian English language educators in public schools and private institutes: a cross-comparison study. *Int. J. Res. Stud. Lang. Learn.* 1, 85–94. doi: 10.5861/ijrsll.2014.557

Karimnia, A., and Mohammdi, N. (2019). The effects of teachers' gender, teaching experience, and brain dominance on their teaching styles. *Int. J. Res. Engl. Educ.* 4, 37–46. doi: 10.29252/ijree.4.1.37

Kazemi, A., and Soleimani, N. (2016). On the relationship between EFL teachers' classroom management approaches and the dominant teaching style: a mixed method study. *Iran. J. Lang. Teach. Res.* 4, 87–103.

Khalid, M., Akhter, M., and Hashmi, A. (2017). Teaching styles of secondary school english teachers and learning styles of their students and relationship of teaching learning style match with students' achievement. *Bull. Educ. Res.* 39, 203–220.

Kothari, T. P., and Pingle, S. S. (2015). Personality traits and teaching style of management teachers: an empirical study. *J. Contempor. Manage. Res.* 9, 16–38.

Lak, M., Soleimani, H., and Parvaneh, F. (2017). The effect of teacher-centeredness method vs. learner-centeredness method on reading comprehension among Iranian EFL learners. Journal of advances in English. *Lang. Teach.* 5, 1–10.

Lowman, J. (1995). *Mastering the techniques of teaching, 2nd Edtn.*, Jossey-Bass: San Francisco.

Lutz, C., Berges, M., Hafemann, J., and Sticha, C. (2018). Piaget's cognitive development in bebras tasks: a descriptive analysis by age groups. Proceeding of international conference on informatics in schools: Situation, evolution, and perspectives. Springer, Cham. October 2018; pp. 259–270.

Mailizar, M. (2018). Investigating Indonesian teachers' knowledge in mathematics teaching. Doctorate Thesis. Faculty of Social, Human and Mathematical Sciences, University of Southampton.

Martin, J. E. (2019). Perceived teaching style and academic growth in an international school setting. Doctorate Thesis. University of Western Kentucky.

Massaada, M. (2016). The analysis of English teachers 'teaching styles and their effects on students 'motivation at SMA Negeri 2 Majene. Doctorate Thesis. Universiti Negeri Makassar.

Mazaheri, S., and Ayatollahi, M. A. (2019). Relationship between Iranian EFL teachers' brain dominance, teaching experience and their teaching style. *Int. J. Foreign Lang. Teach. Res.* 7, 115–126.

Mazeni, I., and Hasmadi, H. (2017). Gaya pengajaran dalam kalangan guru tabika Kemas di Kelantan. *Jurnal Pendidikan Awal Kanak-kanak Kebangsaan.* 6, 1–15.

Mazlini, A., Afian, A. M., Abdul Halim, A., Yuanita, P., Muhamad Izzuddin, A. B., and Siti Mistima, M. (2021). MTeaching practices of the mathematics male and female teachers according to the PISA framework and its relation to their beliefs towards their students mathematics trainer readiness for PAK-21 integration in mathematics teaching and learning. *Rev. Int. Geograph. Educ.* 11, 891–896. doi: 10.33403/rigeo. 8006803

Mohd Effendi Ewan, M. M., Mohammed Afandi, Z., Muhamad Firdaus, M. N., Ahmad Zamri, K., and Nordin, A. R. (2021). The development and psychometric assessment of Malaysian youth adversity quotient instrument (MY-AQi) by combining Rasch model and confirmatory factor analysis. *IEEE Access.* 9, 13314–13329. doi: 10.1109/ACCESS.2021.3050311

Normiati, B., and Abdul, S. A. (2019). The influence of teacher's teaching towards quality of teachers' teaching. *Malays. J. Soc. Sci. Humanit.* 4, 30–42.

Nouraey, P., and Karimnia, A. (2016). The impact of rhythm and gender on spelling output of Iranian EFL learners: a cross-sectional study. *Khazar J. Humanit. Soc. Sci.* 19, 53–63. doi: 10.5782/2223-2621.2016.19.2.53

Nur Farhah, M. S., and Fatimah Wati, H. (2018). Pengaruh personaliti lima faktor terhadap tingkah laku kewargaan organisasi (TKO) dalam kalangan guru. *Jurnal Psikologi Malaysia.* 32, 1–11.

Nur Liyana, M. I., and Zakiah, M. A. (2017). Relationship between motivation and teachers' teaching style among secondary school students' in Kulai. *Man India* 97, 299–307.

Piaget, J. (1936). *The Origins of Intelligence in Children*. International Universities Press: New York.

Rahimi, M., and Asadollahi, F. (2021). Teaching styles of Iranian EFL teachers: do gender, age, and experience make a difference? *Int. J. Engl. Linguist.* 2, 157–164. doi: 10.5539/ijel.v2n2p157

Rosalia, H. (2017). The analysis of English teachers' teaching styles and their effects on students' interest. Doctorate Thesis. University of Malaya, Kuala Lumpur.

Rumainah, I. I., and Faridah, M. K. (2017). Dorongan autonomi guru dalam pembentukan motivasi intrinsik murid. In *Proceedings of the Seminar on Transdisiplin Education (STEd2017)*. Bangi, Selangor: Universiti Kebangsaan Malaysia. pp. 308–314.

Sabado, R. R., and Allan, O. (2019). Junior and senior technical vocational education teachers' performance and teaching styles: a comparison. *Int. J. Latest Res. Humanit. Soc. Sci.* 2, 32–42.

Salhani, I. (2019). Jumlah pelajar mengambil STEM kian merosot. *Berita Harian* 12:18.

Seng, C. L., Sii, N. Y., Isawasan, P., Chen, K. L., and Seng, P. L. (2016). Factors influencing teachers' intention to adopt ICT into teaching using partial least square technique methods. *AIP Conf. Proceed.* 020076. doi: 10.1063/1.5055478

Solomon, D., and Kendall, A. J. (1979). *Children in classrooms: An investigation of person-environment interaction*, Prager Publishers: New York.

Suik Fern, K., and Mohd Effendi Ewan, M. M. (2022). Can a science, technology, engineering, and mathematics (STEM) approach enhance students' mathematics performance? *Sustainability* 14, 379–384. doi: 10.3390/su14010379

Tavakoli, M., and Karimnia, A. (2017). Dominant and gender-specific tendencies in the use of discourse markers: insights from EFL learners. *World J. Engl. Lang.* 7, 1–9. doi: 10.5430/wjel.v7n2p1

Upadhya, P., and Lynch, R. (2019). A comparative study of student motivation and academic achievement in grade 8 science under teacher-centered and student-centered instructional methods at Triamudomsuksa Pattanakarn school, Thailand. *Schol. Hum. Sci.* 11, 108–119.

Wan Faizal, I. (2019). Kerajaan risau penurunan pelajar pilih aliran STEM. *Berita Harian* 27:16.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fpsyg.2022.1042838

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University,
China

REVIEWED BY
Qi Ye,
South China Normal University, China
J. Roberto Sanz Ponce,
Catholic University of Valencia San Vicente Mártir, Spain

*CORRESPONDENCE
Yueyuan Kang
kyymail@163.com
Yiming Zhen
zym961224@163.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 13 September 2022
ACCEPTED 10 October 2022
PUBLISHED 28 October 2022

CITATION
Wang G, Kang Y, Li F, Zhen Y, Chen X and Huang H (2022) An evidence-based study on the current status of Chinese secondary school mathematics teachers' autonomous learning capacity across demographic and contextual factors.
Front. Psychol. 13:1042838.
doi: 10.3389/fpsyg.2022.1042838

An evidence-based study on the current status of Chinese secondary school mathematics teachers' autonomous learning capacity across demographic and contextual factors

Guangming Wang[1], Yueyuan Kang[1]*, Fengxian Li[2], Yiming Zhen[1]*, Xia Chen[1] and Huixuan Huang[1]

[1]Faculty of Education, Tianjin Normal University, Tianjin, China, [2]Department of Education, Inner Mongolia Honder University of Arts and Science, Hohhot, China

Autonomous learning capacity is a key competency that supports teachers' professional development. In this study, a stratified sampling method was used to recruit 396 junior and senior high school mathematics teachers in T city, one of the provincial city in China. A questionnaire with high reliability and validity developed prior to the study by the researchers was employed to measure their autonomous learning capacity and differences across groups. Twelve teachers were then selected for interviews. The results showed satisfactory overall performance. By subdimension, teachers' performance was the best in the development of study plans, followed by evaluation of learning outcomes, while they needed improvement in learning habit formation and proficiency in using learning methods. Furthermore, the analysis of differences across groups indicated that for autonomous learning capacity, female teachers were significantly better than their male peers; junior high school teachers were better compared to those in senior high school; teachers aged 41–50 underperformed those aged 51 and above; teachers who work in rural areas and townships did not perform worse than urban teachers; and those with doctoral degrees did not demonstrate considerable advantage over others. There were no significant differences in the overall autonomous learning capacity across years of teaching and job title groups. However, in the subdimensions, those with 21–30years of teaching experience had lower proficiency in using learning methods and evaluation of learning outcomes, and teachers with senior titles did not demonstrate expected advantages in learning habit formation.

KEYWORDS

mathematics teachers, Chinese teachers, autonomous learning, autonomous learning ability, professional development, teacher learning

Introduction

With the demand for highly qualified talent being stronger than ever before, people have become more dependent on learning for personal development. Some believe that "learning" can be autonomous. According to Holec (1979), autonomous learners are those who take responsibility for various decisions throughout the learning process. Individuals cannot learn and develop without external intervention and guidance, including the education provided by teachers. In assessing the professional competencies of teachers, autonomous learning capacity is one of the key indicators (Kang et al., 2021). As learners, teachers need to regard the development of autonomous learning capacity as a lifelong, continuous, and dynamic process (Smith, 2003), forming a mutually reinforcing positive cycle with lifelong learning (Yurdakul, 2017).

Such capacity is important because of two reasons. For students' learning outcomes, the autonomous learning of teachers contributes to their professional development (Olivier and Wittmann, 2019), and a high level of professional competency is associated with quality teaching (Harris and Sass, 2011; Creemers et al., 2012), which is a predictor of students' academic achievements (e.g., Sanders et al., 1997; Kunter et al., 2013; Kyriakides et al., 2013). For teachers' career prospects, autonomous learning capacity and teaching practices are mutually reinforcing. In other words, autonomous learning capacity is both the enabler for and outcome of successful teaching practices. For example, teachers who teach effectively tend to have a higher level of self-efficacy (Holzberger et al., 2013), which motivates them to strive for further professional development and self-improvement (Posnanski, 2002). Meanwhile, quality teaching is often associated with teachers' characteristics such as personality and educational background (Bowles et al., 2014). Moreover, Certo and Fox (2002) argued that teacher autonomy has a positive impact on perceptions of and attitudes toward teaching. Autonomous teachers are more dedicated to their jobs.

However, studies on mathematics teachers' autonomous learning capacity, especially on secondary school mathematics teachers, have not been given adequate attention. To a certain extent, assessing teachers' competencies is also needed to answer the question of why Chinese students can excel consistently in international mathematics tests, such as the Program for International Student Assessment. Research efforts on Chinese mathematics teachers can initiate strategies to improve teacher competencies. Therefore, this study focused on the autonomous learning capacity of secondary school mathematics teachers in China to assess their competencies and analyze their differences by teacher group.

Literature review

Teachers' autonomous learning capacity

Initiative is a key characteristic of autonomous learning (Stockdale and Brockett, 2011). Autonomous learning is the ability to take control of learning (Holec, 1979). Autonomous learners need to proactively evaluate their learning needs, set learning goals, select and implement appropriate learning strategies by leveraging humanistic and material learning resources, and assess learning outcomes with external assistance or under self-motivation (Loeng, 2020). Commenting on learning autonomy, Mynard and Sorflaten (2006) stated that autonomous learners are often confident. They know their strengths and weaknesses, autonomously make learning decisions, pace learning according to actual situations, plan and set their goals, and assess their learning process and progress. Similarly, Oxford (2003) indicated that autonomous learners are highly motivated and self-efficacious and these internal and external motivations can play an active role in helping them achieve better learning outcomes.

For teachers as autonomous learners, in addition to the general characteristics, there are some teacher-specific attributes. First, teachers' autonomous learning capacity is based on experiences, implying that autonomous learning teachers are proactive, autonomous, and self-directed. Their learning is practice- and problem-oriented and aims at maintaining self-esteem and satisfying needs. In the process of autonomously choosing and integrating new and old knowledge, these teachers fully demonstrate their values (Goodlad, 1990). Moreover, autonomy is a necessary condition for teachers to be creative (Anderson, 2002). Second, teachers' autonomous learning capacity is derived from addressing questions, suggesting that autonomous learning teachers are reflective and creative. From this perspective, Billett (2002) proposed that teacher learning is a process by which teachers develop skills and acquire knowledge and expertise through reflection and action and that self-regulation of a reflective nature contributes to the development of their autonomy (Papamitsiou and Economides, 2019). Finally, teachers' autonomous learning capacity is a part of their daily professional competencies. For example, Kelly (2006) defined teacher learning as the process by which teachers aim to gain expertise and that the process of autonomous learning continues to stimulate lifelong learning (Yurdakul, 2017).

In summary, teachers' autonomous learning capacity involves having autonomy; making plans; taking initiatives; serving professional development needs; and going through the process of autonomous planning, evaluation, and improvement. Accordingly, in this study, teachers' autonomous learning capacity is defined as teachers' capabilities to develop and implement their study plans; adopt appropriate means and methods; make full use of time, space, and other resources; and participate in professional activities such as teacher education, educational teaching, and pedagogical research for continuously improving their knowledge and teaching skills.

Smith and Erdogan (2008) explained that such capabilities can be classified into two dimensions: self-directed development and freedom from control. Pintrich (2004) defined autonomous learning as a proactive and constructive learning process and divided the process into three parts: goal setting and planning, execution and action adjustment, and reflection and cognitive monitoring. It is evident that autonomous learning is a means to tackle shortcomings

TABLE 1 Subdimensions of teachers' autonomous learning capacity and behavioral requirements.

	Subdimensions	Key behavioral requirements
Teachers' autonomous learning capacity	Development of study plans	Teachers are able to understand the requirements of professional standards, undergo self-reflection, and adopt suggestions from colleagues, thus making study plans that address their professional learning needs and setting clear learning goals for the expected outcomes of their professional learning.
	Proficiency in using learning methods	Teachers are able to integrate the following learning activities: reading books and journals, browsing professional websites, observing public classes, watching videos of quality classes, attending special lectures, taking online courses, and taking learning trips.
	Learning habit formation	Teachers are able to develop the habit of taking initiative in learning and form the habit of lifelong learning over time.
	Evaluation of learning outcomes	Teachers are able to evaluate their learning process comprehensively through self-evaluation, learning summaries, scales, and surveys, covering the measures of the initiative and continuity of learning, the effectiveness in applying learning methods and strategies, problems that arise in the learning process, and learning outcomes.

and gain insights about teaching through reflection. From a different perspective Hai and Liu (2018), proposed three analytical dimensions of teacher autonomous learning based on the social cognitive theory: learning motivation, learning strategies and self-monitoring, and argued that teachers' autonomous learning is a process of professional development based on daily work situations and existing knowledge and experience, in which they consciously and actively use various effective methods and self-regulate to ultimately improve their professionalism and effectiveness. In contrast, Borko (2004) proposed three analytical perspectives on teacher learning based on the contextual learning theory, combined with the views about individual and social cognition. First, the focus is on individual teachers, including their learning activities and mental changes. Second, the focus is on teacher groups, or the communities of practice they belong to and are embedded in during the course of their learning activities. Third, both individual teachers and their groups are considered to examine the impact of individual-environment interactions on teacher learning. From the perspectives of learning behaviors and teacher-specific attributes, teacher learning can be defined as a proactive, autonomous, and self-directed learning process while participating in teacher education, educational teaching, and pedagogical research for professional development.

Sebotsa et al. (2019) suggested that teachers' professional development may be difficult to sustain without autonomous learning. They recommended including the dimension of autonomous learning to assess teachers' capabilities in identifying learning needs, setting learning goals, using humanistic and material resources for learning, selecting and implementing appropriate learning strategies, and assessing learning. A similar concern was shared in a discussion about teacher professional development. For example, the National Social Science Fund of China's 2017 Key Educational Bidding Project "Research on Teachers' Key Literacy and Competencies" constructed a double-helix model of teachers' key literacy and competencies and developed a questionnaire based on this model (Wang et al., 2019; Kang et al., 2021). It included teachers' autonomous learning capacity as one of the competencies and measured it in four subdimensions: development of study plans, proficiency in using learning methods, learning habit formation, and evaluation of learning outcomes. Each subdimension was accompanied by a description of key behavioral requirements for clarity (see Table 1). These subdimensions comprehensively measure competencies mostly related to teachers' professional capacity, covering the whole process of instructional preparation, implementation, and evaluation. Additionally, the questionnaire was developed in the context of China, where this study was conducted. Therefore, these subdimensions were used in this study.

Impact of demographic and contextual factors on teachers' autonomous learning capacity

Learning is influenced by three types of factors: personal factors, environmental factors, and behaviors. These factors affect each other reciprocally in a continuous cycle (Bandura, 1962). Among them, the environment can be social or physical, and thus, classified under external factors; while personal factors and behaviors can be categorized under internal factors. For external factors, teacher learning and development take place in professional communities (Shulman and Shulman, 2009; De Jong et al., 2019), and an empirical study revealed that a school's learning beliefs, learning support systems, and learning communities impact teachers' professional learning (Opfer et al., 2011). Moreover, a study of primary and secondary school teachers' motivation for professional development found that external contextual factors such as educational stage, location, and years of teaching experience have significant effects on teachers' motivation for learning and development (Qi et al., 2020). In addition, in a longitudinal study, Day and Gu (2007) found that given the humanistic aspects of teachers' work, not only the external factors across different stages of professional development but also internal conditions can affect their professional learning. For example, personal changes and changes in a work environment can have different effects on professional learning at different stages of career development.

Regarding internal factors, teacher competencies have been graded from the perspective of individual knowledge building (Shulman, 2011) or an individual's positioning within a community (De Jong et al., 2019). Subsequent studies have addressed differences at the individual level. For example, Abdel Razeq (2014) found significant differences in teachers' autonomous learning capacity by gender, with female teachers being better at autonomous learning. Deregözü and Hatipoğlu (2018) identified no differences in German pre-service teachers' autonomous learning capacity by grade level and educational background but revealed statistical differences between them by age and gender. Using data from the 2013 Teaching and Learning International Survey, Chen (2017) added the factor of years of teaching experience and empirically examined teachers' professional development in 36 jurisdictions. The results showed that contextual factors, such as teachers' gender, years of teaching experience, educational background, and professional development needs, have significant effects on teachers' professional learning and development. Besides, a survey of kindergarten teachers in China conducted by Hai and Liu (2018) found significant differences in autonomous learning among teachers with different years of teaching experience, educational background, professional backgrounds, and job titles. The autonomous learning of teachers is determined to a larger extent by individual internal factors than by external factors such as school and school location.

Summarily, teachers' autonomous learning capacity is affected by a variety of internal and external conditions, including gender, age, years of teaching experience, educational background, job title, educational stage, and location. However, the majority of these studies have failed to take into consideration teachers' educational stages or the subject they teach, let alone focus on secondary school mathematics teachers. Moreover, the resultant recommendations for improvement are mostly generic, rather than specific enough for a teacher to incorporate them in their teaching practices for a particular subject. To address the issue, personal factors and external school factors identified in prior studies on differences among teacher groups can be repurposed as demographic and contextual factors. Therefore, this study aimed to examine the following two research questions:

> RQ1: What is the overall status of the autonomous learning capacity of secondary school mathematics teachers?
>
> RQ2: Are there differences in autonomous learning capacity among teacher groups across the contextual factors of gender, age, years of teaching experience, educational background, job title, educational stage, and location? If yes, what are the differences?

Materials and methods

Participants

T city is a provincial city in northern China with good education and economic development. To balance the level of teachers in different schools and to enhance the representativeness of the sample, a stratified sampling method was used to select 430 secondary school mathematics teachers from 8 out of 16 districts of T city as anonymous survey respondents and face-to-face questionnaire completion. Among them, 73 were males and 319 were females. They were aged between 31 and 50 years old, with largely 11–30 years of teaching experience. A majority of them have completed undergraduate education. Most of them work in urban areas. Most of the job titles were intermediate and senior ones (see Table 2 for detailed information about the respondents). Among them, the job title is the result of the evaluation of the level of primary and secondary school teachers by the regional education bureaus in China according to the Evaluation Standards for the Professional and Technical Levels of Primary and Secondary School Teachers in China. In the test, we divided the job titles into three categories, namely not rated or primary, intermediate and senior, according to the level of teachers in T city. A total of 430 questionnaires were distributed. Using polygraph questions, 34 invalid questionnaires with inconsistent responses in the teacher autonomy questionnaire (the second part of the questionnaire) were eliminated, 396 valid questionnaires were obtained, with a return rate of 92.09%.

For a more robust interpretation of the survey results, purposive sampling was used, thus providing more information for this study (Chen, 2000). We selected 12 mathematics teachers and conducted semi-structured interviews using a combination of individual interviews, focus group interviews, phone calls, and online interviews. Moreover, the purposive sampling covered teachers from different groups to ensure representativeness. Specifically, among the 12 mathematics

TABLE 2 Basic information of teachers surveyed.

Demographic variables	Categories	*N*	Percentage (%)
Gender	Male	73	18.39
	Female	319	80.35
	Not available	4	1.26
Age (in years)	30 and below	55	13.89
	31–40	123	31.06
	41–50	153	38.64
	51 and above	64	16.16
	Not available	1	0.25
Years of teaching experience	10 and below	81	20.45
	11–20	143	36.11
	21–30	128	32.32
	31 and above	41	10.35
	Not available	3	0.76
Education	Associate and below	9	2.27
	Bachelor	298	75.25
	Master	83	20.96
	Doctor	5	1.26
	Not available	1	0.25
Job title	Not rated or primary	47	11.87
	Intermediate	183	46.21
	Senior	165	41.67
	Not available	1	0.25
Educational stage	Junior high school	243	61.36
	Senior high school	153	38.64
Location	Urban area	243	61.36
	Township	123	31.06
	Rural area	21	5.30
	Not available	9	2.27

"Not available" indicates an invalid response by the teacher in the first part of the questionnaire for background information.

teachers, some were from junior high schools, while others were from senior high schools, and other teacher groups were also well-represented. Among them, 33.34% were males and 66.66% were females; 25% were aged 30 years and below, 41.67% were aged 31–40 years, and 33.33% were aged 41–50 years; 41.67% had taught for 10 years and below, 33.33% for 11–20 years, and 25% for 21–30 years. Regarding job titles, 25% were at junior level and below, 33.33% were at an intermediate level, and 41.67% were at senior level. In the survey process, first, researchers set up a survey team and trained the members to ensure that they fully grasped the purpose and method of the survey; then, with the consent of the teachers, we introduced the purpose, method and data analysis results of the study to the teachers; finally, we interviewed the teachers according to the interview questionnaire, using shorthand text recording, respecting the interviewees' request not to record their voices, and keeping the interview text strictly confidential and anonymous to fully protect the privacy of the interviewees.

Instruments

The questionnaire used in this study was answered anonymously and consisted of two parts—teachers' basic information and their autonomous learning capacity— with a total of 18 questions. The first part covered the teachers' personal factors and external contextual factors (gender, age, years of teaching experience, educational background, job title, educational stage, and location). The second part was the Teachers' Autonomous Learning Capacity Scale. The questions were mostly drawn from the teachers' autonomous learning capacity items of the Teachers' Key Literacy and Competencies Scale devised by Kang et al. (2021) and structured according to the scale of Williamson (2007). The final version of the questionnaire comprised four subdimensions: development of study plans, proficiency in using learning methods, learning habit formation, and evaluation of learning outcomes. To avoid subjective judgment and biased results related to the "do you agree" questions, experience, means, or situation-based questions were used, mostly about the teachers' behaviors or performance in teacher education, educational teaching, and pedagogical research activities. For example, a question about the development of study plans reads: Which of the following statements best describes your experience in the development of study plans in the recent 5 years? A. Have never made a personal study plan. B. Have included specific study plans in work plans for 1–3 semesters. C. Have included specific study plans in work plans for 4–6 semesters. D. Have included specific study plans in work plans for seven or more semesters.

Using the SPSS 24.0, the reliability test of the scale determined using Cronbach's *a* yielded a value of 0.825, indicating high reliability. Correlation analysis was performed to examine the scale's structural validity. The correlation coefficients between the subdimensions ranged from 0.36 to 0.60, implying a medium correlation between the indicators. The correlation coefficients between the subdimensions and the overall measure ranged from 0.62 to 0.91, indicating a high correlation between the four subdimensions and teachers' autonomous learning capacity (see Table 3), and overall good structural validity of the scale.

The interview outline included four themes: development of study plans, proficiency in using learning methods, learning habit formation, and evaluation of learning outcomes; these can be mapped to the four subdimensions of the scale. To gain a deeper understanding of the interviewers' experience with autonomous learning and its influencing factors, six items were developed based on the outline. Then, they were revised again to incorporate inputs from experts after the interview instrument was drafted. The questions are as follows: (1) Do you make a study plan before autonomous learning? How long is the time horizon covered in the study plan? What are the factors that influence your decision for developing a study plan? (2) Do you use anything as a reference when making a study plan? As a mathematics teacher, what are the areas involved in autonomous learning? (3) In what way do you carry out autonomous learning? What are the factors

TABLE 3 Correlation between teachers' autonomous learning capacity and its subdimensions.

	Autonomous learning capacity	Development of study plans	Proficiency in using learning methods	Learning habit formation	Evaluation of learning outcomes
Development of study plans	0.768**	1			
Proficiency in using learning methods	0.622**	0.368**	1		
Learning habit formation	0.762**	0.452**	0.454**	1	
Evaluation of learning outcomes	0.903**	0.592**	0.420**	0.536**	1

**$p < 0.01$.

that influence your decision on choosing the learning methods? (4) How long is the cycle of your autonomous learning? What steps do you usually take to form a learning habit? What do you think are the reasons why "there is no time for autonomous learning?" (5) Do you assess your own learning? What areas are considered for the assessment? What are the factors that influence your decision for assessing your learning? (6) What should be done to improve teachers' autonomous learning capacity? What are your expected supports from your family, school, and community?

Data analysis

Addressing research question 1, SPSS 24.0 was used for descriptive statistical analysis of the data and for calculating score rates of teachers' autonomous learning capacity and the subdimensions under it. The results presented an overall picture of the autonomous learning capacity of secondary school mathematics teachers. The score rates were calculated as the percentage ratios of the scores to the total points (score rate = score/total points for a dimension). Higher score rates indicate higher levels of teacher capacity in that dimension.

For research question 2, *t*-tests and one-way analysis of variance (ANOVA) were used to test for differences in teachers' autonomous learning capacity across gender, age, years of teaching experience, educational background, job title, location, and educational stage groups. For groups with significant differences, the least significant difference test was used *post hoc* to identify specific differences between the teachers. Furthermore, the effect sizes were calculated to analyze the differences. Specifically, $\eta^2 = 0.01$ (or $d = 0.2$), $\eta^2 = 0.06$ (or $d = 0.5$), and $\eta^2 = 0.14$ (or $d = 0.8$) were used to represent small, medium, and large effects, respectively (Cohen, 1988). Finally, for a robust interpretation of the survey results and a more comprehensive understanding of the factors affecting teachers' autonomous learning capacity, the interview outline was used for the teachers' semi-structured interviews. Two researchers conducted independent qualitative analyses of the interview transcript to summarize the teachers' main views using a bottom-up approach.

Results

Current autonomous learning capacity of secondary school mathematics teachers

Regarding the overall results of the autonomous learning capacity of secondary school mathematics teachers, the mean score was 29.31 with a score rate of 81.42%. The score rate can also be interpreted as the percentage score, which indicates that the overall level of teachers' autonomous learning ability, measured on a percentage scale, reached a corresponding level of 81.42 points, which is a good level. The highest score of 36 had a high frequency, indicating that there were multiple outstanding teachers. The lowest score of 16 had a low frequency, suggesting a desirable result.

By subdimension (see Table 4 for detailed results), the results of mean and median scores greater than 3, negative kurtosis and skewness, and the left-skewed distribution indicated that the scores are generally high and most teachers have good performance. To be specific, teachers had the best performance in the development of study plans ($M = 3.342$), satisfactory performance in the evaluation of learning outcomes ($M = 3.266$), and poor performance in learning habit formation ($M = 3.200$) and proficiency in using learning methods ($M = 3.156$), as shown in Figure 1.

Regarding particular items, in the development of study plans, 97.7% of teachers included specific study plans in their work plans for varying lengths (in semesters), while 2.3% of teachers had never made a personal study plan, indicating that most teachers perform well in this subdimension. In the proficiency in using learning methods, we primarily examined whether teachers were able to conduct autonomous learning in combination with the following methods: reading books or journals, browsing professional websites, observing public classes, watching videos of quality classes, attending lectures, participating in online courses, and visiting field trips. Moreover, we found that 78.3% of teachers were aware of all the above learning methods and used at least half of them in learning, 15.9% of teachers were aware of half of them and used a few for effective learning, and 5.8% of teachers did not know the learning methods and had never used them, indicating a good

TABLE 4 Descriptive results of teachers' autonomous learning capacity by subdimension.

	N	Mean (*M*)	Median	Standard deviation	Variance	Skewness	Kurtosis	Min.	Max.
Development of study plans	397	3.342	3.500	0.677	0.458	−0.758	−0.284	1.00	4.00
Proficiency in using learning methods	397	3.156	3.000	0.894	0.799	−0.802	−0.224	1.00	4.00
Learning habit formation	397	3.200	3.500	0.708	0.501	−0.713	−0.108	1.00	4.00
Evaluation of learning outcomes	397	3.266	3.250	0.653	0.426	−0.549	−0.669	1.25	4.00

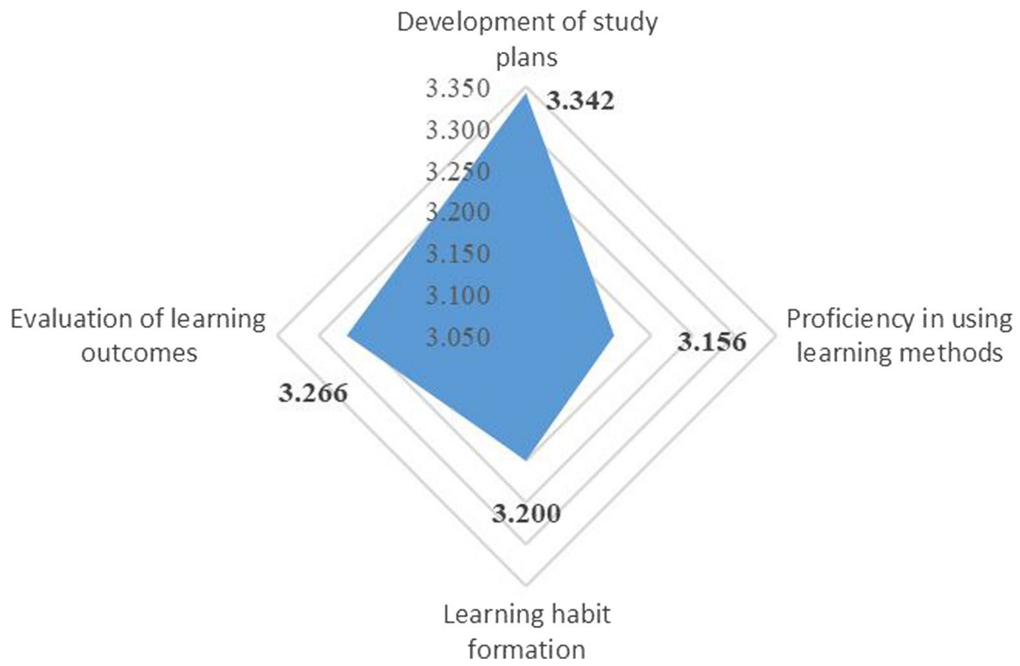


FIGURE 1
Subdimension means of teachers' autonomous learning capacity.

performance by most teachers in this subdimension. In the learning habit formation, 73.2% of teachers reported taking some time to study regardless of whether they are busy at work or not, 22.5% only studied intermittently, and 4.3% had no time to study, indicating that while most teachers have an established study habit, there is still room for further improvement. In the evaluation of learning outcomes, we mainly examined whether teachers can make a comprehensive assessment of the autonomous learning process, including the initiative of learning, continuity of learning, effectiveness of learning methods, effectiveness of learning strategies, and smoothness of the learning process. It was found that 92.9% of teachers could diagnose at least two of the above aspects, and only 7.1% of teachers did not evaluate any aspect of the learning process, indicating that most teachers perform well in this subdimension.

Analysis of differences across teacher groups

To examine the effects of different factors on teachers' autonomous learning capacity, independent samples t-tests were conducted for gender and educational stage. The results showed significant differences with small to medium effect sizes in autonomous learning capacity between male and female mathematics teachers ($t = -2.067$, $p = 0.039 < 0.05$, $d = 0.259$), and between junior and senior high school teachers ($t = 2.796$, $p = 0.006 < 0.01$, $d = 0.292$). Specifically, female teachers outperformed male teachers, and junior high school teachers outperformed senior high school teachers.

Then, one-way ANOVA tests were conducted for age, years of teaching experience, educational background, job title, and

TABLE 5 One-way analysis of variance for teachers' autonomous learning capacity.

	Categories		N	Mean	SD	F	Post-hoc tests	η^2
Age	A	30 and below	55	30.073	5.319	2.713*	C < D	0.020
	B	31–40	123	29.252	4.970			
	C	41–50	153	28.549	5.219			
	D	51 and above	65	30.436	4.1548			
Educational background	A	Associate and below	9	29.223	2.539	2.650*	C > D	0.020
	B	Bachelor	298	29.594	4.975			
	C	Master	83	28.615	5.277			
	D	Doctor	6	24.5623	5.4360			
Location	A	Urban areas	244	28.623	4.967	6.575**	A < B	0.033
	B	Township	123	30.255	5.095			
	C	Rural area	21	31.476	3.945			

*$p < 0.05$ and **$p < 0.01$.

locational factors. Significant differences with small to medium effect sizes were identified for teachers in age ($F = 0.045$, $p = 0.045 < 0.05$, $\eta^2 = 0.020$), educational background ($F = 2.650$, $p = 0.049 < 0.05$, $\eta^2 = 0.020$), and location ($F = 6.575$, $p = 0.002 < 0.01$, $\eta^2 = 0.033$). Specifically, teachers aged 51 years and above performed significantly better than those in the 41–50 age group, teachers who worked in rural areas and townships were significantly better than urban teachers, and teachers with master's degrees and below were significantly better than those with doctoral degrees in autonomous learning capacity (see Table 5). In addition, there was no significant difference in the autonomous learning capacity of mathematics teachers in terms of years of teaching and job title ($p > 0.05$).

To further explore the differences in teachers' autonomous learning capacity across subdimensions, independent samples t-tests were conducted for each of the four subdimensions with respect to gender and educational stage. The results showed that female teachers were significantly better than male teachers in the evaluation of learning outcomes ($t = 2.540$, $p = 0.013 < 0.05$, $d = 0.345$), and junior high school teachers were significantly better than senior high school teachers in learning habit formation ($t = 2.078$, $p = 0.038 < 0.05$, $d = 0.213$), with small to medium effect sizes.

Then, one-way ANOVA tests were conducted for age, years of teaching experience, educational background, job title, and locational factors successively for each of the four subdimensions (see Table 6). Significant differences with small to medium effect sizes were identified in teachers across age ($F = 3.757$, $p = 0.011 < 0.05$, $\eta^2 = 0.028$) and location ($F = 4.084$, $p = 0.018 < 0.05$, $\eta^2 = 0.021$) in development of study plans. Specifically, teachers aged 51 years and above significantly outperformed teachers aged 31–50 years, and rural teachers outperformed urban teachers.

Significant differences with small to medium effect sizes were identified in teachers across age ($F = 3.468$, $p = 0.016 < 0.05$, $\eta^2 = 0.026$), years of teaching experience ($F = 3.213$, $p = 0.023 < 0.05$, $\eta^2 = 0.024$), and location ($F = 4.523$, $p = 0.011 < 0.05$, $\eta^2 = 0.023$) in proficiency in using learning methods. Specifically, teachers aged 41–50 years were less proficient than those aged 31–40 years and over 51 years.

Significant differences with small to medium effect sizes were identified in teachers across educational background ($F = 5.958$, $p = 0.001 < 0.01$, $\eta^2 = 0.044$) and job title ($F = 2.862$, $p = 0.037 < 0.05$, $\eta^2 = 0.021$) in learning habit formation. Specifically, teachers with master's degrees and below performed significantly better than those with doctoral degrees, teachers with bachelor's degrees performed significantly better than those with master's degrees, and teachers with senior titles significantly underperformed those with intermediate and below titles.

Significant differences with small to medium effect sizes were identified in teachers across years of teaching experience ($F = 3.015$, $p = 0.030 < 0.05$, $\eta^2 = 0.023$) and location ($F = 6.029$, $p = 0.003 < 0.01$, $\eta^2 = 0.030$) in evaluation of learning outcomes. Specifically, teachers with 21–30 years of teaching experience significantly underperformed teachers with 10 years and below and 31 years and more of teaching experience, and township teachers performed significantly better than urban teachers.

Qualitative analysis of interview data

Semi-structured interviews were conducted with 12 middle and high school mathematics teachers to assist in interpreting the results of the data analysis. The qualitative analysis revealed that the differences in autonomous learning capacity across groups could be explained by both internal and external factors.

Internal factors

Age and years of teaching experience: Mathematics knowledge base

During the interviews, most teachers agreed that adequate mathematics expertise could increase their efficiency in autonomous learning and acquiring new knowledge and skills. The perception of improved self-efficacy, in turn, could initiate a

TABLE 6 One-way analysis of variance for teachers' autonomous learning capacity by subdimension.

Subdimensions	Categories		*N*	*Mean*	*SD*	*F*	*Post hoc* tests	η^2
Development of study plans	Age (in years)	A. 30 and below	55	3.309	0.773	3.757*	C < D	0.028
		B. 31–40	123	3.333	0.706			
		C. 41–50	153	3.255	0.667			
		D. 51 and above	65	3.585	0.489			
	Location	A. Urban areas	244	3.271	0.693	4.084*	A < C	0.021
		B. Township	123	3.438	0.667			
		C. Rural area	21	3.595	0.436			
Proficiency in using learning methods	Age (in years)	A. 30 and below	55	3.182	0.819	3.468*	B > C	0.026
		B. 31–40	123	3.309	0.780			
		C. 41–50	153	2.981	0.970			
		D. 51 and above	65	3.246	0.919			
	Years of teaching experience	A. 10 and below	81	3.222	0.791	3.213*	B > C	0.024
		B. 11–20	143	3.308	0.824			
		C. 21–30	129	2.985	0.984			
		D. 31 and above	41	3.098	0.944	4.523*	A < C	0.023
	Location	A. Urban areas	244	3.074	0.944			
		B. Township	123	3.244	0.803			
		C. Rural area	21	3.619	0.590			
Learning habit formation	Educational background	A. Associate and below	9	3.333	0.500	5.958**	B > C A > D	0.044
		B. Bachelor	298	3.251	0.672		B > D	
		C. Master	83	3.072	0.770		C > D	
		D. Doctor	6	2.170	0.982			
	Job title	A. Junior and below	47	3.340	0.635	2.862*	B > C	0.021
		IB. ntermediate	183	3.154	0.701			
		C. Senior	165	3.215	0.723			
Evaluation of learning outcomes	Years of teaching experience	A. 10 and below	81	3.414	0.64488	3.015*	C < D	0.023
		B. 11–20	143	3.242	0.62810			
		C. 21–30	129	3.173	0.68982			
		D. 31 and above	41	3.404	0.51782			
	Location	A. Urban areas	244	3.181	0.66001	6.029**	A < B	0.030
		B. Township	123	3.393	0.61987			
		C. Rural area	21	3.512	0.60455			

*$p < 0.05$ and **$p < 0.01$.

virtuous cycle for continuous autonomous learning. The majority of the participants explained that teachers with domain knowledge of mathematics are more likely to understand what they are learning because they can relate new knowledge with what they already know. The network of knowledge means that they better locate anything within the knowledge system, which provides them with a holistic view to approach new knowledge, thus achieving better learning efficiency.

For example, Teacher A said: "I feel that the obvious gap between myself and veteran teachers is that my mathematics-related knowledge (for example, history of mathematics, mathematical analysis, calculation skills) is not as extensive as theirs. I spend much more time than they spend in acquiring the same new knowledge. Also, they are more thoughtful. I often miss the points that veteran teachers may raise. Besides, my views on some of the frontier issues can also enlighten veteran teachers."

Teacher E said: "In a workshop on what 'differences' should be reflected in the approach of 'adopting different methods for the same mathematics lesson,' I thought of a teacher's personal teaching style, handling of teaching content, teaching ideas and methods, and the use of teaching tools. However, a veteran teacher mentioned the importance of paying attention to the students' differences when planning a 'different' lesson. This is something I had not thought of."

The results of the above interviews reveal that while newly recruited young teachers may be inexperienced, they are more

 10.3389/fpsyg.2022.1042838

capable and motivated for self-reflection and autonomous learning, and are more concerned with the frontier of the subject. Veteran teachers, on the other hand, have a greater advantage in terms of their knowledge base in mathematics.

Location: Learning methods

The interviews suggested that appropriate learning methods contribute to mathematics teachers' efficient learning. The rapid development of technologies, including multimedia technology and artificial intelligence, has expanded available applications for mathematics and methods for autonomous learning. As a provincial and a coastal port city in China, T city has few agricultural areas and mountainous areas, a better level of education informatization, and a wide coverage of teacher training and post-service training power, so teachers in rural areas have access to rich learning resources and opportunities. The right approach can effectively improve the efficiency of teachers' autonomous learning and learning outcomes. Choosing appropriate and suitable learning methods is a necessary prerequisite for teachers to engage in autonomous learning.

For example, Teacher D said: "To improve my professional competencies, I use online resources for autonomous learning, such as watching recorded videos for national high-quality lessons and learning information technology and pedagogical skills. I also participate in the mathematics pedagogical research activities organized by the school to learn from the experiences of other teachers."

Teacher C said: "Although our school is located in a remote location, thanks to the development of technologies, China has made many databases publicly accessible. I can observe online quality lessons delivered by mathematics teachers nationwide and learn the way excellent teachers handle their classes."

Teacher E said: "I try to improve my knowledge and professionalism through mathematics teaching reference books. Now that it's easy to get around, sometimes, I also go to downtown schools to attend lectures by experts in mathematics education to get my questions about learning and teaching answered. In addition, I think lesson preparation is also a good way to learn. I acquire or learn new ideas every time I prepare a lesson."

Gender, educational stage, and educational background: Internal motivations for learning

The interviews revealed that internal motivations are a necessary prerequisite for teachers' autonomous learning. What motivates teachers to learn autonomously can be internal, including their personal values and goals for self-improvement. These motivations are catalytic and inspirational, encouraging teachers to enhance their professional knowledge and skills, knowledge base, and determination in overcoming difficulties they encounter in the learning process continuously and sustainably.

For example, Teacher F said: "As a female teacher, I am often able to derive satisfaction from teaching. My self-improvement is driven by good interaction with students in the classroom and the improvement in student achievement. I study regularly and autonomously. I find ways to overcome any difficulties and obstacles I encounter in the learning process, rather than giving up."

Teacher G said: "I teach senior high school mathematics, which is difficult, so I usually spend most of my time teaching and preparing lessons. My main goal is to help students improve their grades in the college entrance examination, so I have little need to study myself and have no set study habit."

Teacher K said: "As a teacher who has experienced systematic research training at the doctoral level, I believe that my mathematical expertise meets the requirements for teaching, so I do not need much additional independent study. There are few mathematics teachers with doctoral degrees, and the teachers around me often talk to me about professional issues. I am not particularly knowledgeable about curriculum standards and other content, and usually do not spend time studying them."

From the results of the above interviews, it is evident that female teachers are able to gain satisfaction from teaching mathematics and have a high internal motivation for learning, a strong sense of autonomous learning, and personal learning needs and pursuits. In contrast, senior high school teachers are more dedicated to teaching, preparing lessons, and helping students improve their academic performance in mathematics and prepare for the college entrance examination. They have weak internal motivation for learning and forming learning habits. Additionally, the doctoral-educated teachers who participated in this study had science and technology academic backgrounds, had not experienced teacher education, and had weaker knowledge of educational expertise, which, combined with the security that comes with a deep foundation of mathematical expertise, prevented them from developing the habit of autonomous learning. This finding corroborates and explains the results of the questionnaire survey that "in autonomous learning capacity, female teachers are significantly better than male teachers, junior high school teachers are better than senior high school teachers, and doctoral teachers underperformed."

External factors

Age and years of teaching experience: Learning time

In the interviews, most teachers mentioned that they had no free time for autonomous learning due to the numerous chores at home and the heavy workload. With little free time at their disposal, they had time for autonomous learning only during semester breaks, but they still had to prepare lessons for the semester. As a result, many teachers' autonomous learning covers only knowledge closely related to the subjects they teach.

Teacher H said: "Classroom work, various interpersonal activities, and regular teaching take up most of the time. In addition, as middle-aged people, they have to take care of their families and children. Thus, they rarely have time for autonomous learning and usually study for completing the tasks assigned by the school."

Teacher B said: "The school conducts a lot of teacher training programs. As middle-aged teachers, we have too many chores in our family life, so we do not have time for autonomous learning."

Job titles and educational background: Reward and evaluation policies

Sound evaluation policies are critical to measure the quality of teachers' autonomous learning. Schools play a major role in teacher evaluation; therefore, a robust evaluation framework is needed in schools to motivate teachers for autonomous learning. The interviews revealed that teacher evaluation policies place too much emphasis on differentiation and selection. Accordingly, teachers focus on these performance indicators, particularly on students' academic achievements. They have few incentives for enhancing their professional competencies through autonomous learning.

For example, Teacher L said: "The school's evaluation policies are based on achievements of students in the class, the papers published by teachers, and the rankings achieved in competitions. Even those who want to learn are targeting promotion or job titles. Teachers are not motivated to learn for improved knowledge and learning capacity."

Teacher K said: "teachers with a doctoral degree have certain advantages in the job title evaluation, as a doctoral degree mathematics teacher, my professional knowledge is relatively more solid, and I usually like to study and research mathematics knowledge."

The results of the above interviews identified a lack of time for learning, limited energy, and work and family factors that weaken autonomous learning capacity among middle-aged teachers. In addition, the interviews revealed that teachers with senior job titles and doctoral degrees are not sufficiently incentivized for enhancing autonomous learning capacity by the reward and evaluation policies.

Discussion

Satisfactory autonomous learning capacity of secondary school mathematics teachers

Considering the findings of Shao (2013) and Sun (2012), this study revealed generally satisfactory autonomous learning capacity of secondary school mathematics teachers. Specifically, among the four subdimensions, teachers' ability in the development of study plans was the best, indicating that most teachers can set clear learning goals and develop study plans based on professional requirements through self-reflection, incorporating suggestions from colleagues, and considering their own professional learning needs. However, the interviews revealed that teachers' planning cycles generally last for a semester. They mostly focus on the next semester's course schedule rather than on professional self-improvement. Moreover, teachers' ability to evaluate learning outcomes is satisfactory, indicating that they can use a combination of assessment tools to evaluate learning outcomes. However, it was revealed that some teachers are overly interested in results and have little appetite for reflecting on the learning process. Finally, improvements are needed in teachers' abilities to form learning habits and gain proficiency in using learning methods, both of which require practice and time. In the interview, we found that although teachers have a desire to learn and improve, many of them report that their study habits are interrupted by external factors due to work commitments and life chores, and they lack time for autonomous learning. In addition, increased job stress can lead to a decrease in their job satisfaction and a lack of time for professional development and enhancement (Naylor, 2001), which can weaken their competency.

Differences in autonomous learning capacity of secondary school mathematics teachers across contextual factors

The results revealed significant differences in teachers' autonomous learning capacity across gender, age, years of teaching experience, educational stage, and location. First, between gender groups, female teachers' autonomous learning capacity is significantly better than that of male teachers, especially in terms of evaluation of learning outcomes. The reason is that female teachers in compulsory educational stages in China have higher job satisfaction than males (Li et al., 2017). As a result, it is easier for female teachers to receive positive feedback in their teaching. Their increased self-efficacy helps them make better career decisions (Ince Aka and Tasar, 2020), motivating them to learn autonomously for professional improvement, which was also verified in the interviews. Self-evaluation of learning outcomes helps them diagnose their shortcomings and find the right direction for learning improvement. In addition, the results of the interviews suggest that teachers' internal motivation contributes to the improvement of their autonomous learning capacity. These results can also explain the finding.

Among educational stages, junior high school teachers have better levels of autonomous learning capacity, especially in learning habit formation. The interviews also revealed that some senior high school teachers mistakenly equate a higher educational stage with higher competencies, resulting in a low level of enthusiasm for their professional advancement. It was also revealed that teachers' lack of internal motivation to learn can affect their planning for autonomous learning, making it difficult for them to form study habits. Moreover, a higher level of educational stage is accompanied by increased pressure due to the college entrance examinations that students in these classes undertake. Senior high school teachers devote most of their energy to education and teaching, instead of paying attention to autonomous learning for self-improvement. In addition, the interviews also showed that the utilitarian orientation of reward

10.3389/fpsyg.2022.1042838

and evaluation policies, as well as less rewarding teacher training activities, occupy too much of the available time for teachers. The situation is worsened by the fact that teachers' professional learning and development are mostly driven by external sources (Lefstein et al., 2020). As a result, teachers have no time for autonomous learning and maintaining learning habits.

Among age groups, teachers aged 41–50 years are less capable of autonomous learning than teachers aged 51 years and above, especially in terms of the development of study plans. In contrast, the former are less adept than those aged 31–40 years in proficiency in using learning methods. This finding is in line that of with previous studies (e.g., Huberman, 1989; Liu and Liu, 2019), and is understandable because teachers aged 41–50 years have transformed from novice to skilled teachers. Given the accumulated work experience, they are prone to burnout, leading to an insufficient passion for learning and development. Meanwhile, the finding that school learning is most conducive to teachers' professional learning and development (Postholm, 2012) implies that middle-aged teachers are in a disadvantaged position for autonomous learning because they need to balance their work and familial responsibilities, which was verified in the interviews. Furthermore, the interviews indicated that teachers' knowledge base in mathematics also affects their autonomous learning capacity. Teachers aged 41–50 years were less knowledgeable and experienced in mathematics than senior teachers and less refreshed than younger teachers. The burnout effects can further decrease their motivation for autonomous learning.

There were no significant overall differences between teachers based on years of teaching experience. However, in the subdimensions, teachers with 21–30 years of teaching experience had low proficiency in using learning methods compared to those with 11–20 years of teaching experience and poor evaluation of learning outcomes compared to those with 31 years and more teaching experience. A previous study revealed that teachers with 6–10 years of teaching experience have the highest level of career attraction and are in the prime of their career development, while teachers over 40 years of age show negative performance and lack the motivation to learn (Liu and Liu, 2019). Huberman (1989) divided the professional life cycle of teachers into seven stages, in which the period of calm and relational distancing can portray the psychological state of teachers with 21–30 years of teaching experience. Many teachers at this stage begin to calm down after experiencing doubt and crisis and are able to complete their classroom teaching with greater ease and self-confidence. However, as career aspirations are gradually achieved, their level of ambition starts declining (Ye, 2001), and the need for self-evaluation decreases. In addition, the interview results suggested that learning methods and internal learning motivations affect the development of this group of teachers' learning capacity. Teachers in a period of calm and relational distancing also have a reduced willingness to explore learning methods. This explains part of the differences with other teachers.

Among location groups, rural or township teachers did not show deficiencies in overall autonomous learning capacity or in the development of study plans, proficiency in using learning methods, and evaluation of learning outcomes. This finding differs from those that suggested that rural teachers need to improve their competencies because they are mediocre (Li and Cui, 2015). One possible reason to explain the inconsistency is that collaboration among teachers contributes to their professional learning and development (Smith et al., 2020). Participants of this study were from the areas of Chinese provincial and coastal cities, with few agricultural and mountainous areas, a good level of education informatization, and a wide coverage of teacher training and post-service training efforts. The rural teachers in the interviews also indicated that they had access to abundant learning resources and the opportunity to learn independently using a variety of learning methods and to collaborate with teachers from other regions. Efforts over the years for quality and balanced development of compulsory education and provision of special training for rural teachers have yielded positive result. Additionally, the job stress of rural teachers is relatively low, giving them more time for autonomous learning. To compensate for the lack of resources in townships and rural areas, teachers in these areas have more moral duties for self-improvement, resulting in higher autonomous learning capacity of rural teachers than those of urban peers across all subdimensions.

Among educational background groups, those with doctoral degrees did not have better autonomous learning capacity, contrary to our expectation, especially in learning habit formation. Some studies have also shown that the effect of educational background on teachers' learning capacity is not significant (Zhu and Jiang, 2021). During the interviews, teachers with doctoral degrees indicated that the mathematical professional learning they had conducted in scientific research was sufficient for their teaching practice and that there was little pressure for independent learning. It is noteworthy that the participants with doctoral degrees were not educated as professional teachers, they lacked teacher education backgrounds. The interviews also revealed that they paid more attention to the learning of mathematics professional theories, but not enough attention to the learning of mathematics teaching and learning theories, and lacked the awareness of mathematics curriculum standards, which made their knowledge in autonomous learning limited to mathematics professional knowledge rather than educational knowledge, and coupled with the high level of superiority brought by high education, these may lower their motivation for learning and development. Furthermore, according to the Education Statistics in T city, the number of teachers with doctoral degrees in mathematics and teacher training backgrounds is very small, which limits the number of people surveyed, and the results may change in the future as the cultivation of doctoral degrees in education in T city deepens, which is the focus of our further in-depth research in the future. In addition, the results revealed that the utilitarian tendency of teacher rewards and evaluation policies can hinder the improvement of teachers' autonomous learning capacity. For example, in the interviews, teachers with doctoral degrees indicated that they have certain advantages in teaching in primary and secondary schools and evaluating teacher job titles in China and that they do not have to continue to do research or spend a lot of time studying for their job title rank after teaching. Therefore, they are more comfortable with their work status and less enthusiastic about learning. This is consistent with prior findings. For example, some studies have suggested that China's teacher evaluation policies should focus on external static indicators such as educational background and job title (Wang and Si, 2011), without considering indicators that reflect the true quality of teachers, leading to a superficial quality evaluation framework (Wang, 2019).

In terms of job titles, there are no differences in the overall autonomous learning capacity among teacher groups. However, in learning habit formation, those with senior job titles do not demonstrate the expected advantages and even underperform those with intermediate job titles. The finding is consistent with existing research conclusions. For example, a previous study found that the learning motivation level of senior job title teachers is the lowest, and the difference with non-senior title teachers is significant (Liu and Liu, 2019). The reason could be because some teachers have become less motivated to work after promotion to senior titles or have moved to non-teaching positions, where they occupy senior title positions but do not teach (Wang and Wu, 2019). Moreover, interviews revealed that job title evaluations focus on exogenous indicators, while teachers at the senior level have reached the culmination of professional development with little hope of further promotion, which can reduce the willingness of some teachers with senior titles to seek autonomous learning and development.

In summary, the findings about differences in the autonomous learning capacity of mathematics teacher groups across demographic and contextual factors can contribute to the literature in this area. The implications are that measures need to be taken to enhance the autonomous learning capacity of particular teacher groups. For example, the unsatisfactory performance of teachers with senior titles and high levels of education may highlight the need for reforming job titles, pay scales, and teacher recruitment systems. For male teachers, those working at higher educational levels, and the middle-aged, targeted training and pedagogical research policies may be advisable.

Future research suggestions

This study investigated the current state of teachers' autonomous learning capacity, examined the differences across teacher groups, and discussed underlying reasons. First, the findings may inform future studies in developing more comprehensive scales, thus exploring factors affecting teachers' autonomous learning capacity from a wider range of perspectives. Second, although this study has identified some significant influencing factors, their mechanisms of action still need further exploration. The relationship between correlated variables can be examined further to gain a clearer understanding of the paths through which teachers' autonomous learning is affected. Finally, teacher policies have an important impact on the construction of the teaching team, the assessment is part of the efforts to provide evidence to inform teacher policies. A question to be addressed is whether artificial intelligence can be employed to conduct big data assessment of autonomous learning capacity comprehensively and in real-time to save effort.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Tianjin Normal University Academic Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

GW, YK, and XC contributed to the construction of concepts, evaluation index system, questionnaire preparation, introduction, and methodology. YZ and XC contributed to the literature review and theoretical background. YK and FL collected and analyzed the data. YK, YZ, and HH wrote the original draft of the manuscript. GW, YK, and YZ revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the key project of the National Social Science Foundation of China (Project No. AFA170008) and Major Project of "The Fourteenth Five-Year Plan of Teacher Education Development and the Future Development of Teacher Education" of China Higher Education Society's Teacher Education Branch in 2020: Research on the Evaluation Index System and Assessment Model of Teachers' Core Qualities and Abilities (Project No. 20ZSJSJYZG02).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdel Razeq, A. A. (2014). University EFL learners' perceptions of their autonomous learning responsibilities and abilities. *RELC J.* 45, 321–336. doi: 10.1177/0033688214547035

Anderson, D. R. (2002). Creative teachers: risk, Responsibility, and Love. *J. Educ.* 183, 33–48. doi: 10.1177/002205740218300104

Bandura, A. (1962). "Social learning through imitation," in *Nebraska Symposium on Motivation, 1962*. (Oxford, England: Univer. Nebraska Press), 211–274.

Billett, S. (2002). Toward a workplace pedagogy: guidance, participation, and engagement. *Adult Educ. Q.* 53, 27–43. doi: 10.1177/074171302237202

Borko, H. (2004). Professional development and teacher learning: mapping the terrain. *Educ. Res.* 33, 3–15. doi: 10.3102/0013189x033008003

Bowles, T., Hattie, J., Dinham, S., Scull, J., and Clinton, J. (2014). Proposing a comprehensive model for identifying teaching candidates. *Aust. Educ. Res.* 41, 365–380. doi: 10.1007/s13384-014-0146-z

Certo, J. L., and Fox, J. E. (2002). Retaining quality teachers. *High Sch. J.* 86, 57–75. doi: 10.1353/hsj.2002.0015

Chen, X. (2000). *Qualitative Research Methods and Social Science Research*. Beijing: Education Science Press.

Chen, C. (2017). Teacher professional development and its influencing factors in an international perspective – an empirical study based on TALIS data. *Int. Comp. Educ.* 39, 84–92. (in Chinese).

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Creemers, B., Kyriakides, L., and Antoniou, P. (2012). *Teacher Professional Development for Improving Quality of Teaching*. Dordrecht: Springer.

Day, C., and Gu, Q. (2007). Variations in the conditions for teachers' professional learning and development: sustaining commitment and effectiveness over a career. *Oxf. Rev. Educ.* 33, 423–443. doi: 10.1080/03054980701450746

De Jong, L., Meirink, J., and Admiraal, W. (2019). School-based teacher collaboration: different learning opportunities across various contexts. *Teach. Teach. Educ.* 86:102925. doi: 10.1016/j.tate.2019.102925

Deregözü, A., and Hatipoğlu, S. (2018). An investigation on prospective German language teachers' autonomous learning level. *J. Lang. Ling. Stud.* 14, 1–10. Retrieved from https://dergipark.org.tr/en/pub/jlls/issue/43213/527571

Goodlad, J. (1990). *Teachers for Our Nation's Schools*. San Francisco, CA: Jossey Bass.

Hai, Y., and Liu, T. (2018). The analysis of developmental level and characteristics of preschool Teachers' self-regulated learning. *J. Educ Stud.* 14, 89–102. (in Chinese).

Harris, D. N., and Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *J. Public Econ.* 95, 798–812. doi: 10.1016/j.jpubeco.2010.11.009

Holec, H. (1979). *Autonomy and Foreign Language Learning*. Oxford: Pergamon Press.

Holzberger, D., Philipp, A., and Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: a longitudinal analysis. *J. Educ. Psychol.* 105, 774–786. doi: 10.1037/a0032198

Huberman, M. (1989). The professional life cycle of teachers. *Teach. Coll. Rec.* 91, 31–57. doi: 10.1177/016146818909100107

Ince Aka, E., and Tasar, M. F. (2020). Prospective science Teachers' views on career planning and their self-efficacy levels for career decision-making. *Int. Online J. Educ. Teach.* 7, 960–975. Retrieved from https://iojet.org/index.php/IOJET/article/view/947

Kang, Y., Zhen, Y., Zhang, X., and Wang, G. (2021). "A study of the current status of teachers' Core qualities and abilities and the importance of their background factors in primary and secondary schools in Tianjin" in *Artificial Intelligence in Education and Teaching Assessment*. eds. W. Wang, G. Wang, X. Ding and B. Zhang (Singapore: Springer Singapore), 213–233.

Kelly, P. (2006). What is teacher learning? A socio-cultural perspective. *Oxf. Rev. Educ.* 32, 505–519. doi: 10.1080/03054980600884227

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., and Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *J. Educ. Psychol.* 105, 805–820. doi: 10.1037/a0032583

Kyriakides, L., Christoforou, C., and Charalambous, C. Y. (2013). What matters for student learning outcomes: a meta-analysis of studies exploring factors of effective teaching. *Teach. Teach. Educ.* 36, 143–152. doi: 10.1016/j.tate.2013.07.010

Lefstein, A., Vedder-Weiss, D., and Segal, A. (2020). Relocating research on teacher learning: toward pedagogically productive talk. *Educ. Res.* 49, 360–368. doi: 10.3102/0013189x20922998

Li, S., and Cui, Y. (2015). Investigation and research on the status of rural Teachers' professional development in the process of new urbanization——based on the empirical analysis of Sichuan, Yunnan, Guizhou and Chongqing. *Educ. Res.* 36, 98–107. (in Chinese).

Li, W., Xu, J., and Ding, X. (2017). An empirical study on job satisfaction of compulsory education teachers——based on a survey of 20 counties in 9 provinces. *Modern Educ. Manag.* 1, 79–84. (in Chinese).

Liu, L., and Liu, J. (2019). Research on Teachers' learning dynamics from the perspective of complex science. *J. Chin. Soc. Educ.* 9, 92–96. (in Chinese).

Loeng, S. (2020). Self-directed learning: a core concept in adult education. *Educ. Res. Int.* 2020, 1–12. doi: 10.1155/2020/3816132

Mynard, J., and Sorflaten, R. (2006). Independent learning in your classroom. *Teach. Learn. Curr. J.* 1, 34–38.

Naylor, C. (2001). "Teacher workload and stress: an international perspective on human costs and systemic failure," in: BCTF Research Report. (ed.) B.C.T. Federation. (Canada: British Columbia Teachers' Federation, Vancouver).

Olivier, J., and Wittmann, G.-E. (2019). Professional development in fostering self-directed learning in German second additional language teachers. *Per Linguam: a Journal of Language Learning = Per Linguam: Tydskrif vir Taalaanleer* 35, 125–142. doi: 10.5785/35-3-870

Opfer, V. D., Pedder, D. J., and Lavicza, Z. (2011). The influence of school orientation to learning on teachers' professional learning change. *Sch. Eff. Sch. Improv.* 22, 193–214. doi: 10.1080/09243453.2011.572078

Oxford, R. L. (2003). "Toward a more systematic model of L2 learner autonomy" in *Learner Autonomy Across Cultures: Language Education Perspectives*. eds. D. Palfreyman and R. C. Smith (London: Palgrave Macmillan UK), 75–91.

Papamitsiou, Z., and Economides, A. A. (2019). Exploring autonomous learning capacity from a self-regulated learning perspective using learning analytics. *Br. J. Educ. Technol.* 50, 3138–3155. doi: 10.1111/bjet.12747

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educ. Psychol. Rev.* 16, 385–407. doi: 10.1007/s10648-004-0006-x

Posnanski, T. J. (2002). Professional development programs for elementary science teachers: an analysis of teacher self-efficacy beliefs and a professional development model. *J. Sci. Teach. Educ.* 13, 189–220. doi: 10.1023/A:1016517100186

Postholm, M. B. (2012). Teachers' professional development: a theoretical review. *Educ. Res.* 54, 405–429. doi: 10.1080/00131881.2012.734725

Qi, Y., Wang, X., and Wu, X. (2020). Teachers' professional development motivation and influencing factors: an exploration based on job characteristics. *Chin. J. Clin. Psych.* 28, 779–782. (in Chinese).

Sanders, W. L., Wright, S. P., and Horn, S. P. (1997). Teacher and classroom context effects on student achievement: implications for teacher evaluation. *J. Pers. Eval. Educ.* 11, 57–67. doi: 10.1023/A:1007999204543

Sebotsa, T., Beer, J. D., and Kriek, J. (2019). "Self-directed learning and teacher professional development: an adapted profile of implementation", in: *Proceedings of Teaching and Education Conferences*. (London: International Institute of Social and Economic Sciences).

Shao, Z. (2013). Investigation and research on self-directed learning of junior high school teachers. Doctoral, Southwest University. (in Chinese).

Shulman, L. (2011). Knowledge and teaching: foundations of the new reform. *Harv. Educ. Rev.* 57, 1–23. doi: 10.17763/haer.57.1.j463w79r56455411%JHarvardEducationalReview

Shulman, L. S., and Shulman, J. H. (2009). How and what teachers learn: a shifting perspective. *J. Curr. Stud.* 189, 1–8. doi: 10.1177/0022057409189001-202

Smith, R. C. (2003). "Teacher education for teacher-learner autonomy", in: *Symposium for Language Teacher Educators: Papers From Three IALS symposia: IALS*, University of Edinburgh), 1–13.

Smith, R., and Erdogan, S. (2008). "Teacher-learner autonomy: Programme goals and student-teacher constructs," in *Learner and teacher autonomy: Concepts, realities and responses*. eds. T. Lamb and H. Reinders (Philadelphia, PL: Benjamins Publishing Company), 83–102.

Smith, R., Ralston, N. C., Naegele, Z., and Waggoner, J. (2020). Team teaching and learning: a model of effective professional development for teachers. *Prof. Educ.* 43, 80–90.

Stockdale, S. L., and Brockett, R. G. (2011). Development of the PRO-SDLS: a measure of self-direction in learning based on the personal responsibility orientation model. *Adult Educ. Q.* 61, 161–180. doi: 10.1177/0741713610380447

Sun, L. (2012). Research on the status of middle school Teachers' autonomous learning and its influencing factors——Taking Tianjin as an example. Doctoral, Northeast Normal University. (in Chinese).

Wang, P. (2019). Research on the quality standards of compulsory education teachers based on policy texts. *J. Chin. Soc. Educ.* 11, 84–88. (in Chinese).

Wang, G., Huang, W., Wu, L., and Wei, Q. (2019). The double helix model of teachers' key competencies and abilities. *Curr. Teach. Material Method* 39, 132–138. (in Chinese)

Wang, P., and Si, X. (2011). Research on teacher resource allocation in the process of urban and rural education integration——taking Shanxi Province as an example. *J. Shanxi Normal Univ. (Phil. Soc. Sci. Edn.)* 40, 156–161. (in Chinese).

Wang, X., and Wu, Z. (2019). Structural contradiction and reform strategies of rural Teachers' professional title evaluation and appointment. *J. Chin. Soc. Educ.* 9, 70–74. (in Chinese).

Williamson, S. N. (2007). Development of a self-rating scale of self-directed learning. *Nurse Res.* 14, 66–83. doi: 10.7748/nr2007.01.14.2.66.c6022

Ye, L. (2001). *A New Probe Into Teachers' Role and Teachers' Development*. Beijing, China: Educational Science Publishing House. (in Chinese).

Yurdakul, C. (2017). An investigation of the relationship between autonomous learning and lifelong learning. *Int. J. Educ. Res. Rev.* 2, 15–20. doi: 10.24331/ijere.309968

Zhu, L., and Jiang, L. (2021). Research on the current situation of primary and secondary school teachers' learning ability development and its promotion strategies. *J. Contin. High. Educ.* 34, 48–54. (in Chinese).

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fpsyg.2022.998586

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Widodo Winarso,
Institut Agama Islam Negeri Syekh Nurjati Cirebon, Indonesia
Shu Zhang,
Beijing Normal University, China

*CORRESPONDENCE
Peijie Jiang
peijiejiang@hunnu.edu.cn
Bin Xiong
bxiong@math.ecnu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 20 July 2022
ACCEPTED 18 October 2022
PUBLISHED 03 November 2022

CITATION
Jiang P, Zhang Y, Jiang Y and Xiong B (2022) Preservice mathematics teachers' perceptions of mathematical problem solving and its teaching: A case from China. *Front. Psychol.* 13:998586. doi: 10.3389/fpsyg.2022.998586

Preservice mathematics teachers' perceptions of mathematical problem solving and its teaching: A case from China

Peijie Jiang[1]*, Yong Zhang[2], Yanyun Jiang[3] and Bin Xiong[4,5]*

[1]School of Mathematics and Statistics, Hunan Normal University, Changsha, China, [2]School of Mathematics, Yunnan Normal University, Kunming, China, [3]The High School Attached to Hunan Normal University, Changsha, China, [4]School of Mathematical Sciences, East China Normal University, Shanghai, China, [5]Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice, Shanghai, China

Preservice mathematics teachers' accurate understanding of mathematical problem solving and its teaching is key to the performance of their professional quality. This study aims to investigate preservice mathematics teachers' understanding of problem solving and its teaching and compares it with the understanding of in-service mathematics teachers. After surveying 326 in-service mathematics teachers, this study constructs a reliable and valid tool for the cognition of mathematical problem solving and its teaching and conducts a questionnaire survey on 26 preservice mathematics teachers. Survey results reveal that preservice mathematics teachers have a good understanding of mathematical problem solving and its teaching and are more confident in the transfer value of problem solving ability. By contrast, in-service teachers are more optimistic that problem solving requires exploration, continuous thinking, and the participation of metacognition. This article concludes that preservice mathematics teachers should focus more on the initiative and creativity of students and put students at the center of education. In addition, teacher educators should provide more teaching practice opportunities for preservice teachers. The findings also show that in-service teachers' understanding of problem solving and its teaching is inferior to that of preservice teachers on some indicators, implying the importance of post-service training for in-service teachers.

KEYWORDS

pre-service teacher, perceptions, problem solving, mathematics teaching, in-service teacher

Introduction

Mathematics learning is essential in cultivating the philosophical thinking and rational spirit of students (Jiang and Xiong, 2021a). Many countries, including China, attach great importance to mathematics education (Jiang and Xiong, 2021b). Problems are the heart of mathematics. Mathematics is developed and perfected by constantly solving various problems, and the mathematician's main reason for existence is to solve problems (Mason, 2016). Therefore, mathematics consists of problems and solutions. Learning how to solve mathematical problems is at the heart of mathematics education (Polya, 2002a). Solving mathematical problems helps students better understand mathematics (Munzar et al., 2021). To a large extent, mathematics education cultivates the problem solving abilities of students (Polya, 1990). Learning to solve problems and think mathematically requires continuous reflection on the nature of this activity (Arcavi et al., 1998). Therefore, the problem solving teaching ability of mathematics teachers is vital. Nowadays, the relevance of problem solving in teaching and learning mathematics has become commonplace (Libedinsky and Soto-Andrade, 2016).

When teachers invite students to solve a problem, they do not know as much about how to solve the problem as many people think (Andrews and Xenofontos, 2015). Improving problem solving requires focusing on some recommendations, mainly for teachers and their education (Zimmermann, 2016). Teachers' knowledge of teaching content affects their classroom practice, which involves student learning and achievement (Peterson et al., 1989). Problem solving is getting from where you are to where you want to be by continuously reformulating the problem until it becomes something you can manage (Kilpatrick, 2016). The cognition of mathematical problem solving affects mathematical problem solving and teaching behavior. Teachers' beliefs about mathematics impact their teaching, and teachers with different views about mathematics teach differently (Philipp, 2007; Lester and Cai, 2016). Teachers are central to advancing the affective atmosphere and social interaction of the class (Pehkonen et al., 2016), and their beliefs have a considerable impact on the nature of classroom practice (Kayan Fadlelmula and Cakiroglu, 2008). Therefore, teachers need to have a good understanding of mathematical problem solving and its teaching.

Preservice mathematics teachers are prospective teachers who will teach mathematics after graduation (Jiang and Jiang, 2022). Many preservice teachers complete advanced mathematics courses with a limited interpretation of critical terms, incorrect beliefs about the nature of mathematics, and a failure to recognize that mathematics stimulates analytical thinking and creativity (Paolucci, 2015). They have difficulty raising and solving problems (Işık and Kar, 2012; Mallart et al., 2018). Teacher education can alleviate negative attitudes or beliefs about mathematics and teaching mathematics to college students preparing to become teachers (Looney et al., 2017). Preservice mathematics teachers with proper training will have better problem solving and problem solving teaching performance (Crespo and Sinclair, 2008; Karp, 2010). They need a teacher preparation program that focuses their attention on the learning of the students they are teaching (Kilpatrick, 2016). They must understand mathematics, teaching, and pedagogy (Register et al., 2022). Proper understanding of mathematical problem solving and its teaching is an essential part of the professional quality of preservice mathematics teachers and can effectively guide their future teaching of mathematical problem solving.

Significant advances have been made in understanding the affective, cognitive, and metacognitive aspects of problem solving in mathematics and other disciplines (Lester and Cai, 2016). Research on the correlation between the use of various problem solving strategies and problem solving success has been plentiful over the last century (Schoenfeld, 2007), and there have been many suggestions for teaching problem solving effectively (Mason, 2016). However, empirical studies of preservice teachers' understanding of mathematical problem solving and its teaching are still rare. The big question facing current mathematical problem solving research and teaching practice is this: How do we make meaningful problem solving a regular feature of mathematics classrooms (Leong et al., 2016)? It is necessary to train many outstanding mathematics teachers, and a feasible method to achieve this is to pay attention to the education of preservice teachers. We should understand the mathematical problem solving and teaching knowledge of preservice teachers. On the basis of this knowledge, we can develop educational strategies to improve their problem solving teaching skills.

Improving the mathematical problem solving teaching ability of teachers requires understanding the mathematical problem solving and teaching perception of in-service and preservice teachers. For high-quality mathematics (problem solving) teaching, preservice teachers think that "developing students' thinking ability" and "mathematical communication ability" is more critical. By contrast, in-service teachers think "learning arrangement" and "building connections" are more important (Clooney and Cunningham, 2017). Age and work experience may shape beliefs related to mathematical problem solving (Metallidou, 2009). It is helpful for the training of preservice teachers to understand how in-service teachers view mathematical problem solving and its teaching. Therefore, comparing the cognition of preservice and in-service teachers toward mathematical problem solving and its teaching is necessary. The cognition of preservice teachers can be better understood by placing them in the background of in-service teachers' perceptions. Issues such as whether their perceptions have something in common, what the differences are, how to narrow the gap, how to further optimize their perceptions, which perceptions can be optimized before they are employed, and which can only be optimized afterward are not only significant for the training of preservice teachers but also help the continuing training of in-service ones.

Mathematical problem solving teaching in China is relatively successful, and Chinese students perform very well in international mathematics competitions, such as the International Mathematical Olympiad (Xiong and Jiang, 2021). A study of 495 Chinese preservice mathematics teachers showed that their beliefs about mathematics teaching are most correlated with their inquiry-based teaching practices (Yang et al., 2020). However, as Cai and Nie (2007) pointed out, mathematical problem solving research in China has been much more content- and experience-based than cognitive- and empirical-based. Empirical studies of problem solving related to preservice mathematics are scarce. Therefore, China's experience is worth examining. The present study aims to answer the following questions:

1. What do preservice mathematics teachers know about mathematical problem solving and its teaching?
2. What is the difference between preservice and in-service mathematics teachers' cognition of mathematical problem solving and its teaching?

Establishing a model and related tools to study preservice teachers' understanding of mathematical problem solving has important implications for future work in mathematics education and teacher education.

Literature review

Mathematical problem solving was once a hot research issue in mathematics education (Lester, 1994; Schoenfeld, 2007). The content covered in the literature review below includes mathematical problem solving, preservice mathematics teachers, and problem solving for preservice mathematics teachers. This section expounds on the background and starting point of this study from these three aspects.

Mathematical problem solving

Problems generally refer to stimulating situations that cannot be responded to with ready-made responses and require that specific barriers be overcome between the given information and the goal. Mathematical problems can generally be divided into problems for construction and problems to prove (Polya, 1945). Schoenfeld (1985) pointed out that mathematical problems are usually divided into two categories: the common practice problem and the problem that students must experience before solving.

George Polya, the founder of the mathematical problem solving theory, described problem solving as follows (Polya, 1962, p. v): "Solving a problem means finding a way out of a difficulty, a way around an obstacle, attaining an aim which was not immediately attainable." In general, solving problems that require effort and exploration improves one's thinking skills. Polya's problem solving table provides a perspective of the problem solving process from four stages: *Understanding the Problem, Devising a Plan, Carrying out the Plan,* and *Looking Back* (Rosiyanti et al., 2021). What Polya proposed was a heuristic approach, and his theories of mathematical problem solving were far-reaching.

Lester (1994) systematically reviewed the research on mathematical problem solving in terms of research content and methods from 1970 to 1984. He summarized the current state of problem solving research and recommended future studies. Schoenfeld (2013) proposed a decision theory based on his four-factor framework for mathematical problem solving. Research on mathematical problem solving is vibrant. The connotation and requirements of mathematical problem solving ability are also constantly changing. There have been many research results on the technology of mathematical problem solving, creativity in mathematical problem solving, and emotion and aesthetics in mathematical problem solving (Amado et al., 2018). In terms of research methods, the research methods of mathematical problem solving include differential analysis, thinking aloud, correlation analysis, and teaching experiments (Bao and Zhou, 2009).

Problem solving is a significant learning activity, but the quality of problem solving teaching needs to be improved. As Lester and Charles (1992) once pointed out, research on mathematical problem solving provides little specific information on problem solving learning. The role of the teacher in teaching is neglected, and little attention is paid to what happens in real classrooms. The research focuses on the individual, on the theory but not on the class, which is the shortcoming of current problem solving teaching research. Lester (2013) reflected on the study of mathematical problem solving teaching and made four assertions: (1) We need to rethink what mathematical problems and problem solving are. (2) We know very little about how to improve students' metacognition through problem solving. (3) Mathematics teachers need not be problem solving experts but must be serious problem solving students. (4) Mathematical problem solving is not always a high-level cognitive activity.

It is worth mentioning that problem solving is a core activity in Chinese mathematics teaching and learning. Mathematical problem solving activities in China have the following characteristics: (1) high level of reasoning, often involving multi-step and complex formal mathematical reasoning; (2) high comprehensive knowledge, with general mathematical problems involving multiple knowledge points; (3) high operation requirements, including high symbolic calculus ability; (4) simple background, in which more emphasis is placed on the connections within mathematics; and (5) presence of various solutions and high problem solving skills. The problem solving characteristics mentioned above are affected by examinations and courses on the one hand and classroom teachings on the other, such as emphasizing "Bianshi" teaching, question-type training, and reduction methods (Bao, 2017).

In conclusion, despite the many research results, we still know very little about how to develop the metacognitive abilities of students through the teaching of mathematical problem solving. The role of teachers in problem solving learning has been neglected. To improve the problem solving ability of students, we must pay attention to the part of teachers. The problem solving teaching ability of mathematics teachers is also crucial. In addition, some empirical research supported by exact facts and data is necessary to study mathematical problem solving teaching.

Preservice mathematics teachers

Increasing attention is being paid to the training of preservice mathematics teachers. Preservice teachers refer to those who want to become teachers. In this study, preservice mathematics teachers are "quasi" mathematics teachers, including undergraduate students in Mathematics education and master's students in curriculum and teaching theory or mathematics teaching (Tong and Yang, 2018; Jiang and Jiang, 2022). The research on preservice mathematics teachers consists of studies on the current situation, curriculum and instruction, skill training, and teaching knowledge research. Many researchers have studied the mathematical knowledge and related beliefs of preservice mathematics teachers as well as their curriculum and teaching knowledge (Pamuk and Peker, 2009; Prescott et al., 2013; Dede and Karakus, 2014; Paolucci, 2015; Lutovac, 2019). Some researchers agree that preservice mathematics teachers must be better prepared for future mathematics teaching (Land et al., 2015; Lau, 2021).

In many countries, preservice mathematics teachers are taught mathematics and mathematics pedagogy in the mathematics department and education department of their universities. The training of preservice mathematics teachers requires interdisciplinary cooperation (Beers and Davidson, 2009; Hodge, 2011; Goos and Bennison, 2018). Some researchers pointed out that the teaching skills of preservice mathematics teachers need to be strengthened (Özgen and Alkan, 2014; Land et al., 2015; Gokalp, 2016). Many researchers (Llinares and Valls, 2009; Özmantar et al., 2010; Baki et al., 2011; Hechter et al., 2012; Caniglia and Meadows, 2018; Saralar et al., 2018) are also concerned about the ability of preservice mathematics teachers to use Information and Communication Technology (ICT).

Compared to their Western counterparts, Chinese preservice mathematics teachers are more familiar with traditional mathematics thinking and teaching and are not competent in TPACK (Xiang and Ning, 2014). After more than 100 years of development, China's formal preservice teacher education has formed some unique models and characteristics (Tong and Yang, 2018). The training program for preservice mathematics teachers in China has two distinct features (Li et al., 2008). The first is that it lays a solid mathematical foundation for normal students to have a higher mathematical literacy. The second is that it pays attention to the review and research of elementary mathematics because everyone believes that a deep understanding of elementary mathematics and solid problem solving ability are fundamental to becoming qualified middle school mathematics teachers. Under the test-oriented education tradition, qualified teachers must have high problem solving skills.

However, many questions about preservice mathematics teachers have not been effectively addressed. The study of how to train preservice mathematics teachers efficiently remains an essential topic in education worldwide.

Problem solving of preservice mathematics teacher

Although some studies do reveal the characteristics of preservice mathematics teachers in problem solving, such studies are not rich enough. Preservice mathematics teachers have many deficiencies in problem solving. In terms of problem solving skills, most preservice mathematics teachers have low problem solving skills (Özgen and Alkan, 2012). For instance, regarding questioning skills, preservice mathematics teachers commit seven types of errors in asking questions about fractional splitting (Işık and Kar, 2012). Some preservice mathematics teachers have difficulty asking questions about everyday life, fitting into the school curriculum at a given educational level, and posing questions that students can self-correct (Mallart et al., 2018). Preservice mathematics teachers can employ problem solving strategies and problem solving, but their use of different techniques is limited (Avcu and Avcu, 2010). Likewise, they have difficulty expressing operations in the mathematical language (Özdemir and Çelik, 2021). Thus, they need a better understanding of problem solving and its teaching through teaching practice.

Preservice mathematics teachers have difficulty choosing specific mathematical problems, and it is rather difficult for them to find unconventional mathematical problem situations (Temur, 2012). Therefore, many problems in problem solving teaching for preservice mathematics teachers need to be discussed in depth. Zsoldos-Marchis (2015) studied the effectiveness of collaborative problem solving in changing the attitudes of preservice elementary teachers toward mathematics. They found that students who used cooperative learning methods had statistically significant positive changes in their enjoyment of mathematics. The students improved their belief in the usefulness of mathematics, preferring to solve unconventional problems. Other researchers explored the learning process of preservice mathematics teachers taking part in a middle school mathematics methods curriculum, noting the need for further research on such programs (Gómez, 2009). Some researchers have tried to paint a picture of problem solving in Chinese mathematics education, revealing the knowledge of Chinese preservice mathematics teachers in problem solving and teaching (Cai and Nie, 2007), but the research is far from extensive. To summarize, research on problem solving and problem solving teaching for preservice mathematics teachers is necessary. Evidence from classrooms, students, and frontline teachers is significant.

Materials and methods

Research methods

Methods

A study in Turkey used a survey research method to understand how Turkish preservice primary school mathematics teachers perceive problem solving in mathematics education (Kayan Fadlelmula and Cakiroglu, 2008). The current study adopted the questionnaire survey method as it can be used to investigate the knowledge of Chinese mathematics teachers on mathematical problem solving and its teaching. First, the researchers designed a preliminary questionnaire based on the literature and then examined the problem solving and teaching knowledge of in-service mathematics teachers through the online teaching and research community (WeChat and QQ) to develop research tools and establish norms. The online survey instrument is *WENJUANXING* (a widely used platform for publishing online questionnaires in China). The researchers presented links to the questionnaires through WeChat and QQ, and in-service mathematics teachers willing to participate in the survey could click the links to fill out the questionnaires. The teachers were fully informed of the purpose of the study, and filling out the online questionnaire was voluntary. Data generated by in-service teachers served as the norm and could be used to develop and validate the validity of research tools. Structural modeling was used to deal with the data. After the research tools were formed, representative preservice mathematics teachers were selected to conduct a questionnaire survey. According to research ethics requirements, the questionnaire survey was conducted only after the preservice mathematics teachers signed the informed consent.

Participants

Chinese mathematics teachers attach great importance to problem solving, must solve many problems, and regard problem solving as one of the most critical teaching tasks (Xiong and Jiang, 2021). This study is part of a more extensive study. In the larger study, the researchers explored how to design curriculum instruction to facilitate the development of problem solving instructional skills among preservice mathematics teachers. Therefore, it is essential to specify the sample for the study. A total of 199 in-service mathematics teachers effectively participated in the online survey, most of whom came from high schools of good quality in Guangxi, China. Therefore, they could serve as representatives of excellent teachers in the province. As they joined the online mathematics teaching and research group independently, these teachers could be considered as having a high interest in mathematics education. They already have some teaching experience and all have been engaged in the teaching of mathematical problem solving. In addition, 127 mathematics competition coaches from all over China also completed the questionnaire online. Chinese mathematics competition coaches have high problem solving skills and problem solving teaching skills.

The participants in this study came from a local key normal university in China. A total of 26 full-time first-year graduate students, 2 males and 24 females, were selected. Two majored in mathematics curriculum and teaching theory and 24 majored in mathematics teaching. They are preservice mathematics teachers and will all be engaged in mathematics teaching after graduation. Most of them have studied mathematics courses such as Mathematical Analysis, Advanced Algebra, Modern Algebra, Functions of Real Variables, Functions of Complex Variables, and Topology at the undergraduate level. They have studied practical courses such as Mathematics Instructional Design and Teaching Skills Training and have experience in micro-teaching and educational practice. However, they are still students and do not have enough teaching experience yet.

Procedure

Questionnaire surveys are one of the most popular data-gathering methods in the social sciences. This study demonstrates the development and use of questionnaires according to the purpose of the study. The basic process of this study was as follows: conduct literature review→design research methods→conduct online surveys on in-service mathematics teachers→design survey tools based on online surveys→conduct surveys on preservice mathematics teachers→collect and analyze data→report results. The study first investigated the problem solving and teaching cognition of in-service mathematics teachers. The researchers developed valid questionnaires based on literature and data on in-service teachers. Then they contacted participants to ask if they would like to participate. With the participants' consent, the researchers committed to protecting their privacy, distributed the electronic questionnaires to them, and technically assisted them in completing the questionnaires. After the participants completed the questionnaires, the researchers retrieved the data. Once the data of preservice mathematics teachers were collected, they were compared with the data of in-service mathematics teachers.

Data collection and analysis

As mentioned earlier in section "Methods," this study collected data through the Internet and surveyed preservice mathematics teachers by using questionnaires. Therefore, the research data can reveal the knowledge of in-service and preservice teachers about mathematical problem solving and its teaching. After the original network data were collected, the data were preprocessed to obtain valid data. The researchers carried out structural equation modeling based on the valid data and then designed a survey tool for preservice mathematics teachers. Descriptive statistics were carried out on the data of both in-service and preservice mathematics teachers, and the similarities and differences between the teachers were compared. The researchers then performed inferential statistics to arrive at more general conclusions. This study used SPSS 22.0 and AMOS 22.0 to process the data. AMOS 22.0 was used to build a structural equation model for understanding mathematical problem solving and its teaching. The maximum likelihood method was used to estimate the model.

Research tools

Beliefs are part of the affective domain of an individual that influences the learning process (Manderfeld and Siller, 2019). Understanding young students' emotional factors and beliefs about mathematics is a complex task (Giaconi et al., 2016), as is understanding teachers' perceptions of mathematical problem solving and its teaching. Many research frameworks have been built on mathematical beliefs (Hannula, 2011, 2012). The teaching beliefs of mathematics teachers refer to their orientations toward teaching mathematics, which involve perspectives regarding instructional activities, the cognitive processes of students, and the purpose of mathematics (Wang et al., 2022). Currently, the transmissive and constructive taxonomy of teaching beliefs is commonly used in existing studies and international assessments, such as TEDS-M (Blömeke and Kaiser, 2014). Unfortunately, the framework for teachers' perception of problem solving and its teaching is relatively rare.

In this research, understanding mathematical problem solving and its teaching refers to the knowledge and essential viewpoints about mathematical problem solving and its teaching. It is the overall reflection of individual minds on mathematical problem solving and teaching. This study characterizes preservice mathematics teachers' cognition of mathematical problem solving and its teaching from three aspects: overall impression of mathematical problem solving, specialized knowledge, and teaching perspective. In particular, impression is the perceptual image of problem solving in the individual's mind, specialized knowledge refers to the individual's rational understanding of problem solving, and teaching perspective is the attitude and orientation of problem solving teaching. The preparation of measurement items is mainly based on students' typical mathematical beliefs and Polya's theories on mathematical problem solving and teaching (Polya, 1945). It is refined and synthesized according to research needs.

Designing a questionnaire means creating valid and reliable questions that address the research objectives, placing them in proper order, and selecting an appropriate administration method. The design of the corresponding measurement items is mainly based on the five-factor cognitive framework of Schoenfeld and is refined and synthesized according to research needs (Schoenfeld, 2016). For example, the five items s8–s12 in the measurement tool correspond to Schoenfeld's five-factor cognitive framework for mathematical problem solving (the coding and content of the item will be mentioned below). Typical mathematical beliefs held by some of the students mentioned by Schoenfeld above are shown in **Table 1**, and these beliefs are imperfect.

Take the teaching viewpoint of mathematical problem solving as an example. Two items are set up to reflect the orientation of preservice mathematics teachers to the teaching of mathematical problem solving. The two items are "Problem solving requires independent thinking; try not to let students discuss with one another" and "Teaching students to solve problems is to tell students the solution." Since problem solving requires independent thinking, cooperative learning promotes problem solving (teaching). Teaching students to solve problems introduces students to problem solving and inspires them to think. The formulation of the above two items is not considered the correct understanding.

Each item is set on a five-point Likert scale: strongly disagree, disagree, neutral, agree, strongly agree. There are positive and negative items. The score is based on the correctness of the options. Each option is assigned the order of 1, 2, 3, 4, 5, and each option of the reverse item is given the order

TABLE 1 Typical student beliefs about the nature of mathematics (Schoenfeld, 1992).

Code	Belief
1	Mathematical problems have one and only one correct answer.
2	There is only one correct way to solve any mathematical problem—usually the rule the teacher has most recently demonstrated to the class.
3	Ordinary students cannot expect to understand mathematics because they wish to memorize it and apply what they have learned mechanically but without understanding.
4	Mathematics is a solitary activity done by individuals in isolation.
5	Students who have understood the mathematics they have studied will be able to solve any assigned problem in five minutes or less.
6	The mathematics learned in school has little or nothing to do with the real world.
7	The formal proof is irrelevant to the processes of discovery or invention.

of 5, 4, 3, 2, 1. The options of the positive item are assigned as 1, 2, 3, 4, and 5 in turn, and the options of the negative item are designated as 5, 4, 3, 2, and 1. The higher the score, the higher the understanding of mathematical problem solving and teaching. Of course, the so-called correct refers to the point of view with a specific basis, in line with the primary trend.

The entire measurement questionnaire initially included 20 measurement items, and after expert judgment removed 8 duplicate items, the final questionnaire consisted of 12 items. The minimum score of the questionnaire is 12 points and the maximum is 60 points. Considering the characteristics of the five-point-option Likert scale, to better analyze the data, the cognitive level is divided into four grades according to the score, of which 48–60 is graded A, 36–47 is graded B, 24–35 is graded C, and 12–23 is graded D.

According to the score characteristics, the researchers set B and above as the qualification level to present results clearly, with A set as an excellent level. Table 2 shows the number of measurement items and the coding of the three secondary indicators of understanding mathematical problem solving and its teaching.

The items of the entire questionnaire are as follows:

- s1. Problems in problem solving refer to the exercises in textbooks, which generally have conventional and mechanized solutions.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s2. Problem solving is calculating the result according to the conventional algorithm.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s3. The four problem solving stages are understanding the problem, making a plan, implementing the plan, and reviewing. After getting the correct answer, you don't have to spend too much time on the retrospective stage.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s4. Problem solving is a skill that has little meaning after a student graduates.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s5. Problem solving requires independent thinking; try not to let students discuss with one another.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s6. Teaching students to solve problems is to teach students ready-made solutions.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s7. Problem solving generally requires mobilizing mathematical thinking methods through exploration, association, and reasoning.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s8. A good problem solver must have a good knowledge structure.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s9. Emotional attitudes have an important impact on problem solving.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s10. Excellent problem solvers can constantly adjust their thinking to solve problems.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s11. Good problem solvers use heuristics well and experiment with different strategies.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree
- s12. Excellent problem solvers have a good sense of experience.
  A. Strongly disagree B. Disagree C. Neutral
  D. Agree E. Strongly agree

Results

Reliability and validity

The cognitive questionnaire for mathematical problem solving and teaching is based on literature, and its content validity has been discussed above. A total of 203 questionnaires were returned, of which 199 were valid questionnaires. Since the questionnaire has 12 items, the ratio of the number of valid questionnaires returned to the number of items in the questionnaire exceeds 10:1, which can ensure the validity of the model significance test (Jiang et al., 2022).

The Cronbach's alpha coefficient of the questionnaire is 0.770. The internal consistency is good because the questionnaire has only 12 items. The KMO value of the questionnaire is 0.798, which makes it suitable for factor analysis. From Table 3, the average score for the test teachers (199 mathematics teachers) is 47.73; they are in the B class but very close to the A class. Their intermediate perception of problem solving and teaching is above passable and very close to

TABLE 2 Problem solving and its teaching cognition measurement framework.

Target	Indicators	Number of items	Code
Cognition	Overall impression	4	s1, s2, s3, s4
	Specialized knowledge	6	s7, s8, s9, s10, s11, s12
	Teaching perspective	2	s5, s6

TABLE 3 Basic statistics for the recognition of in-service teachers.

Statistics	Value
Average	47.73
Median	48.00
Standard deviation	5.356
Minimum	33
Max	60
Upper quartile	51.00
Lower quartile	45.00

good. Their minimum score is 33 (C), a failing knowledge level. Their highest score is 60, with a standard deviation of 5.356. The upper quartile is 51, the median is 48, and the lower quartile is 45. The perceptions of problem solving and teaching of half of the teachers reached an excellent level, and the understanding of problem solving and teaching of more than 75% of the teachers was above the qualified group.

Figure 1 illustrates the frequency distribution of scores, which is approximately an inverted bell. The distribution of scores reflects the characteristics of a normal distribution, and most scored between 43–53. Although the views of in-service mathematics teachers are not necessarily correct, their knowledge of problem solving and its teaching is considered representative. Their views can represent mathematics teachers' relatively proper understanding of problem solving and its teaching. The data of the test teachers provide a reference standard for problem solving and their teaching knowledge level, which can be compared with the subsequent analysis of preservice mathematics teachers.

The chi-square value of the original model after modification is 45.185, and the significance probability value is $p = 0.589 > 0.05$. The model is in good agreement with the sample data. Among them, RMR value is 0.027 < 0.050, GFI value is 0.966 > 0.900, AGFI value is 0.944 > 0.900, PGFI value is 0.594 > 0.500, and NFI value, RFI value, IFI value, TLI value and CFI value are all greater than 0.900. Therefore, the model is adapted.

Since factor loadings in the range of 0.30–0.40 are considered to meet the minimum requirements for explanatory structure (Hair et al., 2009), as shown in Figure 2, most factor loadings in the model are higher than 0.50. In addition, only four factor loading values are lower than 0.50 but higher than 0.40. Although factor loading values greater than 0.50 are generally considered to be of practical significance, the factor loading values of this model all reached the acceptable minimum requirements.

In the above model, the correlation between errors has practical significance because there is a correlation between the corresponding items. Routine and mechanics in item s1 have the same meaning as the regular algorithm in item s2, though the former emphasizes what a problem is. By contrast, the latter emphasizes problem solving, but the two are related. Items s3 and s4 are mainly related in that these two viewpoints are refuted by Polya, the founder of mathematical problem solving theory. The two viewpoints often appear simultaneously and are widely known. The correlation between items s9 and s12 is mainly reflected in their emphasis on non-intellectual factors.

In this study, the questionnaire survey data of 127 problem solving experts were added to the original data of 199 in-service primary and secondary school teachers. These data were obtained through questionnaires initiated by the national middle school mathematics competition coaches on WeChat and QQ groups. This study conducted a confirmatory factor analysis on the questionnaire based on these 326 data.

The results showed that the Cronbach's alpha coefficient of the questionnaire is 0.757, and the reliability of the questionnaire is acceptable. The KMO value of the questionnaire is 0.810, and the questionnaire has good structural validity. The chi-square value of model fit obtained by maximum likelihood estimation is 63.277, and with the significance probability value $p = 0.069 > 0.05$, the model works the sample data. On the indicators of model adaptation, RMR value is 0.027 < 0.050, GFI value is 0.969 > 0.900, AGFI value is 0.949 > 0.900, PGFI value is 0.596 > 0.500, and NFI value (0.947), RFI value (0.928), IFI value (0.987), TLI value (0.982), and CFI value (0.987) are all greater than 0.900. Thus, all values meet the standard of model fitting.

Except for two factor loading values in the model lower than 0.5 (0.42 and 0.45), the other factor loading values are not lower than 0.5, thus meeting the minimum model adaptation requirements. Overall, the questionnaire has acceptable reliability, content validity, and construct validity and can be used to investigate the level of problem solving and teaching knowledge among preservice mathematics teachers.

The above 199 in-service primary and secondary school teachers represent general primary and secondary school mathematics teachers as well. Their data provide a reference standard for mathematical problem solving and teaching cognition for this study.

The single-sample Kolmogorov–Smirnov normality test of the scores of each item and the total score of the 199 middle school teachers in the test are shown in Table 4. The distribution of the scores of each item and the total score of the test teachers does not obey the normal distribution. The subsequent test of the score should use a non-parametric test.

Table 5 presents the cognition of the in-service teachers. The following "more than half" (coded as D) means the proportion is in the interval [50%, 60%), "majority" (coded as C) indicates the ratio is in the interval [60%, 75%), "most" (coded as B) means the proportion is in the range [75%, 90%), and "almost all" (coded as A) represents the proportion is in the interval [90%, 100%).

For items s1, s2, s3, s5, and s6, the majority of the test teachers reached the correct understanding level; for items s4, s9, and s12, most of the test teachers reached the right

10.3389/fpsyg.2022.998586

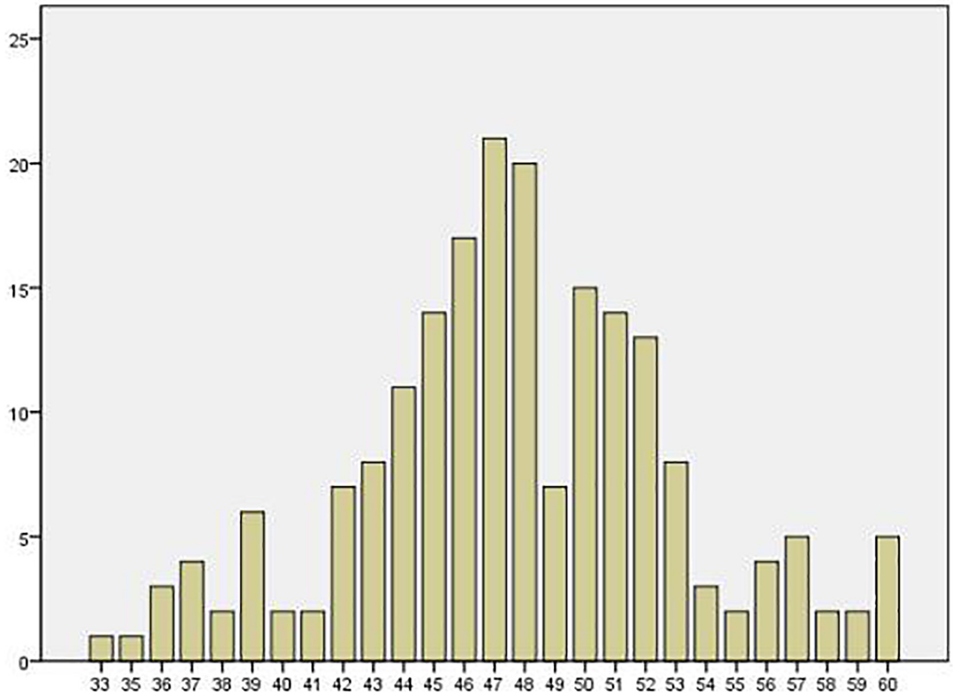


FIGURE 1
Bar graph of in-service teacher scores.

understanding level; and for items s7, s8, s10, and s11, almost all teachers tested reached the level of proper cognition.

Descriptive analysis

Twenty-six questionnaires were sent out in this study and 26 were returned. Therefore, all 26 were valid questionnaires. Recovered data were coded and analyzed using SPSS22. The Cronbach's alpha coefficient of the questionnaire is 0.741, and considering that the questionnaire only had 12 items, the reliability of the questionnaire is acceptable. The average score of the participants is 48.08 points. The intermediate of the participants' cognition of mathematical problem solving and teaching reached a reasonable level. The lowest score is 32, highest is 54, standard deviation is 4.30, upper quartile is 51, median is 48, and lower quartile is 47. Half of the participants understood mathematical problem solving and teaching well. More than 75% of the participants have an understanding above the qualified level and close to the excellent level. One participant has a cognition score of 32, which is not in the qualifying group.

It can be seen from Table 6 that most preservice mathematics teachers have a correct understanding of items s1, s2, s3, s5, s6, s8, and s12. Almost all preservice mathematics teachers have reached the right understanding level for s4, s7, s9, s10, and s11.

The specific options for preservice mathematics teachers are shown in Table 7, which presents their views on each item. Most participants favor collaborative problem solving and believe that teaching students to solve problems is not just about giving them solutions. Most participants felt that common practice problems in textbooks have a different meaning than mathematical problems and that problem solving is not the process of practicing regular algorithms. They attach great importance to the role of problem solving retrospectives and believe that retrospectives are a crucial stage of problem solving. They think that a good knowledge structure is an essential foundation for problem solving and that proper problem solving exercises are necessary. Almost all participants believe that the problem solving process requires mathematical thinking and logical methods and that strategies are crucial in the mathematical problem solving process. They pay attention to emotion and attitude in problem solving, affirm the critical role of metacognitive monitoring and regulation in problem solving, and believe that problem solving skills still have value after graduation.

Difference analysis

The skewness of preservice mathematics teachers (−2.19) was higher than that of in-service mathematics teachers (−0.23), and their scores were skewed to the left. Their average score

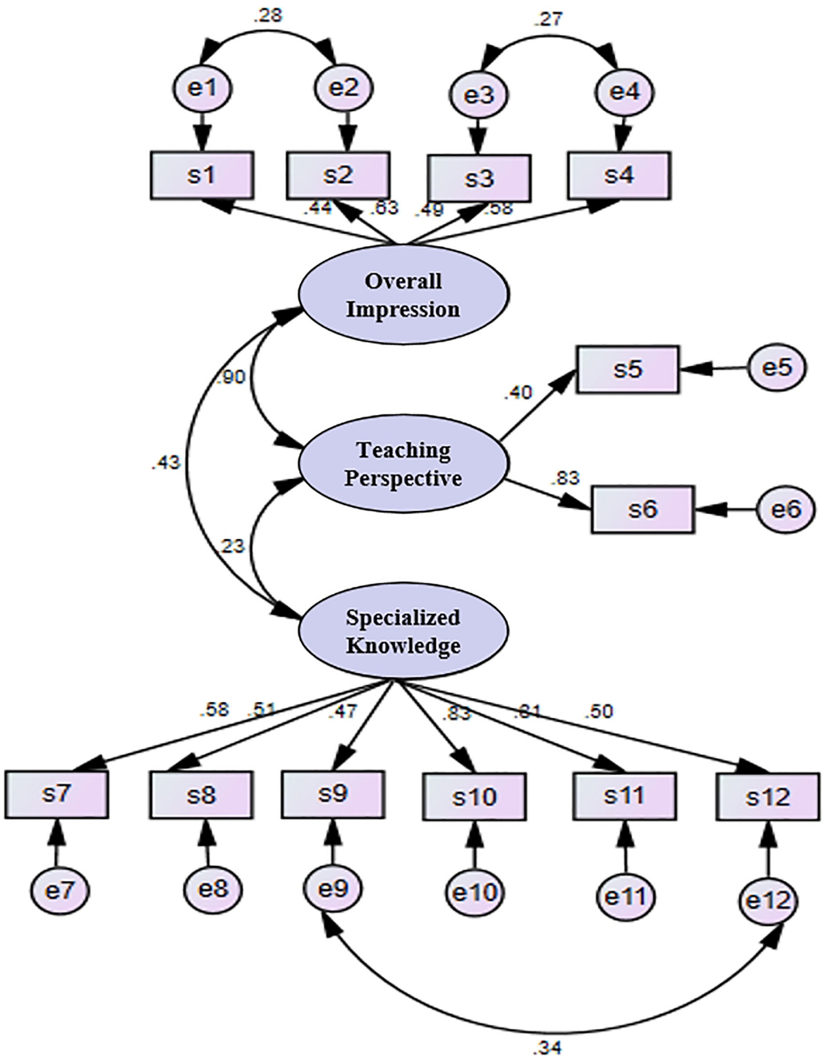


FIGURE 2
Cognitive model of problem solving and its teaching.

TABLE 4 One-sample Kolmogorov–Smirnov normality test.

	s1	s2	s3	s4	s5	s6	s7	s8	S9	s10	s11	s12	Total
p	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.005
Sig.	*	*	*	*	*	*	*	*	*	*	*	*	*

*Significant at the 0.05 level (two-tailed).

TABLE 5 Cognition of in-service teachers.

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
Cognition (%)	64.3	60.8	70.2	83.4	60.8	60.8	94.5	93.0	82.4	97.5	98.0	84.0
Code	C	C	C	B	C	C	A	A	B	A	A	B

(48.08) was slightly higher than that of in-service mathematics teachers (47.73). Their overall awareness of mathematical problem solving and teaching was higher than that of in-service mathematics teachers. The maximum score of the preservice mathematics teachers (54) was lower than that of the in-service mathematics teachers (60), and their minimum score (32) was

TABLE 6 Cognition of preservice mathematics teachers.

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
Percentage	80.7	88.5	84.6	92.3	65.3	69.2	92.3	88.4	98.3	96.2	100	88.4
Code	B	B	B	A	C	C	A	B	A	A	A	B

TABLE 7 Percentage of specific options for preservice mathematics teachers (%).

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
Strongly Agree	3.8	0.0	0.0	0.0	3.8	7.7	11.5	34.6	19.2	34.6	26.9	11.5
Agree	3.8	3.8	3.8	3.8	7.7	11.5	80.8	53.8	73.1	61.5	73.1	76.9
Neutral	11.5	7.7	11.5	3.8	23.1	11.5	3.8	7.7	7.7	3.8	0.0	1.5
Disagree	69.2	80.8	65.4	42.3	61.5	57.7	3.8	3.8	0.0	0.0	0.0	0.0
Strongly Disagree	11.5	7.7	19.2	50.0	3.8	11.5	0.0	0.0	0.0	0.0	0.0	0.0

TABLE 8 Non-parametric independent samples tests (Mann–Whitney *U* test).

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	Total
P	0.484	0.066	0.101	0.030	0.790	0.857	0.019	0.414	0.412	0.040	0.008	0.211	0.316
Sig.				*			*			*	*		

*Significant at the 0.05 level (two-tailed).

lower than that of the test teachers (33). Still, their scores were less volatile (standard deviation: 4.30 < 5.36). Preservice mathematics teachers (51.00, 48.00) were the same as in-service mathematics teachers (51.00, 48.00) in the upper quartile and median scores. However, in the lower quartile, their scores (47.00) were slightly higher than those of in-service mathematics teachers (45.00). Preservice mathematics teachers had a slightly better distribution of scores.

In the scores and total scores of s1, s2, s3, s4, and s9, preservice mathematics teachers were higher than in-service mathematics teachers. They scored lower than in-service mathematics teachers on other items. The scores of each item of preservice and in-service mathematics teachers were concentrated around 4 points. The scores of each item of in-service mathematics teachers were concentrated explicitly in 2, 3, 4, and 5. By contrast, the scores of each item of preservice mathematics teachers were concentrated explicitly in 3, 4, and 5 points. It can also be seen that the distribution of preservice mathematics teachers' scores was good.

The distribution of preservice mathematics teachers' scores did not meet the normality requirement (Kolmogorov–Smirnov normality test, $p = 0.00 < 0.05$). To describe the differences in cognitive level between the subjects and the test teachers in more detail, a non-parametric independent sample test (Mann–Whitney U rank-sum test) was used to analyze the issues. The score distribution of the teacher was tested, and the results are shown in Table 8. Asymptotic significance (two-tailed) is shown in the table at a significance level of 0.05.

There was no statistically significant difference in total scores between preservice and in-service mathematics teachers ($p = 0.316 > 0.05$). Their scores for s1, s2, s3, s5, s6, s8, s9, and s12 were not significantly different (p = 0.484, 0.066, 0.101, 0.790, 0.857, 0.414, 0.412, 0.211). There were significant differences in their scores for s4, s7, s10, and s11 (p = 0.030, 0.019, 0.040, 0.008).

On item s4, "Problem solving is a skill of little significance after students graduate," preservice teachers scored significantly higher than did in-service teachers. They were more convinced that problem solving skills still work after graduation.

On item s7, "Problem solving generally requires mobilizing mathematical thinking methods to go through the process of exploration, association, and reasoning," the scores of preservice teachers were significantly lower than those of in-service teachers, and preservice teachers failed to fully realize that mathematical problem solving is multi-dimensional mathematical thinking that requires the participation in exploration and inquiry. On item s10, "Excellent problem solvers can constantly adjust their thinking to solve problems," preservice teachers scored significantly lower than did in-service teachers. They did not fully recognize the metacognitive monitoring and regulating effect of mathematical problem solving. On item s11, "Excellent problem solvers use heuristics well and experiment with different strategies," preservice teachers scored significantly lower than did in-service teachers on their awareness of the use of heuristics and strategies in problem solving.

There was no significant difference between preservice and in-service teachers in their knowledge of mathematical problem solving and its teaching. Still, preservice teachers scored slightly higher on average. Preservice teachers are more aware that problem solving skills are still valid after graduation. At the same time, they know that they need to participate in mathematical thinking, actively explore, mobilize corresponding strategies, and apply metacognitive monitoring and adjustment in mathematical problem solving and its teaching.

Discussion

First, this study builds a research framework, develops research tools based on existing literature, and tests the reliability and validity of the research framework and research tools. The researchers then use the developed research tools to investigate the mathematical problem solving and teaching perceptions of preservice mathematics teachers and compare the survey results with those of in-service teachers. Finally, the researchers discuss the data results.

Preservice mathematics teachers recognize that problem solving skills are transferable and remain helpful after they graduate. They believe that solving problems requires a process of exploration and effort and fully affirm the importance of emotional attitude in problem solving. They think good problem solvers can constantly adjust their thinking and direction to solve problems as well as use heuristics and other strategies. Problem solving training for preservice mathematics teachers is the core content of mathematics teacher education (dos Santos Morais, 2020). This idea may benefit the good mathematical problem solving and teaching knowledge of preservice teachers. Preservice teachers have positive beliefs about solving math problems, and their views are consistent with the current movement for reform in mathematics education (Kayan Fadlelmula and Cakiroglu, 2008). At the same time, the lack of awareness displayed by in-service teachers relative to preservice mathematics teachers means they need on-the-job training to avoid getting lost in their teaching practice. Some research results point out that some in-service teachers have an insufficient understanding of mathematics teaching, which negatively impact their mathematics teaching (Setoromo and Hadebe-Ndlovu, 2020). The behaviors and attitudes of teachers toward teaching and learning and their knowledge base are the results of the influence of on-the-job training (Ramatlapana, 2009). The participation of in-service teachers in training can facilitate their communication with their peers, help them obtain new information, and update their understanding (Izci and Göktas, 2017). Teacher educators can improve the effectiveness of in-service teacher training by teaching content knowledge orientation and stimulating collaboration among teachers (Selter et al., 2015).

Preservice mathematics teachers also understand the meaning of problems and problem solving, the importance of problem solving review, knowledge structure, and practical problem experience. They recognize the importance of independent thinking for problem solving but ignore the value of collaborative problem solving. Compared with in-service teachers, some preservice teachers think that teaching students to solve problems equates to telling students ready-made solutions. However, as Polya advises mathematics teachers, the best way for students to learn mathematics is to discover it independently (Polya, 1945). Preservice teachers must focus more on the initiative and creativity of students and put them at the center (Polya, 2002b). Many teachers are usually in the early stages of their careers, teaching in ways that are not consistent with their beliefs about teaching (Beswick, 2012). If they want to connect belief to practice, then they need to think about practice (Schoenfeld, 2003). They can change their beliefs about mathematics and its teaching by focusing on how students learn and think about mathematics (Hough et al., 2006). Teachers obtain knowledge by reflecting on the goal accomplishment, learning process, and thinking approach of the students; the efficiency of media; and the recommendation of experts (Sudejamnong et al., 2014). Hence, preservice teachers need to acquire further knowledge in teaching practice. It is not enough that they know what to do but do not know why. Preservice mathematics teachers need to understand the nature of educational philosophy.

Research indicates that many preservice teachers show anxiety about teaching mathematics (Steele et al., 2013) and that teacher training programs have little effect on their beliefs (Dede and Karakus, 2014). However, the results of the current study show no significant difference between preservice and in-service mathematics teachers in their mathematical problem solving and teaching knowledge. Preservice teachers are more confident of the transfer value of problem solving ability, while in-service teachers are more confident that problem solving requires exploration and continuous thinking. According to Polya (2002), the founder of the problem solving theory, teaching students to solve problems introduces them to thinking. The problem solving process is a constant thinking process. Preservice teachers may need to experience more problem solving practices and problem solving teaching practices to appreciate this concept more deeply. They should have the opportunity in their studies to solve (and pose) appropriate problems of similar—sometimes also more demanding—types they use as a teacher later on at school (Zimmermann, 2016). The education and professional development of preservice mathematics teachers span diverse backgrounds, each of which legitimizes different perspectives on mathematics teaching (Ramdhany et al., 2018). Learning by experience is of essential importance in everyday life and the academic field (Fritzlar, 2016). Preservice mathematics teachers need to learn in practice and summarize their experiences independently.

The cognitive framework for mathematical problem solving and its teaching proposed in this paper is valid. Its reliability and validity have been tested and can be used for research. The findings provide a framework for studying

10.3389/fpsyg.2022.998586

preservice mathematics teachers' understanding of problem solving and its teaching. The research results confirm that preservice mathematics teachers understand mathematical problem solving and teaching well. Furthermore, knowing the similarities and differences in understanding between preservice and post-service teachers will help train preservice mathematics teachers better. The findings of this study confirm that the education of preservice mathematics teachers in China, especially at the postgraduate level, is effective in the perception of problem solving teaching (Li et al., 2008). At the same time, the education of preservice teachers should strengthen teaching practices to increase their experience in dealing with practical teaching problems. Every study has its limitations, and this research is no exception. Specifically, the research framework is too simple, the items do not involve an understanding of ICT, the number of items in the questionnaire is too small, and the sample is from only one provincial key normal university. Therefore, the conclusions of this study are not generally applicable. Scientific and careful sampling can minimize the bias caused by selection (Walters, 2021). In addition, teachers with different levels of professional knowledge have different pedagogical focuses on the functions and beliefs of mathematics and demonstrate different teaching methods (Zhang, 2022). Drawing implications for educational practice from the comparison of preservice and in-service mathematics teacher in this study is not straightforward. Future studies should design a richer and more in-depth research framework, design survey tools with higher reliability and validity, and adopt more scientific sampling methods.

Conclusion

Compared with in-service mathematics teachers, preservice mathematics teachers pursuing postgraduate studies have a good understanding of mathematical problem solving and its teaching. Preservice and in-service teachers share much of the same cognition toward mathematical problem solving and its teaching. From the situation of 26 preservice mathematics teachers, it is believed that the training of China's postgraduate-level preservice mathematics teachers has been successful. The difference is that preservice teachers acquire this knowledge through school learning while in-service teachers form more understanding through teaching practice. Moreover, the post-employment knowledge of in-service teachers may cover some of the correct knowledge when they were preservice teachers, which means that on-the-job training throughout their entire career is essential. Preservice mathematics teachers develop well in the cognition that can be established through theoretical study but are insufficient in the knowledge that can only be found through practice. Therefore, more practical opportunities must be created for them. These opportunities to practice include educational traineeships and internships but should not be limited to this, extending toward, for example, consideration of the learning opportunities ICT provides.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Human Subject Protection Committee of East China Normal University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

PJ was the primary author of this research. YZ reviewed the manuscript and made suggestions for revision. YJ analyzed and processed the data for this research. BX gave the overall guidance for this research. All authors contributed to the article and approved the submitted version.

Funding

This study was mainly funded by a grant from the Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice, 18dz2271000. In addition, Hunan Provincial Teaching Reform Research Project (HNJG-2022-0051) supported the cost of proofreading services in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Amado, N., Carreira, S., and Jones, K. (2018). "Broadening research on mathematical problem-solving: An introduction," in *Broadening the Scope of Research on Mathematical Problem Solving: A Focus on Technology, Creativity and Affect*, eds N. Amado, S. Carreira, and K. Jones (New York, NY: Springer International Publishing), 1–12. doi: 10.1007/978-3-319-99861-9_1

Andrews, P., and Xenofontos, C. (2015). Analysing the relationship between the problem-solving-related beliefs, competence and teaching of three Cypriot primary teachers. *J. Math. Teach. Educ.* 18, 299–325. doi: 10.1007/s10857-014-9287-2

Arcavi, A., Kessel, C., Meira, L., and Smith, J. (1998). "Teaching mathematical problem solving: An analysis of an emergent classroom community," in *Research on Collegiate Mathematics Education. III*, eds A. Schoenfeld, J. Kaput, and E. Dubinsky (Washington, DC: American Mathematical Society, Providence), 1–70. doi: 10.1090/cbmath/007/01

Avcu, S., and Avcu, R. (2010). Preservice elementary mathematics teachers' use of strategies in mathematical problem solving. *Procedia Soc. Behav. Sci.* 9, 1282–1286. doi: 10.1016/j.sbspro.2010.12.321

Baki, A., Kosa, T., and Guven, B. (2011). A comparative study of the effects of using dynamic geometry software and physical manipulatives on the spatial visualisation skills of preservice mathematics teachers. *Br. J. Educ. Technol.* 42, 291–310. doi: 10.1111/j.1467-8535.2009.01012.x

Bao, J. (2017). "Mathematical problem solving with Chinese characteristics," in *Mathematics Education in China: Tradition and Reality*, ed. J. P. Wang (Nanjing: Jiangsu Phoenix Education Press), 187–210.

Bao, J., and Zhou, C. (2009). *The Psychological Basis and Process of Mathematics Learning*. Shanghai: Shanghai Education Press.

Beers, D. L., and Davidson, E. M. (2009). A learning community for preservice elementary teachers: A collaboration between mathematics and education. *Primus* 19, 519–540. doi: 10.1080/10511970802067574

Beswick, K. (2012). Teachers' beliefs about school mathematics and mathematicians' mathematics and their relationship to practice. *Educ. Stud. Math.* 79, 127–147. doi: 10.1007/s10649-011-9333-2

Blömeke, S., and Kaiser, G. (2014). "Theoretical framework, study design and main results of TEDS-M," in *International Perspectives on Teacher Knowledge, Beliefs and Opportunities to Learn*, eds S. Blömeke, F. J. Hsieh, G. Kaiser, and W. H. Schmidt (New York, NY: Springer), 19–48. doi: 10.1007/978-94-007-6437-8_2

Cai, J., and Nie, B. (2007). Problem solving in Chinese mathematics education: Research and practice. *ZDM Math. Educ.* 39, 459–473. doi: 10.1007/s11858-007-0042-3

Caniglia, J., and Meadows, M. (2018). Preservice mathematics teachers' use of web resources. *Int. J. Technol. Math. Educ.* 25, 17–34.

Clooney, S., and Cunningham, R. F. (2017). Preservice and inservice mathematics teachers' perspectives of high-quality mathematics instruction. *IUMPST: J.* 2, 1–9.

Crespo, S., and Sinclair, N. (2008). What makes a problem mathematically interesting? Inviting prospective teachers to pose better problems. *J. Math. Teach. Educ.* 11, 395–415. doi: 10.1007/s10857-008-9081-0

Dede, Y., and Karakus, F. (2014). The effect of teacher training programs on pre-service mathematics teachers' beliefs towards mathematics. *Educ. Sci.* 14, 804–809. doi: 10.12738/estp.2014.2.1787

dos Santos Morais, R. (2020). Problem solving as necessary knowledge for teacher training – Mathematics for teaching. *Hist. Mem. Educ.* 11, 239–254. doi: 10.5944/hme.11.2020.23248

Fritzlar, T. (2016). "Memorable Diagonals": Exploratory problems as propositions for doing mathematics," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 157–165. doi: 10.1007/978-3-319-28023-3_10

Giaconi, V., Varas, M. L., Tuohilampi, L., and Hannula, M. (2016). "Affective factors and beliefs about mathematics of young Chilean children: Understanding cultural characteristics," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 37–51. doi: 10.1007/978-3-319-28023-3_3

Gokalp, M. (2016). Investigating classroom teaching competencies of pre service elementary mathematics teachers. *Eurasia J. Math. Sci. Technol. Educ.* 12, 503–512. doi: 10.12973/eurasia.2016.1296a

Gómez, P. (2009). Learning processes in preservice teacher training for secondary school mathematics teachers. *Electron. J. Res. Educ.* 7, 471–498. doi: 10.3389/fpsyg.2022.913536

Goos, M., and Bennison, A. (2018). Boundary crossing and brokering between disciplines in preservice mathematics teacher education. *Math. Educ. Res. J.* 30, 255–275. doi: 10.1007/s13394-017-0232-4

Hair, J. J., Black, W. C., Black, B., Babin, B. J., and Anderson, R. E. (2009). *Multivariate Data Analysis*, 7th Edn. Hoboken: Prentice Hall.

Hannula, M. S. (2011). "The structure and dynamics of affect in mathematical thinking and learning," in *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education*, eds M. Pytlak, T. Rowland, and E. Swoboda (Poland: University of Rzesów), 34–60.

Hannula, M. S. (2012). Exploring new dimensions of mathematics-related affect: Embodied and social theories. *Math. Educ. Res.* 14, 137–161.

Hechter, R. P., Phyfe, L. D., and Vermette, L. A. (2012). Integrating technology in education: Moving the TPCK framework towards practical applications. *Educ. Res. Rev.* 39, 136–152.

Hodge, A. M. (2011). The design and benefits of a teaching seminar for preservice secondary mathematics teachers. *Primus* 21, 263–273. doi: 10.1080/10511970903008949

Hough, S., Pratt, D., and Feikes, D. (2006). "Developing preservice teachers' beliefs about mathematics using a children's thinking approach in content area courses," in *Conference Papers Psychology of Mathematics and Education of North America*, (New York, NY: ACM Digital Library), 1.

Işık, C., and Kar, T. (2012). An error analysis in division problems in fractions posed by pre-service elementary mathematics teachers. *Kuram Ve Uygulamada Egitim Bilimleri* 12, 2303–2309.

Izci, E., and Göktas, Ö (2017). Assessment of in-service training activities for junior high mathematics teachers. *Educ. Res. Rev.* 12, 1220–1229.

Jiang, P., and Jiang, Y. (2022). Online live teaching is effective: An empirical study. *Int. J. Emerg. Technol. Learn.* 17, 73–83. doi: 10.3991/ijet.v17i11.28611

Jiang, P., Jiang, Y., Zhang, X., and Tanu Wijaya, T. (2022). "A survey of Chinese mathematics teachers' views on the educational values of Mathematical Olympiad," in *Proceedings of the 5th International Conference on Big Data and Education*, (New York, NY: ACM Digital Library), 255–260. doi: 10.1145/3524383.3524389

Jiang, P., and Xiong, B. (2021a). Analyze the quality of math kangaroo problems with a content analysis. *J. Phys.* 1875:012015. doi: 10.1088/1742-6596/1875/1/012015

Jiang, P., and Xiong, B. (2021b). "Online live teaching of mathematics gifted students on the tablet," in *2021 the 6th International Conference on Distance Education and Learning*, (New York, NY: ACM Digital Library), 157–161. doi: 10.1145/3474995.3475021

Karp, A. (2010). Analyzing and attempting to overcome prospective teachers' difficulties during problem-solving instruction. *J. Math. Teach. Educ.* 13, 121–139. doi: 10.1007/s10857-009-9127-y

Kayan Fadlelmula, F., and Cakiroglu, E. (2008). *Pre-Service Teachers' Beliefs About Mathematical Problem Solving. Paper Presented at European Conference on Educational Research*. Goteborg: Educational Research.

Kilpatrick, J. (2016). "Reformulating: Approaching mathematical problem solving as inquiry," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 69–81. doi: 10.1007/978-3-319-28023-3_5

Land, T. J., Tyminski, A. M., and Drake, C. (2015). Examining preservice elementary mathematics teachers' reading of educative curriculum materials. *Teach. Teach. Educ.* 51, 16–26. doi: 10.1016/j.tate.2015.05.009

Lau, W. W. F. (2021). Preservice mathematics teachers' professional learning in a pedagogy course: Examining changes in beliefs and confidence in teaching algebra. *Math. Educ. Res. J.* 33, 223–239. doi: 10.1007/s13394-019-00285-y

Leong, Y. H., Tay, E. G., Toh, T. L., Quek, K. S., Toh, P. C., and Dindyal, J. (2016). "Infusing mathematical problem solving in the mathematics curriculum: Replacement units," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 309–325. doi: 10.1007/978-3-319-28023-3_18

Lester, F. (2013). Thoughts about research on mathematical problem-solving instruction. *Math. Enthus.* 10, 245–278. doi: 10.54870/1551-3440.1267

Lester, F. K. (1994). Musings about mathematical problem-solving research: 1970-1994. *J. Res. Math. Educ.* 25, 660–675. doi: 10.2307/749578

Lester, F. K., and Cai, J. (2016). "Can mathematical problem solving be taught? Preliminary answers from 30 years of research," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E.

Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 117–135. doi: 10.1007/978-3-319-28023-3_8

Lester, F. K., and Charles, R. I. (1992). "A framework for research on problem-solving instruction," in *Mathematical Problem Solving and New Information Technologies*, eds J. P. Ponte, J. F. Matos, J. M. Matos, and D. Fernandes (New York, NY: Springer), 1–15.

Li, S., Huang, R., and Shin, H. (2008). "Discipline knowledge preparation for prospective secondary mathematics teachers: An East Asian perspective," in *International Handbook of Mathematics Teacher Education, Volume l: Knowledge and Beliefs in Mathematics Teaching and Teaching Development*, eds P. Sullivan and T. Wood (Rotterdam: Sense Publishers), 63–86.

Libedinsky, N., and Soto-Andrade, J. (2016). "On the role of corporeality, affect, and metaphoring in problem-solving," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 53–67. doi: 10.1007/978-3-319-28023-3_4

Llinares, S., and Valls, J. (2009). The building of preservice primary teachers' knowledge of mathematics teaching: Interaction and online video case studies. *Instr. Sci.* 37, 247–271. doi: 10.1007/s11251-007-9043-4

Looney, L., Perry, D., and Stwck, A. (2017). Turning negatives into positives: The role of an instructional math course on preservice teachers' math beliefs. *Education* 138, 27–40.

Lutovac, S. (2019). Preservice mathematics teachers' narrated failure: Stories of resilience. *Int. J. Educ. Res.* 98, 237–244. doi: 10.1016/j.ijer.2019.09.006

Mallart, A., Font, V., and Diez-Palomar, J. (2018). Case study on mathematics pre-service teachers' difficulties in problem posing. *Eurasia J. Math. Sci. Technol. Educ.* 14, 1465–1481. doi: 10.29333/ejmste/83682

Manderfeld, K., and Siller, H.-S. (2019). Preservice mathematics teachers' beliefs regarding topics of mathematics education. *LUMAT: Int. J. Math. Sci. Technol. Educ.* 7, 65–79.

Mason, J. (2016). "Part 1 reaction: Problem posing and solving today," in *Posing and Solving Mathematical Problems: Advances and New perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 109–113. doi: 10.1007/978-3-319-28023-3_7

Metallidou, P. (2009). Pre-service and in-service teachers' metacognitive knowledge about problem-solving strategies. *Teach. Teach. Educ.* 25, 76–82. doi: 10.1016/j.tate.2008.07.002

Munzar, B., Muis, K. R., Denton, C. A., and Losenno, K. (2021). Elementary students' cognitive and affective responses to impasses during mathematics problem solving. *J. Educ. Psychol.* 113, 104–124.

Özdemir, F., and Çelik, H. C. (2021). Examining problem-solving and problem-posing skills of pre-service mathematics teachers: A qualitative study. *Educ. Q. Rev.* 4:18. doi: 10.31219/osf.io/tp49v

Özgen, K., and Alkan, H. (2012). The relationship between secondary school pre-service mathematics teachers' skills in problem solving dimensions and their learning style characteristics. *Kuram Ve Uygulamada Egitim Bilimleri* 12, 1173–1181.

Özgen, K., and Alkan, H. (2014). An investigation of pre-service mathematics teachers' skills in the development of activities. *Kuram ve Uygulamada Egitim Bilimleri* 14, 1179–1201. doi: 10.12738/estp.2014.3.1866

Özmantar, M. F., Akkoç, H., Bingölbali, E., Demir, S., and Ergene, B. (2010). Pre-Service Mathematics Teachers' Use of Multiple Representations in Technology-Rich Environments. *Eurasia J. Math. Sci. Technol. Educ.* 6, 19–36.

Pamuk, S., and Peker, D. (2009). Turkish preservice science and mathematics teachers' computer related self-efficacies, attitudes, and the relationship between these variables. *Comput. Educ.* 53, 454–461. doi: 10.1016/j.compedu.2009.03.004

Paolucci, C. (2015). Changing perspectives: Examining the potential for advanced mathematical studies to influence preservice teachers' beliefs about mathematics. *Teach. Teach. Educ.* 49, 97–107. doi: 10.1016/j.tate.2015.03.002

Pehkonen, E., Ahtee, M., and Laine, A. (2016). "Pupils' drawings as a research tool in mathematical problem-solving lessons," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 167–188. doi: 10.1007/978-3-319-28023-3_11

Peterson, P. L., Carpenter, T. P., and Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analyses. *J. Educ. Psychol.* 81, 558–569.

Philipp, R. A. (2007). "Teachers' beliefs and affect," in *Second Handbook of Research on Mathematics Teaching and Learning*, ed. F. K. Lester (Reston: National Council of Teachers of Mathematics), 257–315.

Polya, G. (1945). *How to Solve It*. Princeton: Princeton University Press.

Polya, G. (1962). *Mathematical Discovery*. New York, NY: Wiley.

Polya, G. (1990). *Mathematics and Plausible Reasoning, Volume I*. Princeton: Princeton University Press.

Polya, G. (2002a). The goals of mathematical education: Part one. *Math. Teach.* 181, 6–7.

Polya, G. (2002b). The goals of mathematical education: Part two. *Math. Teach.* 181, 42–44.

Prescott, A., Bausch, I., and Bruder, R. (2013). TELPS: A method for analysing mathematics preservice teachers' pedagogical content knowledge. *Teach. Teach. Educ.* 35, 43–50. doi: 10.1016/j.tate.2013.05.002

Ramatlapana, K. (2009). Provision of in-service training of mathematics and science teachers in Botswana: Teachers' perspectives. *J. Math. Teach. Educ.* 12, 153–159. doi: 10.1007/s10857-009-9101-8

Ramdhany, V., Venkat, H., and Christiansen, I. M. (2018). Dissonance and continuity in the Mathematical education and training experiences of pre-service secondary mathematics teachers. *Afr. J. Res. Math. Sci. Technol. Educ.* 22, 186–195. doi: 10.1080/18117295.2018.1474843

Register, J., Fernandes, A., and Pugalee, D. (2022). Supporting preservice mathematics teachers' culturally responsive teaching: A focus on teaching for social justice. *Mathematics* 10:896. doi: 10.3390/math10060896

Rosiyanti, H., Ratnaningsih, D. A., Bahar, H., Iswan, and Faisal. (2021). Application of mathematical problem solving sheets in Polya's learning strategy in social arithmetic material. *Int. J. Early Child. Spec. Educ.* 13, 707–717. doi: 10.9756/INT-JECSE/V13I2.211111

Saralar, I., Isiksal-Bostan, M., and Akyuz, D. (2018). The evaluation of a preservice mathematics Teacher's TPACK: A case of 3D shapes with GeoGebra. *Int. J. Technol. Math. Educ.* 25, 3–21.

Schoenfeld, A. H. (1985). *Mathematical Problem Solving*. Cambridge: Academic Press.

Schoenfeld, A. H. (1992). "Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics," in *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics*, ed. D. A. Grouws (New York, NY: Macmillan Publishing Co, Inc), 334–370.

Schoenfeld, A. H. (2003). How can we examine the connections between teachers' world views and their educational practices? *Issues Educ.* 8, 217–227.

Schoenfeld, A. H. (2007). Problem solving in the United States, 1970–2008: Research and theory, practice and politics. *ZDM Math. Educ.* 39, 537–551. doi: 10.1007/s11858-007-0038-z

Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *Math. Enthusiast* 10, 9–34.

Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *J. Educ.* 196, 1–38. doi: 10.1177/002205741619600202

Selter, C., Gräsel, C., Reinold, M., and Trempler, K. (2015). Variations of in-service training for primary mathematics teachers: An empirical study. *ZDM Math. Educ.* 47, 65–77. doi: 10.1186/s12913-016-1423-5

Setoromo, M., and Hadebe-Ndlovu, B. (2020). An exploration of in-service teachers' understanding of teaching mathematics in grade R classrooms: A case study of grade R in Lesotho University of Kwazulu-Natal. *S. Afr. J. Child. Educ.* 10:9.

Steele, A., Brew, C., Rees, C., and Ibrahim-Khan, S. (2013). Our practice, their readiness: Teacher educators collaborate to explore and improve preservice teacher readiness for science and math instruction. *J. Sci. Teach. Educ.* 24, 111–131. doi: 10.1007/s10972-012-9311-2

Sudejamnong, A., Robsouk, K., Loipha, S., and Inprasitha, M. (2014). Development of teachers' mathematical knowledge for teaching by using the innovation of lesson study and open approach. *Soc. Mind* 04, 317–327. doi: 10.4236/sm.2014.44032

Temur, ÖD. (2012). Analysis of prospective classroom teachers' teaching of mathematical modeling and problem solving. *Eurasia J. Math. Sci. Technol. Educ.* 8, 83–93. doi: 10.12973/eurasia.2012.822a

Tong, L., and Yang, X. (2018). "Pre-service mathematics teacher education," in *The 21st Century Mathematics Education in China New Frontiers of Educational*

Research, eds Y. Cao and F. Leung (New York, NY: Springer), doi: 10.1007/978-3-662-55781-5_17

Walters, W. H. (2021). Survey design, sampling, and significance testing: Key issues. *J. Acad. Librariansh.* 47:102344. doi: 10.1016/j.acalib.2021.102344

Wang, Y., Qin, K., Luo, C., Yang, T., and Xin, T. (2022). Profiles of Chinese mathematics teachers' teaching beliefs and their effects on students' achievement. *ZDM Math. Educ.* 54, 709–720. doi: 10.1007/s11858-022-01353-7

Xiang, K., and Ning, L. (2014). Evaluating Chinese pre-service mathematics teachers' knowledge of integrating technology in teaching. *J. Math. Educ.* 7, 48–58.

Xiong, B., and Jiang, P. (2021). The Chinese experience at the International Mathematical Olympiad. *J. East China Normal Univ.* 2021, 1–14.

Yang, X., Kaiser, G., König, J., and Blömeke, S. (2020). Relationship between pre-service mathematics teachers' knowledge, beliefs and instructional practices in China. *ZDM Math. Educ.* 52, 281–294. doi: 10.1007/s11858-020-01145-x

Zhang, Q. (2022). Understanding Chinese mathematics teaching: How secondary mathematics teachers' beliefs and knowledge influence their teaching in mainland China. *ZDM Math. Educ.* 54, 693–707. doi: 10.1007/s11858-022-01336-8

Zimmermann, B. (2016). "Improving of mathematical problem-solving: Some new ideas from old resources," in *Posing and Solving Mathematical Problems: Advances and New Perspectives*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (New York, NY: Springer International Publishing), 83–108. doi: 10.1007/978-3-319-28023-3_6

Zsoldos-Marchis, I. (2015). Changing pre-service primary-school teachers' attitude towards mathematics by collaborative problem solving. *Procedia Soc. Behav. Sci.* 186, 174–182. doi: 10.1016/j.sbspro.2015.04.100

TYPE Original Research
PUBLISHED 03 November 2022
DOI 10.3389/fpsyg.2022.961702

Organizational support and job crafting with the new math teachers' well-being: The mediating effect of basic psychological needs

Limei Wang[1,2], Fuqiang Peng[3]* and Naiqing Song[1,4]

[1]School of Mathematics and Statistics, Southwest University, Chongqing, China, [2]School of Science, West Yunnan University, Lincang, Yunnan, China, [3]School of Foreign Languages, West Yunnan University, Lincang, Yunnan, China, [4]Centre for Collaborative Innovation of Assessment Toward Basic Education Quality, Southwest University, Chongqing, China

Check for updates

OPEN ACCESS

EDITED BY
Zsolt Lavicza,
Johannes Kepler University of Linz, Austria

REVIEWED BY
Angelica Moè,
University of Padua, Italy
Yiming Cao,
Beijing Normal University, China
Lianchun Dong,
Minzu University of China, China

*CORRESPONDENCE
Fuqiang Peng
pengfuqiang728@163.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 05 June 2022
ACCEPTED 17 October 2022
PUBLISHED 03 November 2022

CITATION
Wang L, Peng F and Song N (2022) Organizational support and job crafting with the new math teachers' well-being: The mediating effect of basic psychological needs. *Front. Psychol.* 13:961702. doi: 10.3389/fpsyg.2022.961702

Enhancement of the teacher well-being level has grown into a general, pressing problem to be solved in the domain of education worldwide. Based on the theoretical perspective of the multi-level dynamically formed mechanical model of occupational well-being, this study initially constructed a mechanism model with the occupational well-being role of organizational support, job crafting, and the occupational well-being of new math teachers at primary and junior high schools, and conducted empirical research using structural equation modeling. The study found out that organizational support, job crafting and basic psychological needs have prominent and positive effects on the occupational well-being of the new math teachers in primary up to junior high schools. It also argues that basic psychological needs may mediate the correlation between organizational support, job crafting, and the occupational well-being of the new math teachers. To sum up, the study findings reveal the mechanisms of the role in organizational support and job crafting on the occupational well-being of new primary up to junior high school math teachers. Also, the findings may be conducive to extending the research on the factors that influence the teacher well-being, notably fostering the study on that in the math teachers in the primary up to junior high schools of China.

KEYWORDS

mathematics teachers, occupational well-being, structural equation modeling, organizational support, job crafting, basic psychological needs

Introduction

Well-being is an ideal state of people's psychological functioning and experience (Ryan and Deci, 2001). Teacher well-being is a response of teachers to their occupation and career from the cognitive, emotional, health, and social aspects (Organization for Economic Co-operation and Development [OECD], 2020), as well as the driving force for development and progress of the teaching community and educational organizations as large (Li and Lu, 2020). The well-being of teachers by profession is first included as a crucial part of the PISA2021 math framework in the OECD report, Teachers' Well-being: A Framework for Data Collection and Analysis (Organization for Economic Co-operation and Development [OECD], 2020). China (2018) has also promulgated the Opinions of the State Council of the Central Committee of the Communist Party of China on Comprehensively Deepening the Reform with the Teachers in the New Era. The country also proposes that "by 2035, the teachers nationwide will enjoy the well-being at the teaching positions, achievements in the cause, and honors amid the society, as an admirable profession". Although there are more countries realizing the essential value of the teacher well-being (Selwyn and Riley, 2015; McCallum et al., 2017; AsiaSociety, 2018; UnitedKingdom, 2018), the teacher well-being at primary and junior high schools generally remains low (Schleicher, 2018) and has led to high teacher-drain rate (Collie and Martin, 2017), serious teacher-shortage (Schleicher, 2018) and little teacher-attractiveness (UnitedKingdom, 2018) at the schools. Relevant statistics indicates that approximately 30% of primary and junior high school teachers in the United States quit within 5 years after their university graduation and the teaching job placement (McCallum et al., 2017); 13 European Union countries are already facing a severe shortage of primary and junior high school teachers (Organization for Economic Co-operation and Development [OECD], 2020); and less than 10% of primary and junior high school teachers in France, Spain and Sweden retain their teaching profession as meritorious and meaningful (Organization for Economic Co-operation and Development [OECD], 2019). These figures showcase an increasingly serious problem of low occupational well-being for primary and junior high school teachers (Schleicher, 2018), and raising the teacher well-being has grown into a common, urgent issue in the educational domain worldwide (Aldrup et al., 2018; Braun et al., 2019; Holzberger et al., 2021). Therefore, it is of great pragmatic significance to explore in depth the influencing factors and their role mechanisms with the occupational well-being of primary and junior high school teachers.

A review of the existing literature reveals that researchers also analyze the factors influencing the teacher well-being from different perspectives. In his study, Warr (1994) finds in his study that the teacher well-being may be affected by age and teaching experience and illustrates a U-shaped curve relevant to the two factors. Mattern and Bauer (2014) also see that rational self-planning could enhance the teacher well-being conquest to some extent. Moè (2016b) confirmed that harmonious passion and teachers enthusiastic, have been found to relate with teacher well-being (Burić and Moè, 2020). It has been proved that when teachers experience wellbeing they are also more motivated (Klusmann et al., 2008; Moè, 2016a; Shoshani and Eldor, 2016). Positive organizational psychology is becoming a heated topic in the study of factors influencing the occupational well-being of primary and junior high school teachers (Schleicher, 2018). Organizational support has received increasing attention as a critical driver of the teacher well-being (Lauermann and König, 2016), promoting the teacher identity, forming positive perceptions and eventually, enhancing the teacher well-being (Wang and Xu, 2008). Different perception levels of the organizational support are related to whether teachers can achieve their psychological satisfaction, which further affects the teacher well-being (Wang and Xu, 2008). Some other researchers focus on the positive role of the organizational support perception in job crafting (Kim et al., 2018). This is believed to have broken through the traditional top-down model designed by the school management for primary and junior high school teachers. Job crafting may enhance their sense of belonging and identity at the schools and in turn, heighten their occupational well-being (Wei et al., 2018). In addition, some researchers discovered that the higher the level of satisfaction of an individual's basic psychological needs is, the more happiness he or she will experience (Wu et al., 2018), and teacher psychological need satisfaction is considered a possible mediator (Moè and Katz, 2020). Teachers' need satisfaction is linked with reappraisal, which is related to the autonomy supportive (Moè and Katz, 2021). In recent years, in attempt to explore the multivariate interaction factors affecting the occupational well-being of primary and junior high school teachers, some researchers have observed a significant, positive correlation between the sense of organizational support plus job crafting and the subjective occupational success in the school teachers (Zhao and Li, 2019), of which, occupational well-being is deemed to be the main expression form of subjective occupational success (Spurk et al., 2019). Based on the aforesaid studies on the well-being of primary and junior high school teachers, researchers would combine organizational support with job crafting. Then, how do organizational support and job crafting affect the occupational well-being of primary and junior high school teachers? What is the role mechanism of basic psychological needs in organizational support, job crafting, and teachers' perceptions of their occupational well-being? It is worthwhile to explore these questions in depth.

Literature review as mentioned earlier facilitates applicable exploration for further study on the teacher well-being at primary and junior high schools. However, the existing studies have not yet fully explained the above issues and still leave room for the following study areas.

Firstly, there is a disconnect with each other between the study on organizational support's facilitation of the teacher well-being and that on the job crafting for the same purpose, with a lack of study on the relationship among the studies on organizational support, job crafting, basic psychological needs, and the teacher well-being, as well as their specific role mechanisms. Some researchers argue that basic psychological needs is a fundamental entry point for the study on the correlation between job crafting and the teacher well-being (Wang et al., 2020), there are still very few studies that introduce basic psychological needs as an intermediate variable to the study on the correlation among organizational support, job crafting, and the teacher well-being.

Secondly, Chinese primary and junior high school teachers also face low levels of the teacher well-being, for which some researchers have analyzed the factors influencing the teacher well-being (Zeng et al., 2021). However, still there is a lack of empirical study on the well-being in the Chinese context.

Thirdly, relevant statistics show that new teachers have relatively low occupational well-being (McCallum and Price, 2010). However, few studies on the teacher well-being have been conducted so far with new teachers at primary and junior high schools. Moreover, there are even much fewer studies focusing on the particular group of teachers of a specific subject at primary and junior high schools (McCallum et al., 2017). Besides, some researchers have also noticed the relatively low happiness of math teachers (Organization for Economic Co-operation and Development [OECD], 2020). Therefore, the study on the well-being of new math teachers is very important.

The contributions of this manuscript are presented in four aspects. First, the multi-level dynamic formation mechanism model of occupational well-being suggests that organizational practices and characteristics have a direct positive influence on occupational well-being, indirectly affecting the well-being by person-environment fit and achieving the well-being through the satisfaction of basic psychological needs. This study intends to explore the correlation between organizational support, job crafting, and the teacher well-being from a multi-level dynamic formation mechanism model of occupational well-being. Second, this study also proposes the hypothesis that basic psychological needs has an intermediary effect on promotion of the teacher well-being via organizational support and job crafting. Third, by constructing a role mechanism model of organizational support, job crafting and the teacher well-being, an empirical study has been conducted using questionnaire data of new math teachers at primary and junior high schools in China. Fourth, the study expands the scope of the study on the teacher well-being from a positive organizational behavioral perspective, opening up new ideas for the government and schools' building and managing the teaching force, and furthering the professional teacher development.

The manuscript is organized as follows. "Literature review and hypotheses development" section presents the theoretical background and review the literature to develop our research hypothesis tested in this study. Next section presents the sample and methodology used to test the hypothesis. "Results" section shows the empirical results, with "Discussion" section discussing these results. Finally, conclusions are made in section "Conclusion".

Literature review and hypotheses development

The organizational support and the teacher well-being

Organizational support is the extent to which an organization values the contribution of its employees and cares about their well-being (Rhoades and Eisenberger, 2002), as well as the overall perception of employees whose work is recognized and whose well-being administered by the employers (Eisenberger and Huntington, 1986). The school teachers' perceptive levels of organizational support are related to a teacher's willingness to work together in behavioral solidarity and to their psychological satisfaction (Eisenberger and Huntington, 1986). Teachers tend to be more satisfied with their profession when they feel supported by their institutions (Tan et al., 2007). And with job satisfaction as a critical factor in well-being, it is clear that increased job satisfaction can significantly enhance occupational well-being (Hur et al., 2016). In addition, some researchers have also found that organizational support to primary and junior high school teachers is remarkably and positively related to various dimensions of well-being, with a significant, positive impact on the teacher well-being (Wang and Xu, 2008). Recently, organizational support has been the strongest predictor of the teacher well-being, enhancing or reducing the effectiveness in terms of psychological capital and emotional exhaustion (Wang et al., 2020). Based on this, the following hypothesis is formulated for this study:

> H1: Organizational support have a remarkably positive effect on the occupational well-being of new math teachers at primary and junior high schools.

The job crafting and the teacher well-being

The essence of job crafting lies in that "employees, on their own initiative, would make bottom-up changes in the content and methods of their work to meet their own needs or those of their groups, for a sense of meaningfulness at work" (Bolino and Grant, 2016; Akkermans and Tims, 2017). Job crafting may help improve the fit between employees and their work environment, enabling the employees to realize their self-worth and create more value for their workplace, thus enhancing their well-being at work (Zhang and Parker, 2019). Furthermore, researchers argue that *t* job crafting may impact positively job satisfaction (Slemp and Vella-Brodrick, 2013; Kim et al., 2018), a significant expression of occupational well-being. Likewise, Wei et al. (2018) argue that teachers' own competence in job crafting would be a key factor to enhance the teacher well-being. Besides, it has also been suggested that there is a positive correlation between job crafting and the employee well-being, and that employees' perceptions of well-being are significantly enhanced from their proactive job redesigning behaviors (Hou et al., 2021). Based on this, the following hypothesis is proposed in this study.

> H2: There was a significant, positive effect of job crafting on the occupational well-being of new math teachers at primary and junior high schools.

The mediating effect of basic psychological needs

Basic psychological needs is derived from the self-determination theory proposed by Deci and Ryan (2012), which states that when people's needs for autonomy are met, they tend to be more actively engaged and creative in their daily activities and be observed to exhibit the individual well-being (Deci and Ryan, 2000). The positive correlation between basic psychological needs and well-being has been affirmed by several studies (Patrick et al., 2007; Wang et al., 2015), which also conclude: the higher level of satisfaction of basic psychological needs in individuals, the greater well-being one may experience (Wu et al., 2018). Within the theoretical framework for the person-environment fit, the self-determination theory expounds the psychological mechanisms by which the fit may affect the well-being at work, and confirms the intermediary role of human psychological need satisfaction (Greguras and Diefendorff, 2009; Yen, 2012; Slemp and Vella-Brodrick, 2013).

Furthermore, organizational support helps improve basic psychological needs (Deci et al., 2016). Also, organizational characteristics (e.g., organizational support) positively or indirectly impact occupational well-being through the person-environment fit resulting from the satisfaction of basic psychological needs (Zou et al., 2015). In addition, some researchers have argued that employees' job crafting have a positive impact on their basic psychological needs (Van Wingerden et al., 2017). Therefore, job crafting can enable employees to meet their psychological needs and increase their perception of happiness. That is, the basic psychological needs are intermediary between job crafting and employee well-being (Hou et al., 2021). Therefore, the study intends to apply the basic psychological requirements of primary and junior high school teachers to the school environment (Huang, 2021), for which the study also develops the following hypotheses:

> H3: Basic psychological needs has a significant positive effect on the occupational well-being of new math teachers at primary and junior high schools.

> H4a: Organizational support has a significant, positive effect on basic psychological needs.

> H4b: Basic psychological needs have a mediating effect in the correlation between organizational support and the occupational well-being of new math teachers at primary and junior high schools.

> H5a: Job crafting has a significant, positive effect on basic psychological needs.

> H5b: The basic psychological needs have a mediating effect in correlation between job crafting and the occupational well-being of new math teachers at primary and junior high schools.

In conclusion, this study construct the theoretical model (see **Figure 1**).

Methodology

Sample and data collection

In this study, we selected new math teachers at primary and junior high schools as the subjects. Then we conducted an online questionnaire survey in which new teachers referred to those who had worked for no more than 3 years, according to the existing literature (Huang, 2021). Before the actual study, the researcher conducted a pre-survey with some new teachers in Yunnan province of China. We revised the questions repeatedly to constitute the formal questionnaire. We mainly took a wide range of samples from the provinces of Yunnan, Guizhou, Guangxi, Guangdong, and Shandong, with a total of 280 questionnaires distributed and 262 returned. As a result, 248 questionnaires were deemed valid, excluding incomplete and invalid ones, with a correct return rate of 88.57%. The results of the demographic analysis of the samples are shown in **Table 1**.

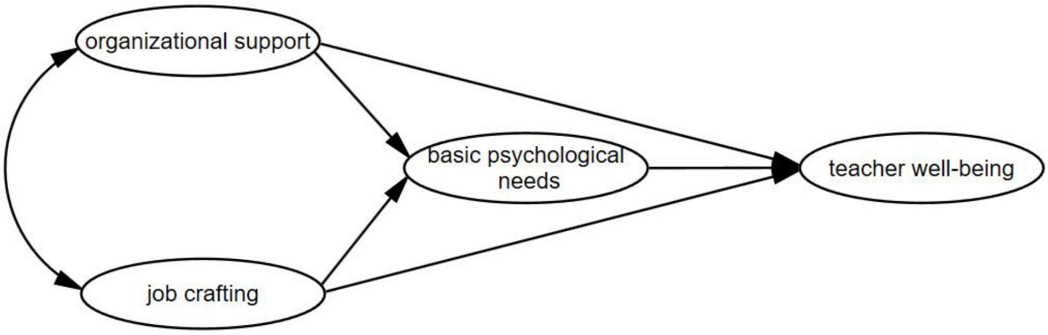


FIGURE 1
The theoretical model.

TABLE 1 Composition of samples (*N* = 248).

Category	Indicator	Time-frequency	Rate-frequency	Category	Indicator	Time-frequency	Rate-frequeny
Gender	Male	87	35.1	Age	Aged 20 or	1	0.4
	Female	161	64.9		Aged 21–23	66	26.6
Teaching experience	≤1 Year	119	48		Aged 24–26	92	37.1
	1–2 Years (excl. 2 Y)	28	11.3		Aged 27–30	23	9.3
	2–3 Years (excl. 3 Y)	16	6.5		Aged 30 or above	66	26.6
	3 Years	85	34.3	Academic qualification	Under junior college	1	0.4
Professional title	Not conferred	145	58.5		Junior college	24	9.7
	Level-III	30	12.1		Undergraduate	222	89.5
	Level-II	41	16.5		Postgraduate	1	0.4
	Level-I	32	12.9	Teaching level	Primary school	168	67.7
School location	County/district/city	61	24.6		Junior high school	80	32.3
	Village/town (ship)	187	75.4				

We took all the samples from primary and junior high school math teachers who had just worked for 3 years or less. 34.8% of the teachers were males and 65.2% females, which conformed with the essential criteria for this study.

Variables and measurement

To ensure the reliability and validity of the measurement instrument, this study intends to apply, as far as possible, well-established scales in the existing literature, with appropriate modifications for the purpose of the measurement questions. Four types of variable scales are stated as follows:

(1) The Organizational Support scale. This study measured the support mainly using the Organizational Support Scale developed by Ling et al. (2006) and Tang (2018), which is found suitable for Chinese primary and junior high school teachers. The scale, which adopts the 5-point Likert scale, has been proven to have good reliability in numerous studies (Gan and Xu, 2018).
(2) The Job Crafting scale. The Slemp and Vella-Brodrick (2013) job crafting questionnaire was used, for example "I will use new skills to improve my work". The 5-point Likert scale has undergone a rigorous process of translation and back-translation, and has been used by many other researchers with good reliability and validity (Liu et al., 2019).
(3) The basic psychological needs scale. The Chinese version of the Basic Psychological Needs Scale in use was proposed by Liu et al. (2013). This scale was revised, following a large number of sampling measurements based on the Gagne Basic Psychological Needs Scale (Gagne, 2003). The revised scale uses the 5-point Likert scale, and has been verified by researchers for reliability and validity (Zhang et al., 2017).
(4) The teacher well-being scale. The Organization for Economic Co-operation and Development [OECD] (2020) first included the Teacher Well-being Scale assessment

in the PISA2021 mathematics framework, providing a holistic, systematic, universally suitable, and scientific conceptual framework and question items for measuring the teacher well-being. Following a rigorous translation and back-translation process, the scale was appropriately adapted and utilized as the teacher well-being scale for this study.

Results

Tests of reliability and validity

SPSS 26.0 was applied to examine the reliability and validity of the scales in this study. The Cronbach's coefficient alpha values for all constructs were more significant than the benchmark value of 0.8 (see **Table 2**), and the deletion of any of the items might not significantly increase the alpha value, which indicates good internal consistency of the questions measured within a single dimension (Nunnally, 1978). The Cronbach'α value of the entire scale in this manuscript reads 0.954 suggesting a high overall reliability.

This study used more mature scales from home and abroad and conducted a pre-survey before the formal distribution of the questionnaire, with the question types revised repeatedly to ensure the accuracy of the questionnaire. This study explored both convergent validity and discriminant validity. Concurrent validity measures whether different items in a scale can reflect the same latent variable and is judged with two indicators: compositional reliability (CR) and average variance extracted (AVE). The judgment criterion, which had been proposed by Fornell and Larcker (1981), was put in use. At CR value of 0.7 or above, the consistency is measured higher, and the value is deemed better. And at an average variance extracted (AVE) of 0.5 or above, the variable has desirable convergent validity. **Table 2** shows that the CR values in this study are more significant than the benchmark value of 0.7 and the AVE values more remarkable than the expected value of 0.5, which indicates that the scale was of satisfactory convergent validity.

Discriminant validity is the presence of a low correlation or significant difference between the potential trait represented by a variable and the trait represented by other variables, obtained

TABLE 2 Testing of reliability and convergent validity.

Variable	Cronbach-α	CR	AVE
Organizational support	0.951	0.953	0.742
Job crafting	0.914	0.916	0.647
Basic psychological needs	0.854	0.860	0.554
New math teachers' well-being at primary and junior high schools	0.846	0.845	0.531

by the square root of the AVE and the correlation between the other variables. In case of a greater square root of the AVE of a variable than the correlation coefficient between that variable and the other variables, discriminant validity between the variables is assessed as good. As shown in **Table 3**, the AVE open root values for each measure are greater than the correlation coefficients between that measure and the other ones. Therefore, the measurement model is readily of good discriminant validity.

Relevant analysis of the variables

The mean value and standard deviation of all the variables in this study were calculated using SPSS 26.0 to work out the correlative coefficients between the variables and obtain the Pearson correlation coefficient matrix (the two-tailed test) for each variable analyzed, which is illustrated in **Table 3**. The results show the significant positive correlations among the variables of institutional support, occupational reshaping, basic psychological needs, and the teacher well-being.

Model fit test

The model fit is the degree of consistency between the theoretical and sample models. The model goodness of fit can be examined using AMOS 24.0. A model fit is assessed as good when the figure is between 0 and 3. As the good fit index GFI, the adjusted AGFI, relative fit index TLI, and comparative fit index CFI exceed 0.9, the closer to 1.0 suggest a better goodness-of-fit between the data and the model, and the greater than 0.8, is an acceptable model. The model is assessed as a good fit given that the variability index RMSEA is less than 0.080 (Wu, 2000).

The testing results are shown **Table 4**, in which = 2.401, GFI = 0.837, AGFI = 0.800, TLI = 0.928, CFI = 0.928, and RMSEA = 0.076. Therefore, the sample model is affirmed to have better goodness of fit.

Path analysis and hypothesis testing

Direct effect testing

Four potential variables and their respective measure items were introduced into the set conceptual data model, and the study hypotheses proposed in the model were tested using the IBM SPSS AMOS 24.0 maximum likelihood estimating approach to examining whether the hypotheses would be supported by the parametric test structure. The final results are presented in **Table 5**, in which the standardized path coefficient of organizational support on the occupational well-being of new math teachers at primary and junior high schools, $\beta = 0.218$ ($t = 4.870$, $P < 0.001$), indicates a significant

TABLE 3 Discriminant validity testing.

Dimension	Organizational support	Job crafting	Basic psychological needs	New math teachers' well-being at primary and junior high schools
Organizational support	**0.861**			
Job crafting	0.544**	**0.804**		
Basic psychological needs	0.679**	0.717**	**0.744**	
New math teachers' well-being at primary and junior high schools	0.663**	0.570**	0.655**	**0.729**

The diagonal bold figure is the arithmetic square root of AVE. **$P < 0.01$.

TABLE 4 Model fit indices.

Fit index	χ^2/df	GFI	AGFI	TLI	CFI	RMSEA
Reference value	<3	>0.8	>0.8	>0.9	>0.9	<0.08
Examined value	2.401	0.837	0.800	0.919	0.928	0.075

positive effect of organizational support on the teacher well-being, by which hypothesis H1 was verified; the standardized path coefficient of job crafting on the aforesaid new math teacher well-being, β = 0.135 (t = 2.023, $P < 0.05$), a significant positive effect of job crafting on the teacher well-being, by which the hypothesis H2 was verified; the standardized path coefficient of basic psychological needs on the teacher well-being, β = 0.209 (t = 2.530, $P < 0.01$), a significant positive effect of basic psychological needs on the teacher well-being, by which the hypothesis H3 was verified; the standardized path coefficient of institutional support on basic psychological needs, β = 0.343 (t = 7.356, $P < 0.001$), a significant positive influence of organizational support on basic psychological needs, by which the hypothesis H4a was tested; and the standardized path coefficient of job crafting on basic psychological needs, β = 0.562 (t = 7.645, $P < 0.001$), a significant positive effect of job crafting on basic psychological needs, by which the hypothesis H5a was verified.

Intermediary effect testing

Based on a simple correlation analysis of the variables, this study used the structural equation modeling to provide insight into the correlation between the variables and thus to examine the intermediary role of basic psychological needs. Bootstrapping was used to sample 5,000 times and 95% confidence intervals were set to test the significance of the mediating effect of organizational support, job crafting and the teacher well-being. Mackinnon et al. (2004) concluded, through experiments, that the bias-corrected method, a non-parametric Bootstrap approach, was optimal. Therefore, only the bias-corrected confidence intervals are reported in this study. With reference to the results of the mediating effects test (see **Table 6**), the intermediary path effect of the organizational support to the occupational well-being of new math teachers at primary and junior high schools was 0.138. So the mediating effect of basic psychological needs was significant (none of the confidence intervals contain 0, $P < 0.05$), and it can be assumed that hypothesis H4b is valid; and the intermediary path effect value of job crafting on the teacher well-being was 0.167, which indicates that the mediating effect of basic psychological needs was significantly present (None of the confidence intervals contain 0, $P < 0.05$), and so hypothesis H5b holds. To sum up, basic psychological needs have a mediating effect in promotion of the teacher well-being with organizational support and job crafting.

After the above path analysis and hypothesis testing, all the influential paths of the model met the criteria of statistical significance. The specific path relationships between the variables of organizational support, job crafting, basic psychological needs and teachers well-being in the research model are shown in **Figure 2**.

Discussion

Compared with prior research findings on the teacher well-being, this study find that in the following three facets.

Firstly, researchers have ever applied the self-determination theory to explain the formation mechanism of occupational well-being (Tims and Bakker, 2010; Yen, 2012; Slemp and Vella-Brodrick, 2013), but there are some overlapping influencing factors on the occupational well-being, such as the interaction with both organizational support and job crafting (Zou et al., 2015). Therefore, this study has found that organizational support and job crafting have a direct positive effect on the occupational well-being of new math teachers at primary and junior high schools, which happens to be consistent with some other researchers' views (Wang and Xu, 2008; Wei et al., 2018). Unlike the previous studies, the study has also found that organizational support and job crafting also indirectly affect teacher well-being through a survey of basic psychological needs. At the same time, this study intended to explore the theory with a multi-level dynamic formation mechanism model of occupational well-being. That may help deepen the understanding of the factors influencing institutional backing and occupational reshaping and thus extend the research on the factors contributing to teacher well-being.

TABLE 5 Path correlation testing.

Hypothesis	Path	Standardized coefficient	Standard error	*t*-value	Significance	Hypothesis tested
H1	Organizational support --→ The teacher well-being	0.218	0.045	4.870	***	YES
H2	Job crafting --→ The teacher well-being	0.135	0.67	2.023	0.043	YES
H3	Basic psychological needs --→ The teacher well-being	0.209	0.083	2.530	0.011	YES
H4a	Organizational support --→ Basic psychological needs	0.343	0.047	7.356	***	YES
H5a	Job crafting --→ Basic psychological needs	0.562	0.074	7.645	***	YES

***$P < 0.001$.

TABLE 6 Intermediary effect testing.

Path	Effect value	95% confidence intervals	*P*
Organizational support --→ Basic psychological needs --→ The teacher well-being	0.138	[0.11, 0.311]	0.035
Job crafting --→ Basic psychological needs --→ The teacher well-being	0.167	[0.05, 0.382]	0.046

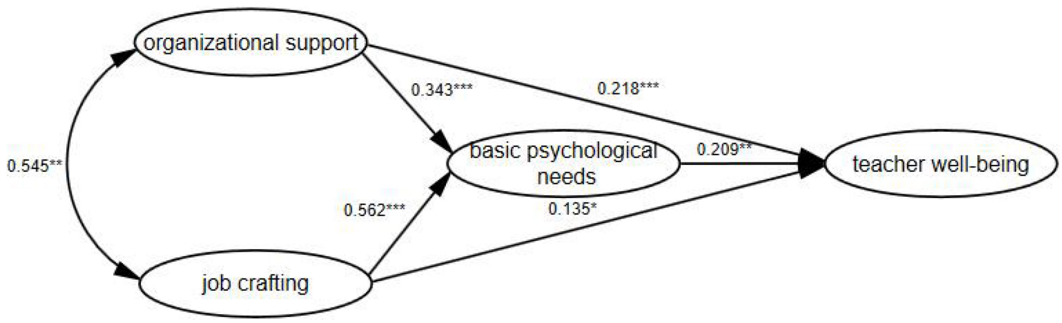


FIGURE 2
The path relationships between the variables of organizational support, job crafting, basic psychological needs, and teachers well-being. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Secondly, Current research on the factors influencing the teacher well-being has focused on work stress (Yang, 2014), psychological capital (Ning, 2017), and job autonomy (Zeng et al., 2021), still leaving room for further studies on the mechanisms influencing the well-being (Wang, 2019). This study found that organizational support, job crafting, and basic psychological needs may contribute to the occupational well-being of new math teachers in primary and junior high schools, with the basic psychological needs mediating effect. These findings could fill the gap in the studies on the factors influencing the occupational well-being of primary and junior high school teachers (Wang, 2019), enriching those on the antecedent variables of the teacher well-being, and eventually, advances on the role of the mechanisms influencing the well-being.

Thirdly, this study also takes into account such factors as the teaching experience of the subjects (Hu and Zhou, 2014). The teacher well-being varies across teaching years and influences different factors (Ling et al., 2016). Besides, it may also change from country to country, in different cultural backgrounds, even in the similar context within China, the teachers may have different levels of well-being (Peng and Yin, 2021). Nonetheless, quite few studies of the teacher well-being have been made with new teachers at primary and junior high schools (Pei and Li, 2015). Moreover, there are fewer studies on teachers instructing a particular subject at primary and junior high schools (McCallum et al., 2017). Therefore, this study investigates the teacher well-being of new math teachers at primary and junior high schools in China to further study teacher well-being.

Conclusion

While applying the theory of multi-level dynamic formation mechanism model of the teacher well-being, this study has conducted an empirical investigation by the structural equation modeling with 248 samples as new math teachers at primary and junior high schools. A preliminary model has been constructed for the roles of organizational support and job crafting in affecting the occupational well-being of the new math teachers at the schools. To this end, the study has found that the correlative roles among the organizational support, job crafting, basic psychological needs and occupational well-being with primary and junior high school teachers, and that there exists a mediating effect of basic psychological needs with organizational support, job crafting, and the teacher well-being.

Concerning practical implications, this study illustrates that organizational support has a significant positive effect on the occupational well-being of the new math teachers at the schools. These teachers' perceptions of the school policies and management, resource support, and humanistic care determine whether they are able to achieve their psychological satisfaction. Furthermore, as their reasonable requests are met and assistance offered promptly upon encountering any problem, the teacher well-being will also be enhanced with higher satisfaction with their school. Therefore, schools should establish appropriate policies to support their teaching and living. For example, schools should understand and help new math teachers to deal with the problems they encounter at work, create a positive working atmosphere for them, and communicate with them before making decisions related to their work; and offer guidance and feedback on their work where necessary. In other words, schools should fully understand the importance of new math teachers' well-being to their schools and practice the well-being specific policies from the perspective of institutional support.

Additionally, as shown in the study, job crafting significantly positively affected the occupational well-being of new math teachers at primary and junior high schools. Also, to meet their needs, based on their own needs and job characteristics, new math teachers can adopt proactive behaviors, to achieve a sustainable person-job fit. They also actively construct their job crafting by re-examining the connections to their tasks, work environment, abilities, and interests and transforming their work. Meanwhile, the schools are advised to develop appropriate job crafting intervention programs considering the specific circumstances of new math teachers to enhance their reshaping capacity. Besides the findings of this study, the schools may also leave some room for bottom-up job crafting by new math teachers, encourage them to exert their initiative, and thus enhance their occupational well-being.

This study remains limitations in several aspects. First of all, the subjectivity of measurement indicators is hardly avoidable. Although its design may reduce bias and errors to a certain extent, this study adopted domestic and international scales and conducted a pilot study before the formal investigation to minimize the impact of subjective errors. However, such errors may still inevitably exist, which will be further minimized in future studies through in-depth interviews based on the Grounded theory and qualitative survey. Secondly, there are quite a few influencing factors on the occupational well-being of new math teachers at primary and junior high schools. This study just investigated three factors, still losing sight of many others, which can be further studied in later studies by introducing possibly sufficient variables from more diverse perspectives. Thirdly, the sample size of this study was only confined to new math teachers at primary and junior high schools in some provinces, which should be extended across the country in China, with a much larger size of samples and rigorously verified error-free data. Additionally, primary and junior high schools are two different contexts. This study does not consider the differences between teachers in further education segments due to time and effort constraints. In future studies, we can consider the differences in the relationship path relationships between the variables of institutional backing, job crafting, basic psychological needs, and teacher well-being. Therefore, the author hopes that future studies on the occupational well-being of new math teachers at primary and junior schools will be further optimized, in both diversity and magnitude.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LW and FP conducted the data analyses, interpreted the results, and drafted the manuscript. NS provided feedback and co-wrote the final submission. LW, FP, and NS conceived the idea of the study, designed the study, contributed to the manuscript revision, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akkermans, J., and Tims, M. (2017). Crafting your career: How career competencies relate to career success via job crafting. *Appl. Psychol.* 66, 168–195. doi: 10.1111/apps.12082

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., and Trautwein, U. (2018). Student misbehavior and teacher well-being: Testing the mediating role of the teacher-student relationship. *Learn. Instr.* 58, 126–136. doi: 10.1016/j.learninstruc.2018.05.006

AsiaSociety (2018). "New Challenges and opportunities facing the teaching profession in public education," in *The 2018 International Summit on the Teaching Profession*, (Lisbon).

Bolino, M. C., and Grant, A. M. (2016). The bright side of being prosocial at work, and the dark side, too: A review and agenda for research on other-oriented motives, behavior, and impact in organizations. *Acad. Manag. Ann.* 10, 599–670. doi: 10.1080/19416520.2016.1153260

Braun, S. S., Roeser, R. W., Mashburn, A. J., and Skinner, E. (2019). Middle school teachers' mindfulness, occupational health and well-being, and the quality of teacher-student interactions. *Mindfulness* 10, 245–255. doi: 10.1007/s12671-018-0968-2

Burić, I., and Moè, A. (2020). What makes teachers enthusiastic: The interplay of positive affect, self-efficacy and job satisfaction. *Teach. Teach. Educ.* 89:103008. doi: 10.1016/j.tate.2019.103008

China (2018). *Opinions of the state council of the central committee of the communist party of china on comprehensively deepening the reform with the teachers in the new era*. Beijing: Xinhua News Agency.

Collie, R. J., and Martin, A. J. (2017). Teachers' sense of adaptability: Examining links with perceived autonomy support, teachers' psychological functioning, and students' numeracy achievement. *Learn. Individ. Differ.* 55, 29–39. doi: 10.1016/j.lindif.2017.03.003

Deci, E. L., and Ryan, R. M. (2000). The "What" and "Why" of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inq.* 4, 226–268. doi: 10.1207/S15327965PLI1104_01

Deci, E. L., and Ryan, R. M. (2012). "Motivation, personality, and development within embedded social contexts: An overview of self-determination theory," in *The Oxford Handbook Of Human Motivationed*, ed. R. M. Ryan (Oxford: Oxford University Press), 85–107. doi: 10.1093/oxfordhb/9780195399820.013.0006

Deci, E. L., Ryan, R. M., Gagne, M., Leone, D. R., Usunov, J., Kornazheva, B. P., et al. (2016). Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: A cross-cultural study of self-determination. *Pers. Soc. Psychol. Bull.* 27, 930–942. doi: 10.1177/0146167201278002

Eisenberger, R., and Huntington, R. (1986). Perceived organizational support. *J. Appl. Psychol.* 3, 500–507. doi: 10.1037/0021-9010.71.3.500

Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104

Gagne, M. E. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement1. *Motiv. Emot.* 27, 199–223. doi: 10.1023/A:1025007614869

Gan, W., and Xu, X. (2018). Organizational support social capital and second-generation of migrant workers' turnover intention. *J. Manag. Sci.* 31, 79–89.

Greguras, G. J., and Diefendorff, J. M. (2009). Different fits satisfy different needs: Linking person-environment fit to employee commitment and performance using self-determination theory. *J. Appl. Psychol.* 94, 465–477. doi: 10.1037/a0014068

Holzberger, D., Maurer, C., Kunina-Habenicht, O., and Kunter, M. (2021). Ready to teach? A profile analysis of cognitive and motivational-affective teacher characteristics at the end of pre-service teacher education and the long-term effects on occupational well-being. *Teach. Teach. Educ.* 100:103285. doi: 10.1016/j.tate.2021.103285

Hou, Y., Wu, Q., and Gui, J. (2021). Research on the relationship between job crafting and employees well-being: a cross-level model based on selfdetermination theory. *Mod. Manag. Sci.* 3, 31–41.

Hu, Y., and Zhou, Y. (2014). Effective ways of promoting the growth of new teachers——case study of four primary school new teachers in the city. *Teach. Educ. Res.* 26, 54–60.

Huang, C. (2021). *Action Research on Teachers' Support of Middle Class Children's Constructive PlayBased on Self-Determination Theory*, Ph.D thesis, Chengdu: Chengdu University.

Hur, E., Jeon, L., and Buettner, C. K. (2016). Preschool teachers' child-centered beliefs: Direct and indirect associations with work climate and job-related wellbeing. *Child Youth Care Forum* 45, 451–465. doi: 10.1007/s10566-015-9338-6

Kim, H., Im, J., and Qu, H. (2018). Exploring antecedents and consequences of job crafting. *Int. J. Hosp. Manag.* 75, 18–26. doi: 10.1016/j.ijhm.2018.02.014

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., and Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *J. Educ. Psychol.* 100, 702–715. doi: 10.1037/0022-0663.100.3.702

Lauermann, F., and König, J. (2016). Teachers' professional competence and wellbeing: Understanding the links between general pedagogical knowledge, self-efficacy and burnout. *Learn. Instr.* 45, 9–19. doi: 10.1016/j.learninstruc.2016.06.006

Li, G., and Lu, L. (2020). PISA 2021 assessment of teachers' occupational wellbeing: Framework and characteristics. *China Exam.* 11, 48–60.

Ling, H., Wang, Y., Zhang, J., and Li, H. (2016). Comparative studies on well-being among novice, proficient and expert teachers. *Chin. J. Clin. Psychol.* 24, 56–59.

Ling, W., Yang, H., and Fang, L. (2006). Perceived Organizational Support(POS) of the employees. *Acta Psychol. Sin.* 38, 281–287.

Liu, J., Lin, L., Yuan, L., Wei, C., Ying, Z., and Chen, X. (2013). Reliability and validity of the chinese version of the basic psychological needs scale. *Chin. Ment. Health J.* 27, 791–795.

Liu, Y., Yang, D., and An, Y. (2019). Research on the relationship between challenging-obstructive stress and work well-being: The mediating effect of job remolding. *Contemp. Econ. Manag.* 41, 77–84.

Mackinnon, D. P., Lockwood, C. M., and Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivar. Behav. Res.* 39, 99–128. doi: 10.1207/s15327906mbr3901_4

Mattern, J., and Bauer, J. (2014). Does teachers' cognitive self-regulation increase their occupational well-being? The structure and role of self-regulation in the teaching context. *Teach. Teach. Educ.* 43, 58–68. doi: 10.1016/j.tate.2014.05.004

McCallum, F., and Price, D. (2010). Well teachers, well students. *J. Stud. Wellbeing* 4, 19–34. doi: 10.21913/JSW.v4i1.599

McCallum, F., Price, D., Graham, A., and Morrison, A. (2017). *Teacher Wellbeing: A review of the literature*. Sydney: Association of Independent Schools of NSW.

Moè, A. (2016a). Does displayed enthusiasm favour recall, intrinsic motivation and time estimation?. *Cogn. Emot.* 30, 1361–1369. doi: 10.1080/02699931.2015.1061480

Moè, A. (2016b). Harmonious passion and its relationship with teacher well-being. *Teach. Teach. Educ.* 59, 431–437. doi: 10.1016/j.tate.2016.07.017

Moè, A., and Katz, I. (2020). Self-compassionate teachers are more autonomy supportive and structuring whereas self-derogating teachers are more controlling and chaotic: The mediating role of need satisfaction and burnout. *Teach. Teach. Educ.* 96:103173. doi: 10.1016/j.tate.2020.103173

Moè, A., and Katz, I. (2021). Emotion regulation and need satisfaction shape a motivating teaching style. *Teach. Teach. Theory Pract.* 27, 370–387. doi: 10.1080/13540602.2020.1777960

Ning, J. (2017). *Research On The Influence Of Psychological Capital On The Happiness Of Primary And Secondary School Teachers, And The Intermediary Role Of Job Identity And Job Satisfaction*, Ph.D thesis, Shanghai: Qufu Normal University.

Nunnally, J. C. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.

Organization for Economic Co-operation and Development [OECD] (2019). *TALIS 2018 Results(Volume1): Teachers And School Leaders As Lifelong Learners*. Berlin: OECD Publishing.

Organization for Economic Co-operation and Development [OECD] (2020). *Teachers' Well-Being: A Framework For Data Collection And Analysis*. Berlin: OECD Publishing, doi: 10.1787/c36fc9d3-en

Patrick, H., Knee, C. R., Canevello, A., and Lonsbary, C. (2007). The role of need fulfillment in relationship functioning and well-being: A self-determination theory perspective. *J. Pers. Soc. Psychol.* 92, 434–457. doi: 10.1037/0022-3514.92.3.434

Pei, M., and Li, X. (2015). A review of overseas research on teacher well-being. *Teach. Educ. Res.* 27, 93–98.

Peng, Z., and Yin, J. (2021). Work remodeling research: Twenty years of review and outlook. *Modernization Manag.* 41, 111–116.

Rhoades, L., and Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *J. Appl. Psychol.* 87, 698–714. doi: 10.1037/0021-9010.87.4.698

Ryan, R. M., and Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annu. Rev. Psychol.* 1, 141–166. doi: 10.1146/annurev.psych.52.1.141

Schleicher, A. (2018). *Valuing our Teachers and Raising their Status: How Communities Can Help*. Berlin: OECD Publishing. doi: 10.1787/9789264292697-en

Selwyn, J., and Riley, S. (2015). *Measuring Well-Being: A Literature Review*. Bristol: University of Bristol.

Shoshani, A., and Eldor, L. (2016). The informal learning of teachers: Learning climate, job satisfaction and teachers' and students' motivation and well-being. *Int. J. Educ. Res.* 79, 52–63. doi: 10.1016/j.ijer.2016.06.007

Slemp, G., and Vella-Brodrick, D. (2013). The job crafting questionnaire: A new scale to measure the extent to which employees engage in job crafting. *Int. J. Wellbeing* 2, 126–146.

Spurk, D., Hirschi, A., and Dries, N. (2019). Antecedents and outcomes of objective versus subjective career success: Competing perspectives and future directions. *J. Manag.* 45, 35–69. doi: 10.1177/0149206318786563

Tan, X., Qin, Q., and Pan, X. (2007). The relationships among employees' perceived organizational support. Job satisfaction and job-quitting intention. *Psychol. Sci.* 2, 441–443.

Tang, J. (2018). *The Research on the Mechanism of Effects of Middle School Teachers' Perceived Organizational Support on Their Job Satisfaction and Stress*, Ph.D thesis, Nanjing: Nanjing University of Aeronautics and Astronautics.

Tims, M., and Bakker, A. B. (2010). Job crafting: Towards a new model of individual Job redesign. *SA J. Ind. Psychol.* 36, 1–9. doi: 10.4102/sajip.v36i2.841

UnitedKingdom (2018). *Teacher Recruitment And Retention Strategies*. Beaverton, OR: Ringgold, Inc, 26.

Van Wingerden, J., Derks, D., and Bakker, A. B. (2017). The impact of personal resources and job crafting interventions on work engagement and performance. *Hum. Resour. Manag.* 56, 51–67. doi: 10.1002/hrm.21758

Wang, J., Hao, X., and Li, Y. (2020). The relationship between organizational support and special education teachers'professional well-being: Analysis of chain mediating effect. *Chin. J. Clin. Psychol.* 28, 1281–1284.

Wang, L., and Xu, C. (2008). Impact of primary and secondary school teachers' perceived organizational support on their happiness and job burnout. *Chin. J. Clin. Psychol.* 16, 574–575.

Wang, L., Tao, T., Fan, C., and Gao, W. (2015). Does psychological need satisfaction perceived online enhance well-being? *Psych J.* 4, 146–154. doi: 10.1002/pchj.98

Wang, W. (2019). *The Primaryand Secondary School Teacher'Soccupational Well-Being: The Structure Andinfluencing Factors*, Ph.D thesis, Hangzhou: Zhejian Gnormal University.

Warr, P. B. (1994). A conceptual framework for the study of work and mental health. *Work Stress* 8, 84–97. doi: 10.1080/02678379408259982

Wei, X., Zhou, Y., and Yang, J. (2018). The impact of job remodeling on the professional happiness of primary school teachers. *Contemp. Educ. Sci.* 5, 67–71.

Wu, C., Rong, S., Zhu, F., Chen, Y., and Guo, Y. (2018). [Basic psychological need and its satisfaction]. *Adv. Psychol. Sci.* 26, 1063–1073. doi: 10.3724/SP.J.1042.2018.01063

Wu, M. (2000). *Structural Equation Model: the operation and application of AMOS (The 2nd Editioned.)*. Chongqing: Chongqing University Press.

Yang, L. (2014). Influence of work stress on primary school teachers' professional happiness: The intermediary role of professional identity. *Chinacadult Educ.* 24, 136–138.

Yen, W. (2012). *Person-Environment Fit: Work-Related Attitudes And Behavioral Outcomes In Continuing Care Retirement Communities*, Ph.D thesis, Ann Arbor: ProQues.

Zeng, L., Wang, Y., Zeng, C., Huang, Y., and Zhao, S. (2021). The influence of job autonomy on primary and secondary school teachers' well-being: The chain mediating effect of job crafting and work family balance. *J. Psychol. Sci.* 44, 705–712.

Zhang, F., and Parker, S. K. (2019). Reorienting job crafting research: A hierarchical structure of job crafting concepts and integrative review. *J. Organ. Behav.* 40, 126–146. doi: 10.1002/job.2332

Zhang, L., Liu, L., Jin, T., and Zhang, L. (2017). Intermediary effect of basic psychological needs in regulating sense of security and online deviant behaviar of college students. *J. Nejiang Normal Univ.* 32, 23–27.

Zhao, X., and Li, F. (2019). Relationship among Perceived Organizational Support, Job Crafting and Subjective Career Success of Primary and Secondary School Teachers. *Teach. Educ. Res.* 31, 15–21.

Zou, Q., Zuo, B., and Dai, T. (2015). Happiness at work: Definition, measurement levels and causal models. *Adv. Psychol. Sci.* 23, 669–678. doi: 10.3724/SP.J.1042.2015.00669

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 15 November 2022
DOI 10.3389/fpsyg.2022.1009142

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University,
China

REVIEWED BY
Tommy Tanu Wijaya,
Beijing Normal University, China
Zhemin Zhu,
Beihua University, China

*CORRESPONDENCE
Yi Zhang
13609399385@126.com

SPECIALTY SECTION
This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 01 August 2022
ACCEPTED 14 October 2022
PUBLISHED 15 November 2022

CITATION
Wu X, Zhang Y, Wu R, Tang X and Xu T (2022) Cognitive model construction and assessment of data analysis ability based on CDA.
Front. Psychol. 13:1009142.
doi: 10.3389/fpsyg.2022.1009142

Cognitive model construction and assessment of data analysis ability based on CDA

Xiaopeng Wu[1], Yi Zhang[2,3]*, Rongxiu Wu[4], Xiuxiu Tang[5] and Tianshu Xu[3]

[1]Faculty of Education, Northeast Normal University, Changchun, China, [2]School of Mathematics and Statistics, Qiannan Normal University for Nationalities, Duyun, China, [3]School of Mathematical Sciences, East China Normal University, Shanghai, China, [4]Science Education Department, Harvard-Smithsonian Center for Astrophysics, Harvard University, Cambridge, MA, United States, [5]College of Education, Purdue University, West Lafayette, IN, United States

Ability of data analysis, as one of the essential core qualities of modern citizens, has received widespread attention from the international education community. How to evaluate students' data analysis ability and obtain the detailed diagnosis information is one of the key issues for schools to improve education quality. With an employment of cognitive diagnostic assessment (CDA) as the basic theoretical framework, this study constructed the cognitive model of data analysis ability for 503 Grade 9 students in China. The follow-up analyses including the learning path, learning progression and corresponding personalized assessment were also provided. The result indicated that first, almost all the students had the data awareness. Furthermore, the probability of mastering the attribute *Interpretation and inference of data* was relatively low with only 60% or so. Also, the probabilities of mastering the rest of attributes were about 70% on average. It was expected that this study would provide a new cognitive diagnostic perspective on the assessment of students' essential data analysis abilities.

KEYWORDS

data analysis ability, cognitive diagnostic assessment, math education, ability assessment, cognitive model

Introduction

Importance of data analysis ability

With the arrival of big data era, substantial changes have taken place in different fields of society, such as the emergence of artificial intelligence, machine learning, precision medicine, and computational education (François and Monteiro, 2018). Data has become an essential asset in various professions for daily use, and data analysis ability correspondingly becomes a necessity for our work and life (Dani and Joan, 2004). In fields such as business, economics, and others, they increasingly rely on data analytics rather than experience or intuition only to make decisions in management (Bryant et al., 2008). Due to the popularity of data in daily life and workplace, data literacy has gradually become an ability that everyone shall have instead of an ability mastered by only few senior personnel in some specific industries (Borges-Rey, 2017; Sharma, 2017). Prado and Marzal (2013) have provided a list of standards for data literacy, which include (1) Determining the context of data production and reuse, as well as the value, category and format of data; (2) Figuring out when you need data and obtain it appropriately; (3) Properly evaluating data and data sources; (4) Using certain plans, measures, system architecture and appropriate evaluation methods to determine the appropriate methods to operate and analyze the data; (5) Visualization of data analysis results; (6) Using the analysis results to learn, make decisions, or solve certain problems. This list covers almost all required in data literacy and has become one of the basic qualities necessary for the future use. Moreover, data literacy also plays a crucial role in mathematics. Generation and solution of some certain mathematical problems are based on data analysis as well. Good data analysis ability can effectively help students find and solve problems. Therefore, it is of far-reaching significance to focus on cultivating students' data analysis ability.

Definition of data analysis ability

Commonly cultivated by means of statistical disciplines or statistical content, data analysis ability tightly relates with statistics and can be extracted from statistical literacy. However, statistical literacy so far has not been well defined and there is no consensus achieved on its definition among statistics educators, statisticians, and researchers globally (Kaplan and Thorpe, 2010; Schield, 2010; Ridgway et al., 2011; English, 2014). Moore and Cobb (2000) took "change" as the core element of statistical thought and believed that "change" was everywhere in the process. Wallman (1993) stated that statistical literacy was the ability to understand and critically evaluate the statistical results that permeate our daily lives, as well as the ability to understand the contributions of statistical thinking in the public, professional, personal, and private spheres. Through the literature, we can see that there exist common core elements in both the statistical literacy and data analysis ability. Pedagogically, data analysis offers an opportunity for students to explore openly. Conceptually, data analysis examines patterns, hubs, clusters, gaps, dissemination, and variation in data. Philosophically, some advocates of data analysis recommend introducing data in a non-probabilistic environment, while others suggest establishing a link between data analysis and the notion of probability. In the latter view, both data and contingency were considered in the framework of a systematic study of probability (Shaughnessy et al., 1996). In this study, we extracted the data analysis ability from the statistical literacy as the evidence for its definition. Specifically, we defined data analysis ability as the ability to perceive data from daily life, to consciously collect and organize data, to represent data according to different needs, and to rationally analyze and interpret data through operations.

Assessment of data analysis ability

In the assessment of data analysis ability, Graham et al. (2009) built a framework representing children's statistical thinking based on the Cognitive Diagnostic Model (CDM) and other relevant research. This framework provided the theory for the characterization of children's statistical thinking and planning guidance for data processing. Reading (2002) established a four-level statistical thinking analysis framework, which were *Data Collection*, *Data Tabulation and Representation*, *Data Reduction*, and *Interpretation and Inference*. Using the Structure of the Observed Learning Outcome (SOLO, Biggs and Collis, 2014) classification framework, he analyzed the typical responses of students in different levels and drew corresponding conclusions. Mooney (2002) portrayed statistical thinking into four dimensions: *Describing data*, *Organizing and Generalizing data*, *Representing data*, and *Analyzing and Interpreting data*. Still based on the SOLO classification framework, four levels were described for each aspect: *Idiosyncratic*, *Transitional*, *Quantitative*, and *Analytical*. What's more, they also pointed out the necessity of establishing learning trajectories that connect students with different levels of statistical thinking. Due to the high expectation in statistics teaching (Jacobe et al., 2014), effective assessment tools were in a great necessity to evaluate learners' understanding of statistical concepts more precisely.

As we can see, people have sufficiently realized the significance of data analysis and put the cultivation of students' data analysis ability in a more centered position. Although there is no clear definition of data analysis ability yet, a lot of research has been conducted on its attribute division, which lays the foundation for future research. Currently, the assessment of students' data analysis ability has mostly constructed the students' levels of abilities according to the traditional Classical Testing Theory (CTT). It exhibits great advantages in understanding the overall status of students' data analysis ability, however, it cannot provide students with more fine-grained diagnostic information, which is the key to promoting students' development (Huff and Goodman, 2007). The new Cognitive Diagnostic Assessment (CDA), a measurement tool developed in the most recent years combining both cognitive psychology and modern psychometrics, can analyze students' knowledge or skill attributes involved in the process of answering questions from the perspective of cognitive psychology, and integrate attributes into the measurement model. The individual's psychological cognitive process is measured to determine students' mastery of attributes. CDA provides students with more detailed diagnostic information and supports a further in-depth study of students' cognitive process (Leighton and Gierl, 2007). In this study, CDA was applied to construct a cognitive model of students' data analysis ability. Through an examination of 503 Grade 9 students in China, the probability of mastering the attributes of students' data analysis ability can be obtained. On this basis, the students' learning path, learning progression, and personal assessment were then explored, providing a reference for the further study of students' data analysis ability.

Construction of cognitive model of data analysis ability

Sorting out the attributes of data analysis ability

In the attribute classification of data analysis ability, data processing includes organization, description, presentation, and analysis of data, which strongly depends on the representation of various graphics and icons (Shaughnessy et al., 1996). Jones et al. (2000) formed a model composed four attributes of data analysis ability, namely, *Data organization*, *Data representation*, *Data analysis*, and *Data interpretation*. In this model, the continual process of data analysis was systematically presented. What's more, they also made detailed operational definitions for each attribute and coded them accordingly. Mooney (2002) portrayed statistical thinking as four dimensions: *Data description*, *Data organization and generalization*, *Data representation*, *Data analysis*, and *interpretation*. Reading (2002) concluded a few steps in general, which included *Data collection*, *Data tabulation and presentation*, *Data summarization*, *interpretation*, and *inference* in the established statistical thinking framework. Arican and Kuzu (2020) divided data analysis literacy into four aspects: *(a) Data representation and interpretation, (b) Sample interpretation, (c) Statistical methods selection, (d) Understanding and application.* These studies provided the basis for the classification of data analysis ability attributes.

The expert method

The construction of the cognitive model, especially the construction of the relationship between the attributes in the cognitive model, is more complex. Normally, expert method is used to analyze the cognitive process of students in a certain field, to obtain the relationship between different attributes. In this study, five experts were selected, including three middle school math teachers who were awarded the title of Guizhou Provincial or Municipal Famous Teachers, with rich teaching experience and knowledge of students. Another expert was a mathematics education researcher who focuses on the assessment of data analysis ability and can inspect the cognitive model of data analysis ability from the theoretical level. The last expert we invited was a doctoral student majoring in mathematics education, who has been engaging in the teaching of data problem-solving for a long time, and can examine the cognitive model of data analysis ability from the perspective of assessment.

Through open questionnaires, experts were required to enumerate the attributes of data analysis ability and draw the structural relationships between the attributes. Results of the expert method were presented in Table 1.

Based on the common elements in Table 1 and the existing definitions of data analysis ability, the attributes of data analysis ability were extracted, which were: (1) Data awareness; (2) Data collection and sorting; (3) Data representation; (4) Data concentration; (5) Dispersion of Data; (6) Data interpretation and reasoning. According to the relationship between the extracted elements in Table 1, the structure model of data analysis ability was obtained in Figure 1.

Cognitive diagnosis assessment of data analysis ability

Subjects

The subjects of this study were 503 Grade 9 students from 10 classes in two middle schools in Guizhou Province, China. Guizhou is a relatively underdeveloped province in terms of education. Based on the annual proportion of admission to the secondary school entrance examination, the two schools selected belong to the upper middle level in Guizhou Province. The test was 1 h long and all Grade 9 students in both schools took the test. Informed consent was obtained from all the students and teachers before the implementation of test.

Assessment tool

Construction of the Q-matrix

According to the method of constructing the Q-matrix in the Rule Space Model (RSM), the adjacency matrix was obtained according to the cognitive model shown in Figure 1. The preliminary Q-matrix can be further obtained through Boolean algebra (Tatsuoka, 2009), which was presented in Table 2.

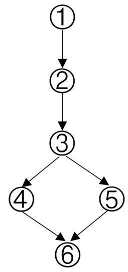
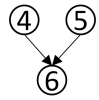
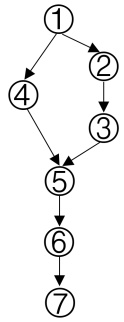
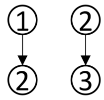
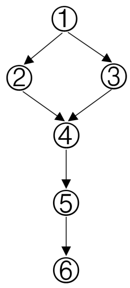
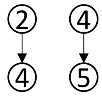
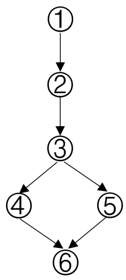
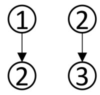
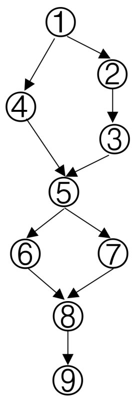
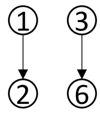
According to Table 2, the test should be composed of 8 items. However, considering the requirements for the number of tasks completed by the students in the test, the preliminary Q-matrix in Table 2 was expanded, and 2 to 3 items were designed for the same examination mode each. The final Q-matrix including 22 items was formed, as shown in Table 3.

Formation of assessment tools

CDA can not only reflect the internal relationship between the test items and cognitive attributes, but also demonstrate the relationship between the subject's knowledge state and attributes (de la Torre and Chiu, 2016). One of the outstanding characteristics of CDA is its assessment structure. Through a more operational and internally consistent Q-matrix, subjects' unobservable cognitive state can be linked to the observable item responses, which goes beyond the simple two-way list format

10.3389/fpsyg.2022.1009142

TABLE 1 Analysis of the expert survey of the cognitive model of data analysis ability.

No	Attribute	Structure	Common elements	Relationship
1	①Ask a question ②Collect data ③Description of data ④Amount of data concentration ⑤Discrete amount of data ⑥Make a decision		②Data collection and sorting ④The degree of data concentration ⑤The degree of dispersion of data ⑥Inference and interpretation of data	
2	①Data awareness ②Data collection ③Data representation ④Understand the randomness of data ⑤Analysis means or strategy choice ⑥Interpretation of results ⑦Result presentation		①Awareness of data ②Data collection and arrangement ③Display of data ⑥Reasoning and interpretation of data	
3	①Collect and organize data ②Descriptive data ③Analyze the data ④Degree of data concentration ⑤Degree of dispersion of data ⑥Experience the randomness of data		②Data collection and arrangement ④The degree of data concentration ⑤The degree of dispersion of data	
4	①Awareness of data ②Data collection ③Display of data ④Data extrapolation ⑤Calculate the relevant amount of data ⑥Using data to solve problems		①Awareness of data ②Data collection and arrangement ③Display of data ⑥Reasoning and analysis of data	
5	①Data awareness ②Collect data ③Collate the data ④Understand the data ⑤Display of data ⑥Data inference ⑦Interpret the data results ⑧Explain the data conclusion ⑨Awareness of randomness		①Awareness of data ②Data collection and arrangement ③Display of data ⑥Reasoning and interpretation of data	

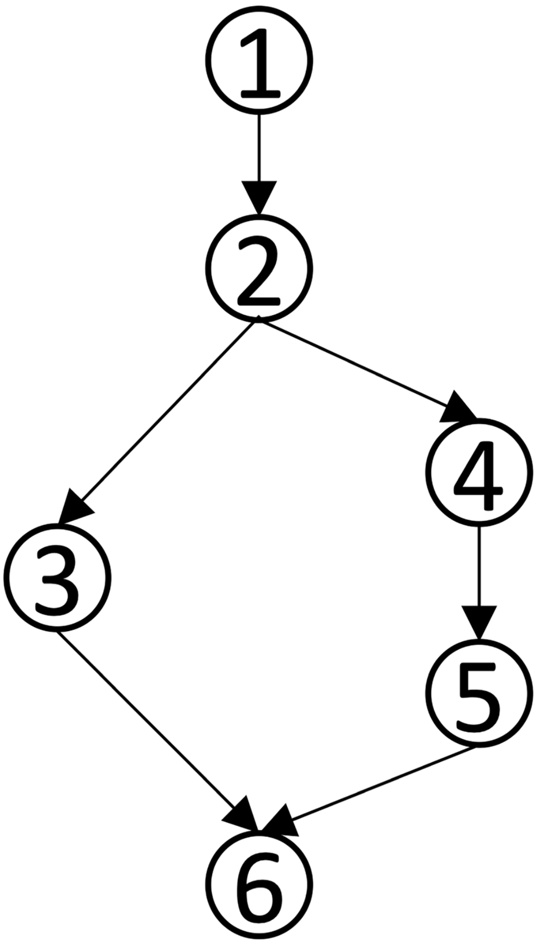


FIGURE 1
Cognitive model of data analysis ability.

TABLE 2 Preliminary Q-matrix of data analysis ability.

	①	②	③	④	⑤	⑥
T1	1	0	0	0	0	0
T2	1	1	0	0	0	0
T3	1	1	1	0	0	0
T4	1	1	0	1	0	0
T5	1	1	0	1	1	0
T6	1	1	1	1	1	1
T7	1	1	1	1	0	0
T8	1	1	1	1	1	0

(Wu et al., 2020). In this study, based on the final Q-matrix in Table 3, we selected 60 items from one of the large-scaled assessment tests *Trends in International Mathematics and Science Study* (TIMSS) as the first round of items selected. Through coding the attributes of these items by the five experts, 22 items with high label consistency were finally selected as the assessment tools.

TABLE 3 Final Q-matrix of data analysis ability.

	A1	A2	A3	A4	A5	A6
Item1	1	0	0	0	0	0
Item2	1	1	0	0	0	0
Item3	1	1	0	1	0	0
Item4	1	1	0	1	0	0
Item5	1	1	0	1	1	0
Item6	1	1	1	1	0	0
Item7	1	1	1	1	0	0
Item8	1	1	1	1	0	0
Item9	1	1	1	1	1	0
Item10	1	1	1	0	0	0
Item11	1	0	0	0	0	0
Item12	1	1	0	0	0	0
Item13	1	1	0	0	0	0
Item14	1	1	0	1	1	0
Item15	1	1	0	1	0	0
Item16	1	1	1	0	0	0
Item17	1	1	1	1	1	0
Item18	1	1	1	0	0	0
Item19	1	1	0	1	1	0
Item20	1	1	1	1	1	0
Item21	1	1	1	1	1	1
Item22	1	1	1	1	1	1

Selection of CDM

Comparison and selection of models play a vital role in CDA process. A large number of cognitive diagnostic practices have shown that choosing an appropriate CDM is an important prerequisite for accurate diagnosis and classification of subjects (Templin and Henson, 2010). Since the theory of CDA was put forward, hundreds of measurement models varying in assumptions, parameters, mathematical principles, and actual conditions have been developed. In order to obtain a more suitable model, using the G-DINA package in the R language, this study evaluated the parameters of common models, such as Deterministic Input; Noisy 'And' Gate (DINA; Haertel, 1989; Junker and Sijtsma, 2001; de La Torre, 2009), Deterministic Input; Noisy 'or' Gate (DINO; Templin and Henson, 2006, 2010), Reduced Reparametrized Unified Model (RRUM; Hartz, 2002), Additive Cognitive Diagnosis Model (ACDM; de La Torre, 2011), Generalized Diagnostic Model (GDM; von Davier, 2014), Log-linear Cognitive Diagnosis Model (LCDM; Henson et al., 2009), Linear Logistic Model (LLM; Maris, 1999), Generalized DINA (G-DINA; de La Torre, 2011), and Mixture Model (von Davier, 2010). Based on the Q-matrix in Table 3 and test data, the relevant parameters of the CDMs were estimated, as shown in Table 4.

Two criteria are generally considered in the selection of CDMs, which are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The number of parameters represents the load in the model evaluation process. The smaller the value of AIC and BIC, the smaller the load and the better the model fitting (Vrieze, 2012). Through the comparisons of AIC and

TABLE 4 Statistical comparison of parameters in different models.

Model	Number of parameters	Deviation	AIC	BIC
DINA	107	10735.15	10949.15	11400.75
DINO	107	10793.88	11007.88	11459.48
RRUM	162	10507.75	10831.75	11515.49
LLM	162	10503.48	10827.48	11511.22
ACDM	162	10523.89	10847.89	11531.63
GDM	121	10584.73	10826.73	11337.43
LCDM	406	10521.61	11333.61	13047.17
GDINA	447	10183.07	11077.07	12963.68
Mixed	125	10668.92	10918.92	11446.5

BIC in Table 3, the three indicators of the GDM were the smallest, which represented the most appropriate model fit to the data.

GDM model

The GDM is a model that adapts to multi-level response variables and has two or more skill levels. It extends the commonly used IRT model to a multivariate, multiskilled classification model (von Davier and Yamamoto, 2004). Like other CDMs, Q-matrix is an effective part of the model. Its general form can be applied to non-integer, multi-dimensional and multi-attribute skills. It provides a more general way to specify the skill mode and the Q-matrix interact. The model (von Davier, 2008) is presented as:

$$P_{ig}(x|a) = P\left(x|\beta_{ig}, a, q_i, \gamma_{ig}\right) = \frac{\exp\left[\beta_{xig} + \sum_{k=1}^{K} x\gamma_{ikg} q_{ik} a_k\right]}{1 + \sum_{y=1}^{m_i} \exp\left[\beta_{yig} + \sum_{k=1}^{K} y\gamma_{ikg} q_{ik} a_k\right]}$$

where $P\left(x|\beta_{ij}, a, q_i, \gamma_{1g}\right)$ indicates that the distribution of observation variable x is in a given condition $(a_1, \cdots a_K)$, difficulty parameter β_{ig}, guessing parameter γ_{ig} and Q-matrix q_i. $q_{ik} a_k = b_k(q_i, a)$, however, $b(q_i, a) = (q_{i1}a_1, \cdots, q_{iK}a_K)$ in IRT, therefore, the k_{th} element of b $b_k(q_i, a) = q_{ik} a_k$

$$b_k(q_i, a) = \begin{cases} a_k, & \text{if } q_{ik} = 1, \\ 0, & \text{if } q_{ik} = 0. \end{cases}$$

where $q_{ik, i=1,\cdots,I,\ k=1,\cdots,K}$ is a $I \times K$ matrix with true value q_i. This matrix associates I observed variables with K unobserved (skills) variables to determine these variables in a specific model in cognitive diagnosis. GDM is a general diagnostic model suitable for both dichotomous and multichoice data, which can model multi-dimensional mixed binary and sequential skill variables.

Quality of assessment tool

Reliability of the test

Reliability represents the stability and reliability of measurement, which is one of the most important indicators of tool evaluation in the test. As a new generation of assessment, CDA has its own uniqueness that its reliability mainly focuses on the consistency index of attribute retest (Templin and Bradshaw, 2013). Similar to the Subkoviak method of decision consistency in the CTT standard reference test, this indicator is calculated by correlating the probabilities of the attribute's mastery of the same subjects in two successive measurements, with the assumption that the probabilities of the attributes mastered by the subjects remain the same (Templin and Bradshaw, 2013). The reliabilities for different models were shown in Table 5.

According to the statistics in Table 5, the test–retest reliability was estimated by repeatedly extracting from examinee's posterior distribution to simulate repeated testing (Templin and Bradshaw, 2013). The test–retest consistency index was acceptable, with the average value reaching 0.8938. The reliability of each attribute was above 0.78, and most of them were above 0.85, indicating that the GDM model was reliable to use for the current dataset.

Item fit

Item fit is also the focal point in the CDM analysis. Studies have shown that whether the test data of a CDM fits the items or not directly determines the accuracy of the model's diagnostic effect (Song et al., 2016). The conventional method to examine the fitting effect of items is the chi-square test, however, since the characteristics of the CDA do not conform to the hypothesis of the chi-square test, and the preconditions of the chi-square test do not meet, the traditional chi-square test are not appropriate to evaluate the fitting effect of items in CDM (Batanero and Díaz, 2010). In CDA, RMSEA is used to measure the fitting effect of test items by mainly comparing the square root error of observed response and predicted response under the different potential classifications. The calculation formula of RMSEA for item *j* is:

$$RMSEA_j = \sqrt{\sum_k \sum_c \pi(\theta_c) \left(P_j(\theta_c) - \frac{n_{jkc}}{N_{jc}} \right)^2}$$

where $\pi(\theta_c)$ represents the classification probability of the potential trait level of type c, and P_j represents the probability estimated by the item response function. n_{jkc} represents the expected number of people at the k_{th} dimension of the c_{th} potential trait level in the j_{th} item, and N_{jc} represents the expected number of people at the c_{th} potential trait level. Through the calculation of residual information, the residual information of the test items was shown in Table 6.

The closer the RMSEA value is to 0, the smaller the deviation of the fit and the better the fitting effect. The critical value of

TABLE 5 Reliability of data analysis ability.

Templin reliability index						
A1	A2	A3	A4	A5	A6	Mean
1	0.9275	0.7869	0.9285	0.8542	0.8658	0.8938

RMSEA is normally set to 0.1. Values greater than 0.1 for RMSEA is an indication of poor item fit (Oliveri and von Davier, 2011). According to this standard, the GDM model was still the one with the best fit in the test items since the RMSEA values for all the test items were less than 0.1. Only items 3, 7, 9, 13, 17, 18, 19, 20, 21, 22 had the values slightly greater than 0.1. Therefore, the GDM model still had the most appropriate item level fit in this study.

Assessment result

Probability of mastery of attributes of data analysis ability

The probability of each student's attribute mastery can be obtained through the assessment of the model in CDA. To further analyze the gender differences in students' mastery of attributes, comparison analyses of the attribute mastery probability were conducted on genders and the mastery probabilities of the six attributes were obtained as shown in Figure 2.

Figure 2 showed that the mastery probability of the attribute *Know data* was the highest, reaching 100%, which indicated that almost all students had the basic data awareness. The probability of mastering these four attributes, *Data collection and collation, Data presentation, Concentration of data,* and *Dispersion of the data*, were quite similar at about 70%. Among these four, the probabilities of mastering attributes of *Data collection and collation* and *Dispersion of the data* were slightly higher. Last, the mastery probability of *Interpretation and inference of data* was the lowest, about 60%.

Overall, in Figure 2, what we can observe was that there was no obvious gender difference in attribute mastery probability, especially the probability of mastering *Know data*, *Data presentation* and *Dispersion of data* were almost the same for both male and female students. In terms of *Data collection and collation*, *Concentration of data* and *Interpretation and inference of data*, the probabilities of male students mastering attributes were slightly higher than these for female ones, showing certain advantages for male students.

Construction of learning progression for data analysis ability

With a combined method of learning progression construction with Item Response Theory (IRT), students' abilities were calculated in each type of knowledge state, and students in different knowledge state categories were divided (Wu et al., 2021). The ability value 1.37 was the highest for the knowledge state (111111),

TABLE 6 RMSEA information of the test items.

Item	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
RMSEA	0.0176	0.0406	0.1041	0.0839	0.0989	0.0807	0.1151	0.0812
Item	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16
RMSEA	0.1028	0.0673	0.0684	0.0486	0.1107	0.0942	0.0799	0.0692
Item	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	\	\
RMSEA	0.1857	0.1103	0.4106	0.2119	0.1923	0.2511	\	\

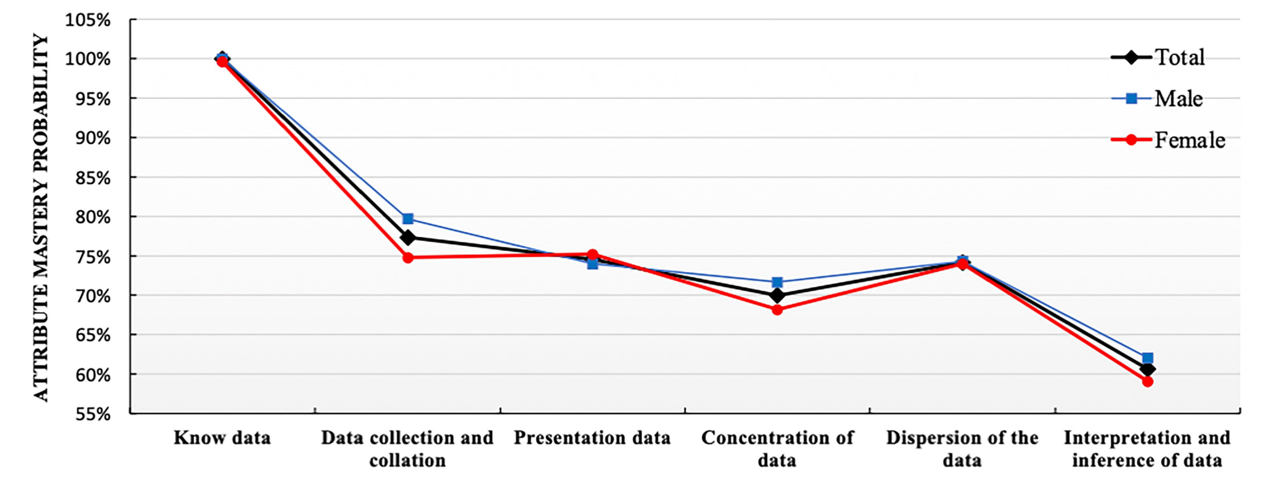


FIGURE 2
Distribution diagram for attribute mastery probability of data analysis ability.

and the ability value −2.41 was the lowest for the knowledge state (010000). We have divided the interval (−2.5, 1.5) into 5 levels and the diagram for learning progression was shown in Figure 3.

According to the attributes of the knowledge states in the different stages in Figure 3, the learning progression of the ladder were defined, and the divisions of learning progression level were organized in Table 7. It provided a more reliable theoretical basis for student learning, teacher teaching, textbook compilation, and test compilation.

Personalized analysis of data analysis ability

Accurately depicting the knowledge state of each student is the greatest advantage of CDA. To fully illustrate the fine-grained information that CDA can provide for each student, three students numbered 337, 476 and 424 were selected for the analysis. The radar chart of their attribute mastery was shown in Figure 4.

The characteristic of the three students shown in Figure 4 was that they correctly answered the same number of questions. If through the traditional evaluation methods, these students were considered to have the same total score, showing that there was no difference among them. However, in CDA, differences were still obvious in Figure 3. Student No. 337 had mastered attributes ①, ②, ④, ⑥; student No. 476 had mastered attributes ①, ②, ③, ⑤, ⑥, but not fully mastered attributes ② or ⑥, whose probability of attribute mastery was approximately 80%; and student No. 424 had mastered the attributes ①, ②, ④, ⑤, ⑥. Therefore, what we can conclude was that these three students not only differed in the type of attribute mastery, but also in the number and degree of attribute mastery. These results offered more fine-grained information for students' personalized learning.

Discussion

This study constructed a reasonable cognitive model for students' data analysis ability based on the expert method. Results showed that students had a good attribute mastery of *Know data*, but students' mastery of *Interpretation and inference of data* was relatively poor. Data analysis ability can be clearly divided into five levels of learning progression and students with the same scores differed obviously in the knowledge state.

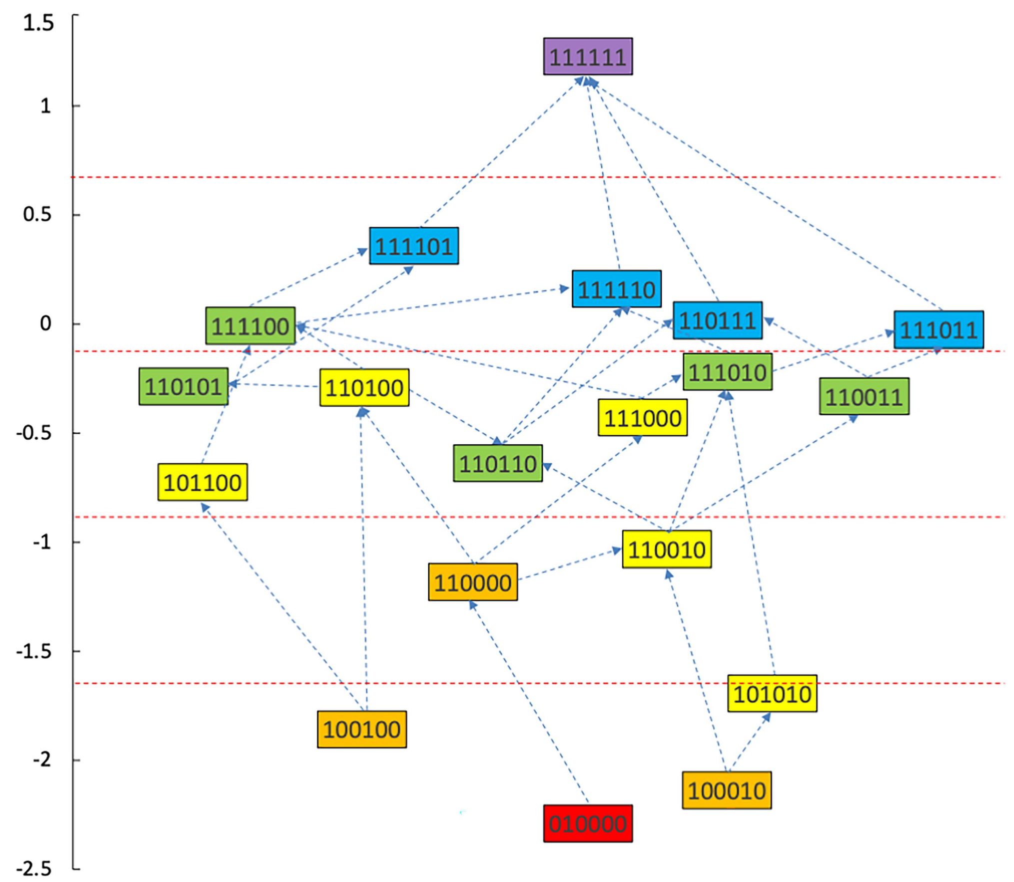


FIGURE 3
Advanced diagram of data analysis ability progression.

TABLE 7 Divisions of learning progression levels of data analysis ability.

Level	Content definition	Attribute mastery
1	Students can pay attention to the core data in problem solving, and be able to distinguish different data, recognize the role of data in problem solving, and have a certain sense of data	Preliminary mastery of attributes ①
2	Students are able to collect data according to reasonable methods, and carry out preliminary sorting and statistics on the collected data, and carry out preliminary data management.	On the basis of mastering ①, further master the attributes ②
3	Students can properly represent the data, such as using histograms, line graphs, etc., and can freely convert between different representations; can calculate the average, mode, and median of the data and experience different quantities specialty	On the basis of mastering ① and ②, further master ③ and ④ attributes
4	Students can calculate the degree of dispersion of data, such as variance, standard deviation, range, etc., based on the amount of concentration, and appreciate the meaning of the degree of dispersion of data. Infer the meaning expressed by the data and draw useful conclusions.	On the basis of mastering the attributes ① to ④, grasp the attributes ⑤ and ⑥ initially
5	On the basis of mastering the concepts and operations of integers, measurement, plane geometry, data, probability, and preliminary statistics, students have further mastered the basic concepts and operations of elementary algebra, and can solve practical problems related to equations, inequalities, and functions	Master all attributes

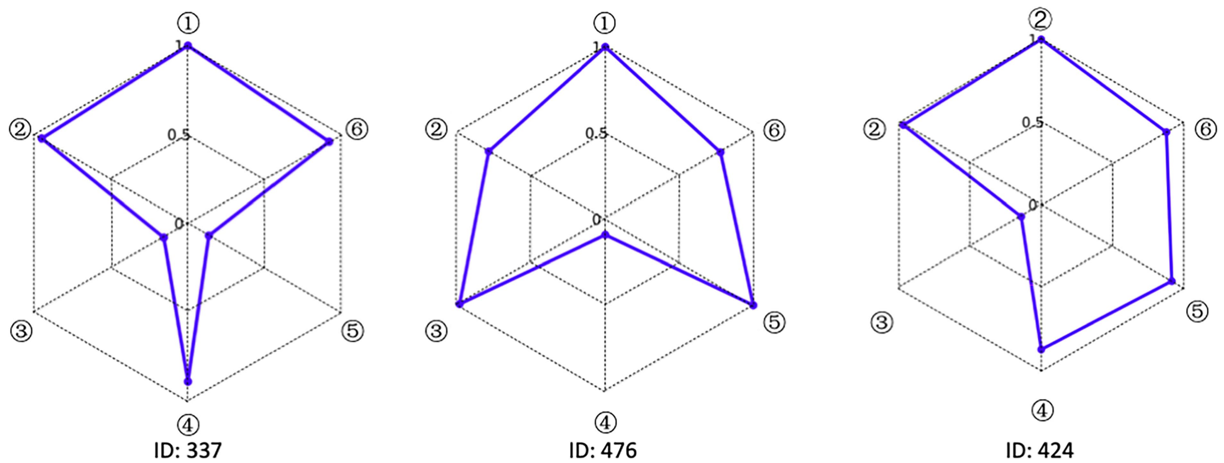


FIGURE 4
Radar chart of the mastery results of students' data analysis ability.

As one of the essential core qualities of modern citizens, the importance of data analysis ability has been recognized by all walks of life. The accurate assessment of data analysis ability is a topic that is worthy further exploration. As a new generation of assessment theory, CDA is designed to detect students' specific knowledge structure or operational skills in a certain field, so as to provide students with more fine-grained diagnostic information on cognitive strengths and weaknesses (Leighton and Gierl, 2007). It is essentially a diagnosis of cognitive attributes, and the construction of attributes plays a vital role in the assessment process (Wu et al., 2020). However, most of the extant research have only considered the division of cognitive attributes without taking their prerequisite relationship into account. Starting from expert method, this study constructed a cognitive model while having the prerequisite relationship between attributes into account. The model that was formed has provided a theoretical basis for clarifying the relationship between the attributes of data analysis ability and further guiding the teaching and students' personal assessment. It provided a more standardized research method for the assessment of students' data analysis abilities, which was more in line with the research paradigm of CDA.

In this study, we also explored the learning path and learning progression of students' data analysis ability. The learning path depicted student's thinking and learning in a specific field of mathematics, reflecting student's learning process through a series of instructive tasks. These tasks were designed to promote the development of students' mental process and thinking level (Clements and Sarama, 2012). The learning path reflects the trend of students' actual learning progress, rather than focusing on subject knowledge, which distinguishes the learner's logic from the subject logic and plays an intermediary role in the selection of learning goals and methods (Corcoran et al., 2009). The realization of different learning objectives needs to be supported by the corresponding learning paths, and different learning paths will determine the choice of learning methods. In the selection of learning methods, learning path distinguishes the student's "voice" from the subject's perspective, emphasizes the development of students' cognitive order, and further clarifies the importance of learners in guiding future teaching, curriculum and assessment (Confrey, 2006). Students are all in different levels of learning paths, with various learning resources and learning environment around, therefore, a selection of a suitable learning path and learning method according to their individual conditions and backgrounds will be more appropriate. Learning path helps learners choose appropriate teaching activities, tasks, tools, interaction and evaluation methods, and promotes students to gradually master increasingly complex concepts (Confrey et al., 2009). It combines the evaluation results of each student with the corresponding learning mode, extracts the formative evaluation results from the summative evaluation data, and provides a basis for students to formulate their personalized learning plans.

This study provided a more complete and standardized research method, constructed a cognitive model of data analysis ability, and made contributions to the theory as well as methods to a certain degree. However, due to the limited material and financial resources, this study also had some shortcomings inevitably. First, the sample size available was limited. With only approximately 500 students in China, the result of the study lacked generalizability to some extent. In addition, any assessment cannot make without the discipline itself. The evaluation content should be analyzed based on the characteristics of the discipline (Zhang et al., 2019). Last, a longitudinal study with similar method and approach is recommended in the future to verify the result reliability and validity (Zhan et al., 2019; Wu et al., 2020).

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to XW, 18198689070@126.com.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XW designed the study and wrote this manuscript. YZ contributed to the manuscript writing and the continued revision provided by the reviewers. XT collected the data and wrote this manuscript. RW and TX have revised the language of the article and provided comments. All authors contributed to the article and approved the submitted version.

Funding

This work was support by 2021 Humanities and Social Science Research project of Ministry of Education of China: Cognitive Diagnosis and Evaluation of Mathematics Core Literacy (21YJC880102). Teacher Education "JIEBANGLINGTI" Project of Northeast Normal University: Learning Progression Construction and Learning Path Analysis Based on Cognitive Diagnosis (JSJY20220305).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arican, M., and Kuzu, O. (2020). Diagnosing preservice teachers' understanding of statistics and probability: developing a test for cognitive assessment. *Int. J. Sci. Math. Educ.* 18, 771–790. doi: 10.1007/s10763-019-09985-0

Batanero, C., and Díaz, C. (2010). Training teachers to teach statistics: what can we learn from research? *Statist. et enseignement* 11, 83–20. doi: 10.1111/j.1467-9639.1989.tb00069.x

Biggs, J. B., and Collis, K. F. (2014). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)* Hobart, Tasmania, Austraila: Academic Press.

Borges-Rey, E. L. (2017). "Data literacy and citizenship: understanding 'big Data'to boost teaching and learning in science and mathematics," in *Handbook of research on driving STEM learning with educational technologies*. eds. M. S. Ramírez-Montoya (IGI Global), 65–79. doi: 10.4018/978-1-5225-2026-9.ch004

Bryant, R., Katz, R. H., and Lazowska, E. D. (2008). Big-data computing: creating revolutionary breakthroughs in commerce, science and society. Available at: https://cra.org/ccc/wp-content/uploads/sites/2/2015/05/Big_Data.pdf (Accessed October 20, 2022).

Clements, D. H., and Sarama, J. (2012). "Learning trajectories in mathematics education," in *Julie Sarama hypothetical learning trajectories* eds. H. Douglas, D. H. Clements and H. Sarama (NewYork: Routledge), 81–90.

Confrey, J. (2006). "The evolution of design studies as methodology. A aparecer en RK Sawyer," in *The Cambridge handbook of the learning sciences*. ed. R. K. Sawyer (Thessaloniki, Greece: Cambridge, UK: Cambridge University Press), 135–151.

Confrey, J., Maloney, A., Nguyen, K., Mojica, G., and Myers, M. (2009). "Equipartitioning/splitting as a foundation of rational number reasoning using learning trajectories," in *Paper presented at the 33rd conference of the International Group for the Psychology of mathematics education* eds. M. Tzekaki, M. Kaldrimidou, and H. Sakonidis, (Thessaloniki, Greece: PME).

Corcoran, T. B., Mosher, F. A., and Rogat, A. (2009). Learning progressions in science: an evidence-based approach to reform. Available at: http://repository.Upenn.Edu/cpre_research reports/53 (Accessed October 20, 2022).

Dani, B. Z., and Joan, G. (2004). "Statistical literacy, reasoning, and thinking: goals, definitions, and challenges," in *The challenge of developing statistical literacy, reasoning and thinking*. eds. D. Ben-Zvi and J. Garfield (Dordrecht: Springer)

de La Torre, J. (2009). DINA model and parameter estimation: a didactic. *J. Educ. Behav. Stat.* 34, 115–130. doi: 10.3102/1076998607309474

de La Torre, J. (2011). The generalized DINA model framework. *Psychometrika* 76, 179–199. doi: 10.1007/s11336-011-9207-7

de la Torre, J., and Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika* 81, 253–273. doi: 10.1007/s11336-015-9467-8

English, L. D. (2014). "Promoting statistical literacy through data modelling in the early school years," in *Probabilistic thinking. Advances in mathematics education*. eds. E. Chernoff and B. Sriraman (Dordrecht: Springer)

François, K., and Monteiro, C. (2018). Big data literacy. Paper presented at the looking back, looking forward. Proceedings of the 10th international conference on the teaching of statistics, Kyoto, Japan, July.

Graham, A. T., Pfannkuch, M., and Thomas, M. O. J. (2009). Versatile thinking and the learning of statistical concepts. *ZDM* 41, 681–695. doi: 10.1007/s11858-009-0210-8

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *J. Educ. Meas.* 26:301. doi: 10.1111/j.1745-3984.1989.tb00336.x

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*, Illinois, America: University of Illinois at Urbana-Champaign.

Henson, R. A., Templin, J. L., and Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika* 74, 191–210. doi: 10.1007/s11336-008-9089-5

Huff, K., and Goodman, D. P. (2007). "The demand for cognitive diagnostic assessment," in *Cognitive diagnostic assessment for education: Theory and applications*. eds. J. P. Leighton and M. J. Gierl (Cambridge, UK: Cambridge University Press), 19–60.

Jacobe, T., Foti, S., and Whitaker, D. (2014). Middle school (ages 10–13) students. Understanding of stat- istics. In K. Makar, SousaB. de and R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the ninth international conference on teaching statistics (ICOTS9, July, 2014)*, Flagstaff, Arizona, USA. Voorburg: International Statistical Institute.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., and Putt, I. J. (2000). A framework for characterizing children's statistical thinking. *Math. Think. Learn.* 2, 269–307. doi: 10.1207/S15327833MTL0204_3

Junker, B. W., and Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Appl. Psychol. Meas.* 25, 258–272. doi: 10.1177/01466210122032064

Kaplan, J. J., and Thorpe, J. (2010). "Post-secondary and adult statistical literacy: assessing beyond the classroom" *in Data and context in statistics education: towards an evidence-based society. Proceedings of the eighth international conference on teaching statistics*. Voorburg, The Netherlands: International Statistical Institute.

Leighton, J. P., and Gierl, M. J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educ. Meas. Issues Pract.* 26, 3–16. doi: 10.1111/j.1745-3992.2007.00090.x

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika* 64, 187–212. doi: 0033-3123/1999-2/1995-0422-A, 10.1007/BF02294535

Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Math. Think. Learn.* 4, 23–63. doi: 10.1207/S15327833MTL0401_2

Moore, D. S., and Cobb, G. W. (2000). Statistics and mathematics: tension and cooperation. *Am. Math. Mon.* 107, 615–630. doi: 10.1080/00029890.2000.12005247

Oliveri, M. E., and von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychol. Test Assess. Model.* 53, 315–333.

Prado, J. C., and Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri* 63, 123–134. doi: 10.1515/libri-2013-0010

Reading, C. (2002). *Profile for statistical understanding. Paper presented at the proceedings of the sixth international conference on teaching statistics*, Cape Town, South Africa.

Ridgway, J., Nicholson, J., and McCusker, S. (2011). "Developing statistical literacy in students and teachers," in *Teaching statistics in school mathematics-challenges for teaching and teacher education: A joint ICMI/IASE study*. eds. C. Batanero, G. Burrill and C. Reading (Dordrecht: Springer), 311–322.

Schield, M. (2010). "Assessing statistical literacy: take care," in *Assessment methods in statistical education: An international perspective*. eds. P. Bidgood, N. Hunt and F. Jolliffe (Milton: John Wiley & Sons), 133–152.

Sharma, S. (2017). Definitions and models of statistical literacy: a literature review. *Open Rev. Edu. Res.* 4, 118–133. doi: 10.1080/23265507.2017.1354313

Shaughnessy, J. M., Garfield, J., and Greer, B. (1996). "Data handling," in *International handbook of mathematics education*. eds. A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick and C. Laborde, vol. *4* (Dordrecht, The Netherlands: Kluwer), 205–238.

Song, L., Wang, W., and Dai, H. (2016). Overall and item fitting indicators under the cognitive diagnostic model. *Psychol. Explor.* 36, 79–83.

Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*, NewYork: Routledge.

Templin, J., and Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *J. Classif.* 30, 251–275. doi: 10.1007/s00357-013-9129-4

Templin, J. L., and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychol. Methods* 11, 287–305. doi: 10.1037/1082-989X.11.3.287

Templin, J., and Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*, New York: Guilford Press.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *Br. J. Math. Stat. Psychol.* 61, 287–307. doi: 10.1348/000711007X193957

von Davier, M. (2010). Hierarchical mixtures of diagnostic models. *Psychol. Test Assess. Model.* 52, 8–28.

von Davier, M. (2014). The log-linear cognitive diagnostic model (LCDM) as a special case of the general diagnostic model (GDM). *ETS Res. Rep. Series* 2014, 1–13. doi: 10.1002/ets2.12043

von Davier, M., and Yamamoto, K. (2004). A class of models for cognitive diagnosis. Paper presented at the 4th spearman conference, Philadelphia, PA.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol. Methods* 17, 228–243. doi: 10.1037/a0027127

Wallman, K. K. (1993). Enhancing statistical literacy: enriching our society. *J. Am. Stat. Assoc.* 88, 1–8. doi: 10.1080/01621459.1993.10594283

Wu, X., Wu, R., Chang, H.-H., Kong, Q., and Zhang, Y. (2020). International comparative study on PISA mathematics achievement test based on cognitive diagnostic models. *Front. Psychol.* 11:2230. doi: 10.3389/fpsyg.2020.02230

Wu, X., Wu, R., Zhang, Y., Arthur, D., and Chang, H. H. (2021). Research on construction method of learning paths and learning progressions based on cognitive diagnosis assessment. *Assess. Edu.: Princip. Policy Pract.* 28, 657–675. doi: 10.1080/0969594X.2021.1978387

Zhan, P., Jiao, H., Liao, D., and Li, F. (2019). A longitudinal higher-order diagnostic classification model. *J. Educ. Behav. Stat.* 44, 251–281. doi: 10.3102/1076998619827593

Zhang, Q., Bian, Y., Chen, P., and Zhang, J. (2019). The cognitive diagnostic assessment of lower primary school Students' Chinese character learning. *Educ. Res.* 44, 76–85.

frontiers | Frontiers in Psychology

TYPE Review
PUBLISHED 29 November 2022
DOI 10.3389/fpsyg.2022.1039761

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Trung Tran,
Vietnam National University, Hanoi, Vietnam
Qi Ye,
South China Normal University, China

*CORRESPONDENCE
Qiaoping Zhang
zqiaoping@eduhk.hk

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 08 September 2022
ACCEPTED 07 November 2022
PUBLISHED 29 November 2022

CITATION
Xu X, Zhang Q, Sun J and Wei Y (2022) A bibliometric review on latent topics and research trends in the growth mindset literature for mathematics education. *Front. Psychol.* 13:1039761. doi: 10.3389/fpsyg.2022.1039761

A bibliometric review on latent topics and research trends in the growth mindset literature for mathematics education

Xiaoyu Xu[1], Qiaoping Zhang[1]*, Jin Sun[2] and Yicheng Wei[1]

[1]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong, Hong Kong SAR, China, [2]Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong, Hong Kong SAR, China

Embracing a growth mindset is essential to students' academic improvement. This manuscript aims to better understand the existing literature on the role and effects of the growth mindset in mathematics teaching and learning. It provides an updated perspective on the research regarding the growth mindset in mathematics education. The dataset comprises 85 journal articles published from 2012 to 2022 retrieved from the Web of Science (WOS) and Scopus databases. The current study applies a methodology based on bibliometric analysis techniques. The analysis reveals and corroborates several patterns from the research trends, journals, countries, and authors that have significant impacts on the research field. The findings show that USA, UK, and Norway are the most productive countries in publishing research on the topic. Moreover, the results of the thematic analysis indicate that the topics discussed among most of the articles in the dataset include engagement, implementation, persistence, children, fluid intelligence, and skills. The longitudinal trends in research themes based on study keywords illustrate an evolution in the research from the concept of mindsets to implicit theories on the growth mindset alongside academic achievement. Lastly, this study also provides an overview of the conceptual structure underlying studies on the growth mindset, which offers valuable insights into potential research topics for academics and practitioners seeking to explore the growth mindset in the future.

KEYWORDS
growth mindset, mathematics education, bibliometric analysis, student learning, mathematics learning, bibliometric review

10.3389/fpsyg.2022.1039761

Introduction

Entering the early 21st century, we have witnessed lots of changes and evolvements. Amid globalization, the growth of information and communication technology (ICT), and knowledge sharing, the content of education is changing (Trilling and Fadel, 2009; Voogt and Roblin, 2012). To cope with our ever-changing society, education should equip students with essential skills to enable them to thrive and succeed in their future. In 2018, the Organisation for Economic Co-operation and Development [OECD] (2018) proposed the Learning Compass 2030 framework, emphasizing the importance of critical thinking, meta-cognitive skills, learning-to-learn skills, and the ability to learn attitudes and values.

In education, students' mindset strongly influences their learning performance (Dweck, 2017a). A student's mindset refers to the attitudes, beliefs, and expectations they have about a course or subject (Chew, 2014; Dweck, 2017b). These types of mindsets can be a boon or a barrier to learning. Students with a fixed mindset tend to believe intelligence is fixed, and that they are born with a particular set of skills and cannot change them (Dweck, 2006). Children with a fixed mindset are concerned with how they will be judged, and they want to make sure they succeed (Dweck, 2017b). In contrast, students with a growth mindset see intellectual ability as a malleable trait that could be cultivated and enhanced through personal effort and guidance (Dweck, 2015). They are concerned with improving. For these children, success is about stretching themselves (Dweck, 2017b). These students are more likely to embrace intellectual challenges as opportunities to learn and grow, and to be more resilient in the face of setbacks (Yeager and Dweck, 2020). The National Council of Teachers of Mathematics (2014) reported that "believing in, and acting on, growth mindsets versus fixed mindsets can make an enormous difference in what students accomplish" (p. 64). When it comes to mathematics learning, mindset is of particular importance. Students with mathematical problem-solving and critical thinking skills are among the strongest performers with a growth mindset in overall mathematical achievement (Organisation for Economic Co-operation and Development [OECD], 2013). To nurture a growth mindset for students, we should also examine teachers. Studies have revealed that teacher mindsets can influence students' mindset and directly affect their achievement (Ostroff, 2016; Ronkainen et al., 2019). Kamins and Dweck (1999) observed that when teachers believed in their students' ability in achieving success, the students could stretch their limits and exceed expectations. However, in this research area, many empirical studies have focused on students, and more research needs to be conducted on how to develop a growth mindset in teachers (Guidera, 2014).

Although positive correlations or influences have been found between students' growth mindset and their performance (Dweck, 2017a), to what extent and in what kinds of learning areas these effects hold are unclear. A more holistic analysis of specific empirical studies is required. Notably, most studies lack a comprehensive understanding of the entire growth mindset in mathematics education. For example, whether a growth mindset approach exists for other learning and teaching topics in mathematics education remains unanswered. Furthermore, to the best of our knowledge, no previous efforts have been made to conduct a bibliometric review of the literature in this field, a widely used mathematical and statistical tool for quantitative research (Pritchard, 1969; Chen et al., 2018). Thus, this article aims to fill this gap by systematically reviewing the literature with the bibliometric method, and to summarize current literature findings. We examine empirical studies on the growth mindset in mathematics education over the last 10 years and provide a more detailed picture of latent topics, development trends, collaborative organizations, and annual topic distributions. This study also further discusses the representative research work and also suggests a possible pathway for future research.

The current review investigates the following research questions:

(1) Which countries/regions were major contributors to growth mindset research in the last decade?
(2) What were the primary research topics for the growth mindset in mathematics education and their significance to our society (i.e., students, teachers, school, broader society)?
(3) How did research topics evolve through the years?
(4) What could be the possible research directions in the future?

Literature review

To better understand the development of the growth mindset in mathematics education, we first briefly introduce some key concepts in growth mindset research. This section analyzes three aspects, including the definition of the growth mindset, interventions for the growth mindset, and the growth mindset in mathematics education, to review and describe state-of-the-art research in mathematics education.

Definition of the growth mindset

Mindset can be understood as the influence of past thinking on current thinking. It is a collection of beliefs related to continual learning and the malleability of intelligence (Dweck, 2006). Mindset could be classified into two types: the fixed mindset and the growth mindset (Dweck et al., 1995;

10.3389/fpsyg.2022.1039761

Dweck, 2006). A person with a fixed mindset believes that intelligence is a stable, unchangeable trait. Conversely, a person with a growth mindset believes that intelligent skills could be cultivated and developed through effort.

Previous studies have interpreted the effect of different mindsets on student learning and teachers' teaching, especially when they struggled with problems or failure. Students with a fixed mindset tend to avoid challenges, quit when they encounter challenges, and ultimately achieve less academic success (Dweck, 2006; Smiley et al., 2016). Conversely, when students learn with a growth mindset, they can improve with effort and guidance. They are more willing to accept challenging work and persevere through obstacles by exploring new tactics or increasing their efforts. Those students realize and appreciate the importance of trial and error, where they can learn from mistakes and alter their tactics (Dweck, 2006; Boaler, 2015). A mindset might change with different contexts and over time. Teachers' understanding and explanation of mindset theory could help students change their mindset toward learning mathematics and promote their positive beliefs and attitudes toward the subject (Boaler, 2015).

Interventions for promoting the growth mindset

Replicated studies (Sisk et al., 2018; Yeager et al., 2019) show that mindset treatments have a positive impact on student learning achievements. Hence, changing students' mindset from a fixed type to a growth type becomes crucial. Yeager and Dweck (2020) identified instructional interventions that assist struggling students in tracking their progress. By adopting a specific program or a series of steps to target an academic need, these interventions were expected to help kids with learning troubles in subjects such as mathematics. Moreover, Yeager and Dweck (2020) proposed that any intervention should describe actionable steps for developing a growth mindset. For instance, individuals can train their brains by attempting challenging schoolwork. They may also benefit from hearing about notable people or colleagues with a development mindset. Nevertheless, interventions should not be passive actions; they should require individual reflection. For instance, as part of an intervention program, students may compose a brief essay about how they have developed their abilities through challenges and how they want to adopt a growth mindset in their future endeavors. Students may also compose a letter or write what they would communicate to their peers; this exercise can determine which students have a fixed mindset. Ultimately, interventions should not merely highlight the effort but also show that learning abilities have the potential for improvement. This does not mean that learning abilities can be readily altered or considerably modified, but that the potential for change exists (Yeager and Dweck, 2020). Vongkulluksn et al. (2021) mentioned that the learning process should be highlighted rather than the results of learning. Students should learn to acquire and generalize strategies and resources that they could apply in future work. Teachers could play their part in helping students to go through failure or setbacks and appreciate them as part of the learning process. Failure would offer crucial feedback on improvements and help build knowledge (Dweck, 2017a). Feedback is vital and should be matched with the learning objectives that students are aiming to achieve.

The growth mindset in mathematics education

Having a growth mindset would help students understand that they could improve their mathematical abilities with effort. Holding a growth mindset in mathematics learning meant that students could leverage a particular thinking procedure to solve mathematical problems and were willing to attempt the task various times despite setbacks. This kind of mindset would gradually transform into a habitual response (Dweck, 2008). Solving problems multiple times through trial and error cultivated a growth mindset, which helped students learn mathematics and strengthened their belief in the possibility of growth in their intelligence (Dweck, 2015). Students with a growth mindset believed that the more they learnt, ranging from mathematical principles to calculation methods, the better their mathematical thinking skills, driving a virtuous cycle in their continuous learning (Ng, 2018). They understood that their objectives in learning mathematics were to think, understand, and grow. However, when students treated mathematical problems as just a series of short questions, they could not appreciate their own cognitive development in small steps and the wider applications of learning mathematics (Boaler, 2015). They perceived that there were only fixed methods for solving particular mathematical problems.

Several studies on students' mathematics learning attitudes considered the students' mindset to be an important factor in developing their problem-solving skills. The present study addresses the growth mindset in mathematics learning. The growth mindset in mathematics learning refers to how an individual thinks while learning mathematics, reflecting their number sense, logical thinking ability, judgment ability, and speculative ability (Hakim and Nurlaelah, 2018). A growth mindset highlighting the learning process was significant for developing students' problem-solving skill, and beneficial to their continuous pursuit of learning (Dweck, 2006; Boaler, 2015). People with a growth mindset in mathematics learning believed that their mathematical abilities could be developed through learning and training, and their intelligence was malleable rather than fixed.

Turning to other studies, Daly et al. (2019) proposed that students' mindsets could produce either positive or

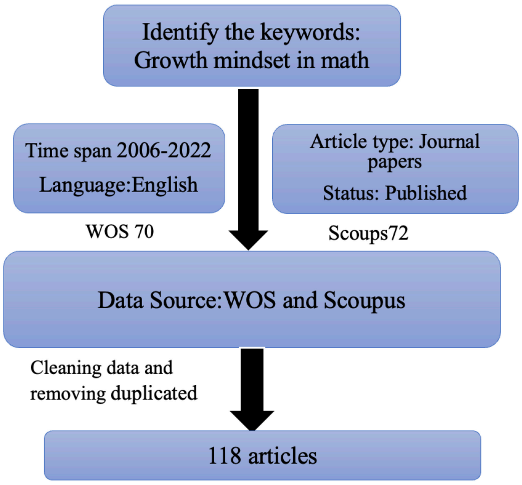


FIGURE 1
Process of data collection.

negative effects on their mathematics learning. The positive effects included students building certain mathematical thinking patterns which could be applied to solving new problems. Therefore, when conditions remained unchanged, the existing thinking patterns could help students quickly process the numbers and formulas, and then associate and mobilize their learned knowledge and skills to quickly respond to the environment. A positive effect enabled people to quickly extract familiar information from the original cognitive structure and choose the correct direction of thinking, thus contributing to the development of new knowledge. In Eichhorn's (2016) study on Indian primary students' number sense, it was found that, to some extent, the students' negative mindset will limit the divergence of their thinking, making it difficult for them to think flexibly in new environments, leading them to be easily influenced by their old thinking. In this sense, a growth mindset should be even more crucial for helping students change their current way of thinking.

Lastly, the growth mindset also has important implications for the development of subjective task values, including intrinsic value, utility value, and attainment value. The growth mindset places greater emphasis on mastery-oriented or learning goals, while the fixed mindset prefers to endorse performance goals (Blackwell et al., 2007; Burnette et al., 2013). Learning goals emphasize the importance of improving individual ability and expanding skill sets. In contrast, performance goals emphasize the importance of demonstrating a high ability (performance approach) and avoiding the external perception of low ability (performance-avoidance) (Elliot and Harackiewicz, 1996). When facing a challenging task, individuals with fixed mindsets would worry about their own incompetency in performing the task, which in turn undermines their intrinsic interest or enjoyment during the process (Dweck, 2008; Stipek and Gralinski, 1996). To conclude, the growth mindset is crucial to mathematics education as it helps students learn and teachers teach. Learning with the growth mindset in mathematics reflects an active learning method for acquiring mathematical knowledge, where the students position themselves to make sense of what they learn.

Therefore, this bibliometric study attempted to systematically review how the growth mindset in mathematics education has developed in recent decades so as to refresh our understanding of the gist of the literature and identify future research directions. A total of 85 studies were examined in this study. The latent topics, representative research work, development trends, collaborative organizations, and annual topic distributions will be discussed in detail.

Methodology

Statistical bibliography is useful in revealing the development of a discipline (Pritchard, 1969). Bibliometrics uses quantitative analysis and statistics to describe pattern relationships within the research topic (Hawkins, 1981; Chen et al., 2018). Bibliometric techniques can identify current research areas and provide a roadmap for further research (Luo et al., 2022). To assess and analyze the journal impact factors of articles, the current study also processed the qualitative data in the literature. The Web of Science (WOS) and Scopus databases were used. The WOS is a broader platform for scientific information, while Scopus is a comprehensive bibliographic database that provides article abstracts and citations of peer-reviewed scientific literature. Combining the two databases is significantly beneficial for reviewing the literature (Echchakoui, 2020).

Data collection

Data were collected from articles published from 2006 to 2022 retrieved from the WOS and Scopus databases. The search strings "growth mindset in math" and "growth mindset in mathematics education" were used to screen titles, abstracts, keywords, and citations to ensure relevance. After removing duplicates, the final sample comprised 85 journals and articles published from 2012 to 2022. Bibliometric analysis was conducted using techniques available in the software RStudio. The analysis reveals and corroborates several patterns in the research trends, journals, countries, and authors that have significantly impacted research on the growth mindset in mathematics education. The present dataset thus provides an updated perspective on research regarding the growth mindset in mathematics education. **Figure 1** illustrates the process of data collection.

Data analysis

The data search in the WOS and Scopus identified 85 articles relevant to the three research questions of this study. First, to determine trends in growth mindset research (RQ1), the number of articles published for each year between 2012 and 2022 was tallied and fitted on a curve. Next, the bibliometric analysis identified themes and networks among the major contributors to growth mindset research (RQ2), enabling the analysis and visualization of collaborations between researchers, as well as relationships between prolific countries/regions and institutions. Lastly, the topics of the 85 articles (RQ3) were extracted from their abstracts using the biblioshiny package for the **R** programming language (Chen et al., 2020). Structural topic modeling then enabled us to incorporate information into our model and understand how articles addressing the same topic may use different word choices in their discussions of the topic.

Results

Our search strings provided the flexibility for capturing various terms used to refer to the growth mindset in mathematics; however, they also yielded irrelevant search outcomes (e.g., research about STEM education) that had to be filtered out from the final sample. **Figure 2** presents the results of our analysis; the table on the left displays key statistics in terms of article and citation counts, countries/regions of origin, and topics identified. The line graph in the middle illustrates the annual count for relevant articles on growth mindset research. It shows that significantly fewer articles related to the growth mindset in mathematics education had been published before 2012; from that year on, academic interest increased as the research topic evolved. Our findings also reveal that the USA, the UK, and Norway were the most productive countries in generating research on the growth mindset in mathematics education (see **Figure 3**). Based on the results of the analyses of themes and keywords, our findings uncover more detail about how to describe the growth mindset in mathematics education. In sum, our findings provide a representative overview of the growth mindset studies in mathematics education, offering valuable research insights for academics and practitioners looking to explore the growth mindset in the future.

Basic summary statistics

Bibliometric indicators were employed to summarize the dataset. The present bibliographic collection includes 85 articles from 2012 to 2022. The majority were journaled articles, early access publications, and conference papers, with only one review paper in the dataset. Regarding the sources for the articles, the table in **Figure 2** shows that 58 periodicals and books were represented. The table also shows that the annual growth rate for the number of articles was 29.24%, with 313 authors represented in the given time span. Turning to the keyword data, the number of author keywords (DE) was 273 words, while the number for the Keywords Plus indicator (ID) was 339. The larger number for ID compared to DE was expected because the former is a more broadly descriptive metric. Lastly, the average number of co-authors per document was 4.19, and the proportion of international co-authorships was 20%.

The line graph in the middle of **Figure 2** depicts the relationship between article publications and year, which illustrates a rising trend. The year 2021 was the year with the most articles published, numbering 23. A similar rising trend is observed for the annual citation count illustrated in the line graph on the right, with 2019 having the highest number of average citations.

Factors relevant to country of origin

To better understand the relationships between the country of origin of the documents, the keywords included in their abstracts, and their authors, we created the three-field plot (a type of Sankey diagram) illustrated in **Figure 3**. The plot indicates the relationships between the top countries and keywords identified among the datasets. The left column ranks the top 10 countries in published articles, namely, the USA, France, the UK, China, Norway, Korea, Germany, Italy, Finland, and Australia. The middle yellow column ranks the top 14 keywords, namely, "growth mindset," "mindset," "motivation," "academic achievement," "anxiety," "implicit theories," "implicit theories of intelligence," "STEM," "mathematics," "mindsets," "adolescence," "stereotype threat," "mathematics achievement," and "grit." Lastly, the right column ranks the top authors according to the number of published articles they have.

The plot shows that articles from Finland and Australia included "motivation" as a keyword, while articles from Korea, China, and Norway included "growth mindset" as a keyword. Articles from the USA targeted the broadest set of keywords, covering 9 out of 14 terms. Finally, the top authors represented in the plot include Murphy, H. Lee, Fink, J. Lee, Canning, Bong, Duckworth, and Frey.

Figure 4 illustrates the distribution of multi-country publications (M) and single-country publications (SCPs) in the dataset by country; MCPs refer to articles with at least one co-author representing a different country than that of the corresponding author. The bar chart shows that the top four countries represented in MCPs are China, Norway, USA, and Germany, while the top four countries represented in SCPs are USA, UK, Korea, and Italy.

Figure 5 illustrates the top 10 most prolific institutions publishing articles on growth mindset research in mathematics

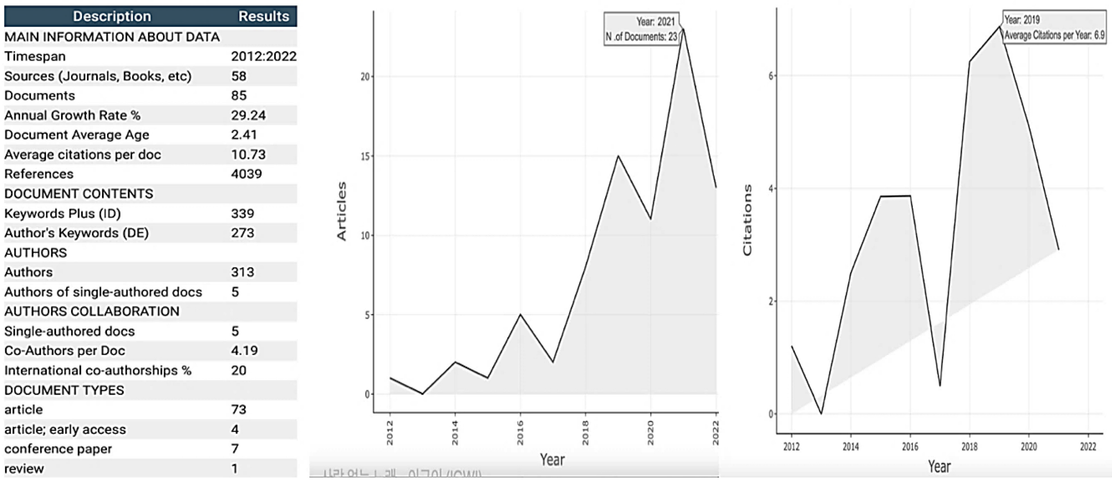


FIGURE 2
Summary statistics **(left)**, trends in the article **(middle)**, and citation **(right)** counts.

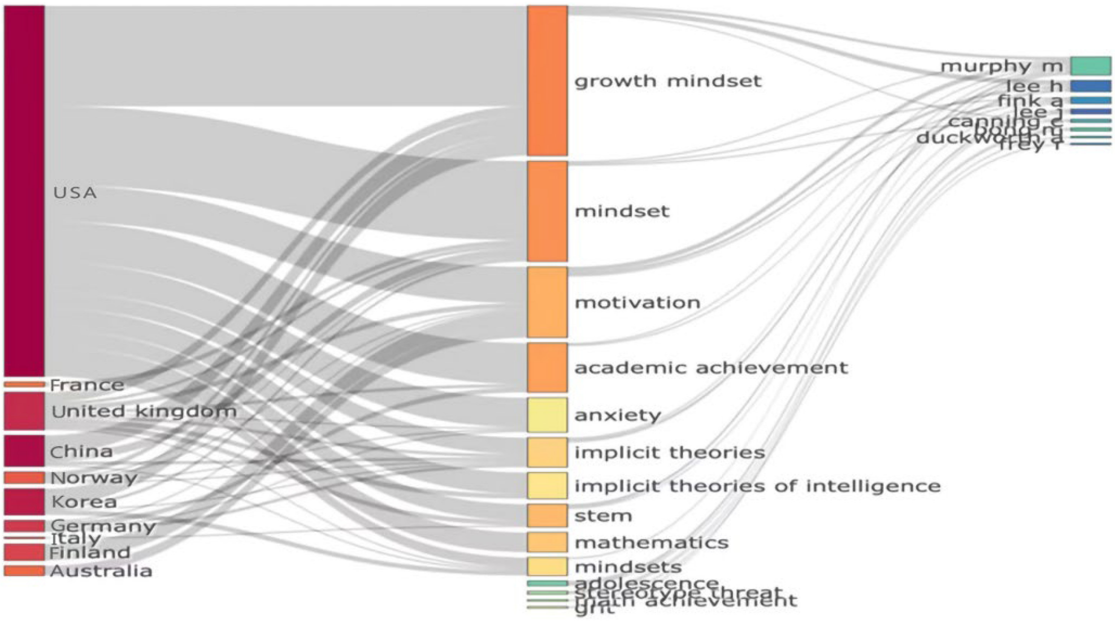


FIGURE 3
Three-field plot for countries of publications, keywords in abstracts, and authors.

from 2012 to 2020, of which 9 are in the US and 1 is in Korea. The top three institutions are Stanford University, Stanford Graduate School of Education, and Korea University.

Sources and authors

In Figure 6, the frequency plot on the left shows the cumulative occurrences of the top five publication sources represented in the dataset. Occurrences generally increased between 2012 and 2022, with individual sources displaying different degrees of fluctuation. The most impactful source was the annual conference proceedings for the American Society for Engineering Education (*ASEE Annu. Conf. Expo.*), followed by four journals including *British Journal of Educational Psychology* (*Br. J. Educ. Psychol.*), *Frontiers in Psychology* (*Front. Psychol.*), *International Journal of STEM Education* (*Int. J. STEM Educ.*), and *Journal of Youth and Adolescence* (*J. Youth Adolesc.*). The frequency plot offers one indicator for the impact of the sources over time by illustrating periods of increase and plateauing. For

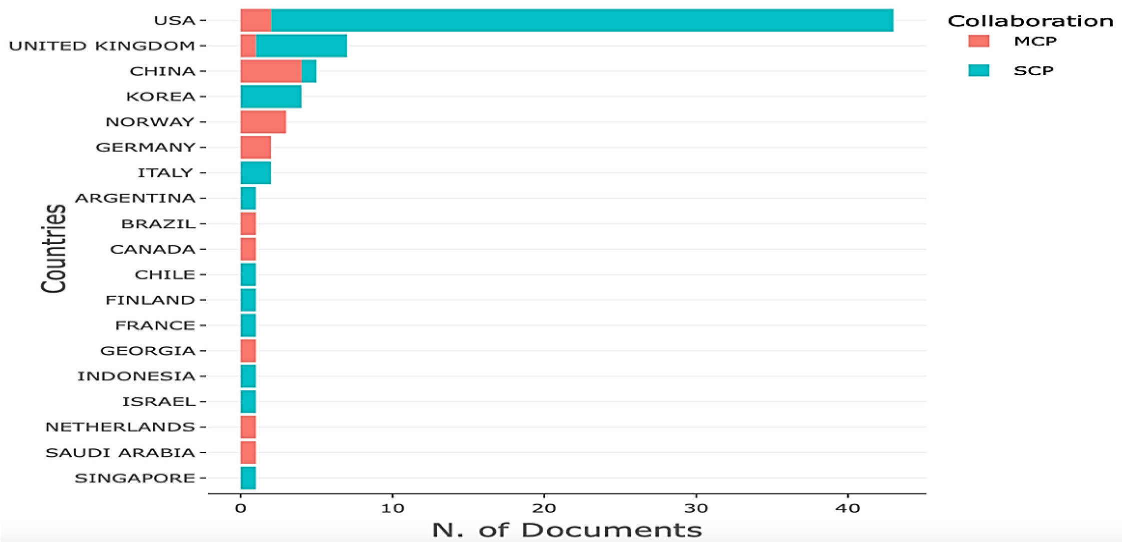


FIGURE 4
Multi-country (MCPs) and single-country (SCPs) publications by country.

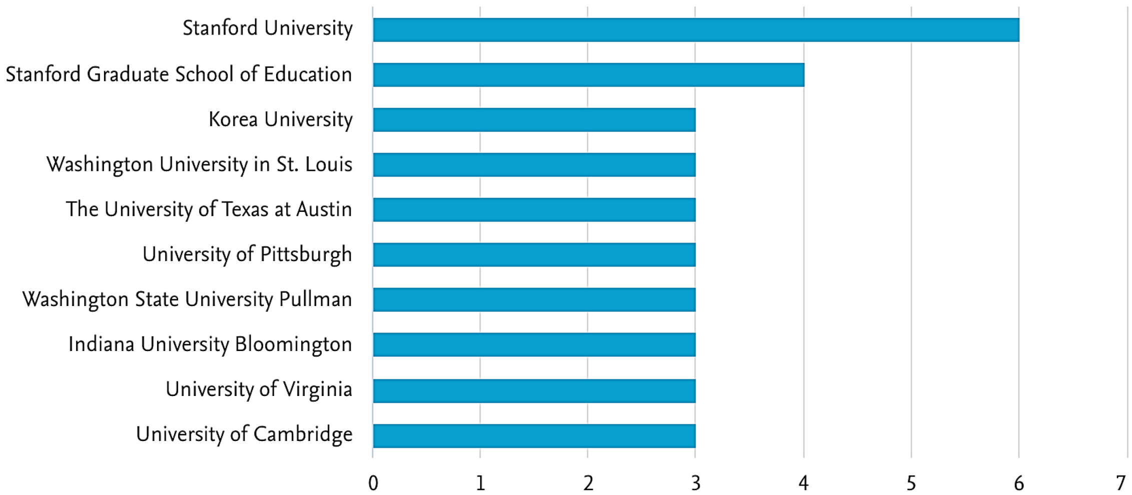


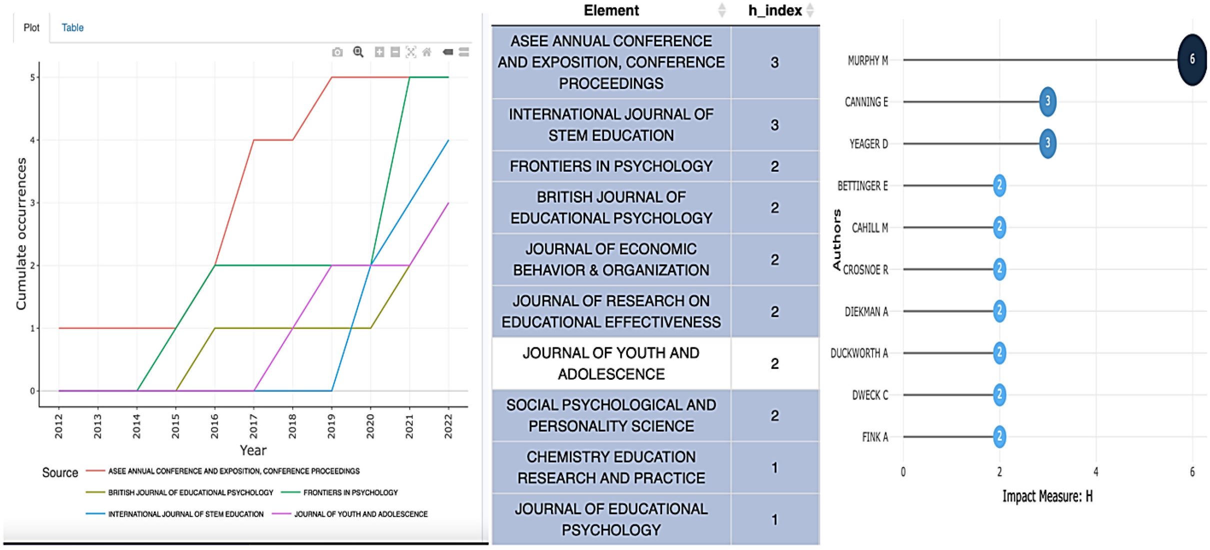
FIGURE 5
Top 10 most prolific institutions.

instance, the number of documents published in *Front. Psychol.* and *J. Youth Adolesc.* remained at two during the years 2016–2020 and 2019–2021, respectively. Similarly, the number for *Br. J. Educ. Psychol.* stagnated at only one document during the years 2016–2020.

Another indicator for discovering the impact of sources is the *h*-index (Hirsch index), which measures the number of published articles (*h*) by an author or journal that have been cited at least *h* times (Aria and Cuccurullo, 2017). The table in the middle of **Figure 6** shows *ASEE Annu. Conf. Expo.* and *Int. J. STEM Educ.* to be the top two sources, each with an *h*-index of 3. Notably, the two journals with an *h*-index of 1—*Chemistry Education Research and Practice* and *Journal of Educational Psychology*—include articles published by the top three most impactful authors by *h*-index listed in the graph on the right of **Figure 6**, namely, Murry (6), Canning (3), and Yeager (3).

Keyword and topic distributions

Among the different groups of significant terms related to publications, keywords indicate essential concepts found in the abstract and main text of articles while also functioning as search terms that help readers easily find key themes, providing



Element	h_index
ASEE ANNUAL CONFERENCE AND EXPOSITION, CONFERENCE PROCEEDINGS	3
INTERNATIONAL JOURNAL OF STEM EDUCATION	3
FRONTIERS IN PSYCHOLOGY	2
BRITISH JOURNAL OF EDUCATIONAL PSYCHOLOGY	2
JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION	2
JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS	2
JOURNAL OF YOUTH AND ADOLESCENCE	2
SOCIAL PSYCHOLOGICAL AND PERSONALITY SCIENCE	2
CHEMISTRY EDUCATION RESEARCH AND PRACTICE	1
JOURNAL OF EDUCATIONAL PSYCHOLOGY	1

FIGURE 6
Relationships between authors, publication sources, and *h*-index.

more information to guide data searches. Words included in an abstract provide an overview of the content of the manuscript and a guide to its essential written components, capturing the core of an article. Lastly, the titles themselves convey the principal topics of studies and highlight the significance of their findings to attract readers (Chen et al., 2018). Consequently, identifying author keywords, associated phrases, and important terms in titles and abstracts is essential for understanding the essence of a group of articles.

Figure 7 illustrates word clouds generated using the bibliometric package in RStudio, with larger-sized words representing terms that appeared more often across articles. In the author keyword cloud, the top words are "mathematics," "motivation," "mindset," "STEM," and "implicit theories." Additional important terms in the "Keywords Plus" cloud include "stereotype threat," "intelligence," "achievement," "students," "beliefs," "performance," and "science." Lastly, other important terms found in the title and abstract word clouds include "growth", "intervention", "learning", and "math". The thematic connections among these words may represent latent trends in research on the growth mindset (Figure 3).

While word clouds are effective in visualizing keywords, they are insufficient for understanding the connection between these important terms and the topics they address. Figure 8 presents a conceptual map that attempts to depict the connections between concepts and ideas using multidimensional scaling (MDS). In the map, words that have similar distributions along the two dimensions appear closer together (Aria and Cuccurullo, 2017). This map depicts the average position of all column profiles. The distributions of keywords on the map illustrate the two core topics of mathematics and education, with keyword clusters related to the topics falling within the respective polygons that represent them, namely, pink for mathematics and blue for education. The size of the polygons and the number of keyword points within them show that mathematics is the more broadly addressed topic, including 53 keywords such as "mindset conception," "impact," "test performance," "growth mindset," and "ability." Conversely, the blue triangle representing education only includes six keywords: "education," "professional aspects," "students," "engineering education," "teaching," and "technology".

To further understand the themes addressed by the topics and keywords included in the articles, mixed methods thematic network analysis can be conducted to clarify the relationships between concepts and terms. It has been discovered that a single broad overarching thread deriving from a growth mindset connects further to the keyword, thereby establishing a thematic link between the nature of the study where the research is being carried out. Figure 9 presents a bubble chart generated to visualize the various themes addressed by the articles. The distances of thematic network bubbles from the central axes are functions of their relevance to growth mindset research in mathematics education and their degree of development. The chart shows that thematic networks are distributed across all four quadrants. Highly developed and isolated themes in the upper-left quadrant include engagement, implementation, persistence, children, fluid intelligence, and skills. Declining themes in the lower-left quadrant include computers, individual differences, and technology, while emerging themes in the same quadrant include adolescents, stereotype threats, and academic achievement. Lastly, the basic transversal themes

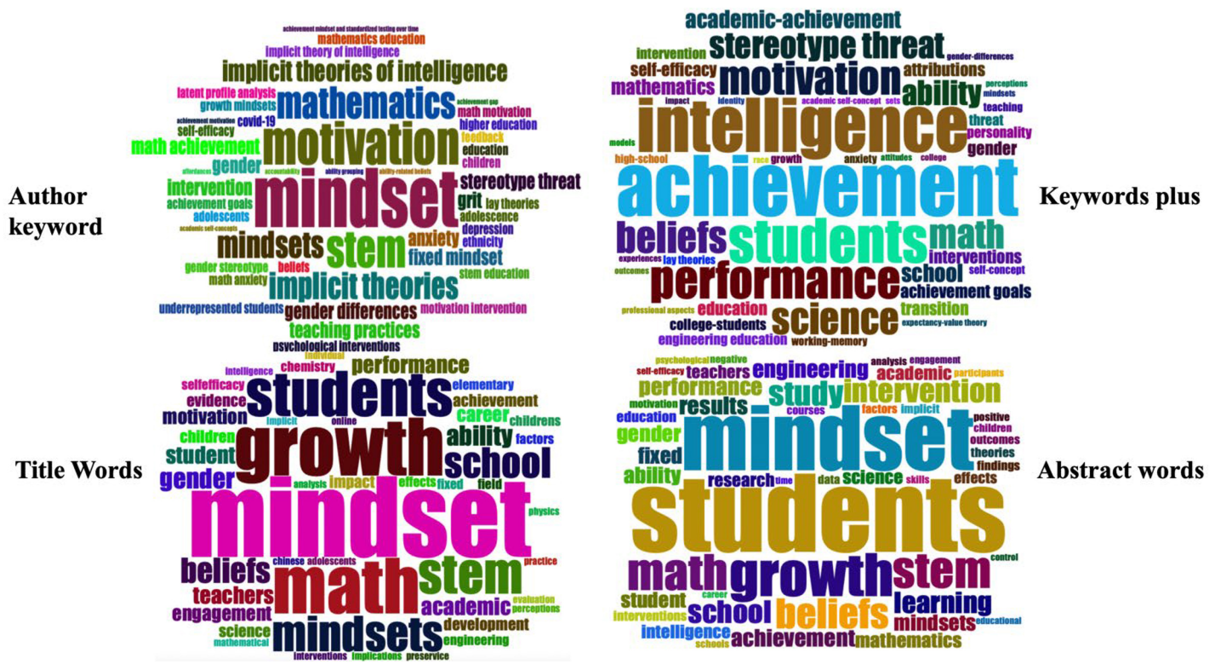


FIGURE 7
Clockwise from **(top left)**: word clouds representing author keywords, "Keywords Plus" phrases, words in abstracts, and words in titles.

in the lower-right quadrant include education, teaching, and entrepreneurship education. Only a few thematic bubbles straddle two quadrants, such as engineering education, professional aspects, and underrepresented minorities in the bottom quadrants. This indicates that there are still a few topic contents related to the growth mindset.

In the graph showing the distribution and development of research topics over time, a shift over the ten years can be noticed (see **Figure 10**). While topics in a given research field may have something in common, they can also branch out into other areas of study. Building on the resurgent interest in mapping scientific paradigms (e.g., using a flow structure), **Figure 10** presents a flow diagram indicating the longitudinal evolution of research themes included in the dataset. The left side lists the top keywords in growth mindset research during 2012–2020, namely, "children," "implicit theories," "students," and "science." The right side lists the top keywords for 2021–2022, namely, "performance," "implicit theories," "stereotype threat," "achievement goals," "individual differences," and "anxiety." The flows that connect the two lists of keywords illustrate an evolution in the research from the concept of mindsets to implicit theories with the growth mindset alongside academic achievement. Notably, "implicit theories" has branched out into "individual differences" and "anxiety", while "students" have fed back into "implicit theories." This visualization of keyword trends offers more explicit details about the evolution of growth mindset research, specifically illustrating the merger of "implicit theories" and "students".

Discussion

This section discusses the results of our research and introduces the potential implications for future practice and research.

Growth mindset research in different countries

As regards the results of factors relevant to the country of origin (section "Factors relevant to country of origin"), they are used to visualize the structure, description, and monitoring of published research in a particular field (Garfield et al., 1964). **Figure 3** depicts a three-field plot hosting the corresponding author's country, the name of the authors, and keywords that define the central theme of the article. **Figures 4**, **5** refer to the institutions wherein the authors are either employed or are professionally associated to undertake the research in their respective field of study. They all unveil several factors relevant to growth mindset research in different countries by examining keywords, institutions, and sources. Describing and analyzing the topic of the growth mindset from an empirical perspective, we surveyed important bibliometric factors related to countries, sources, authors, and keywords. We also identified associations between the emergence of growth mindset research and several important keywords such as motivation, intelligence, and performance. Our findings indicate that the USA has the

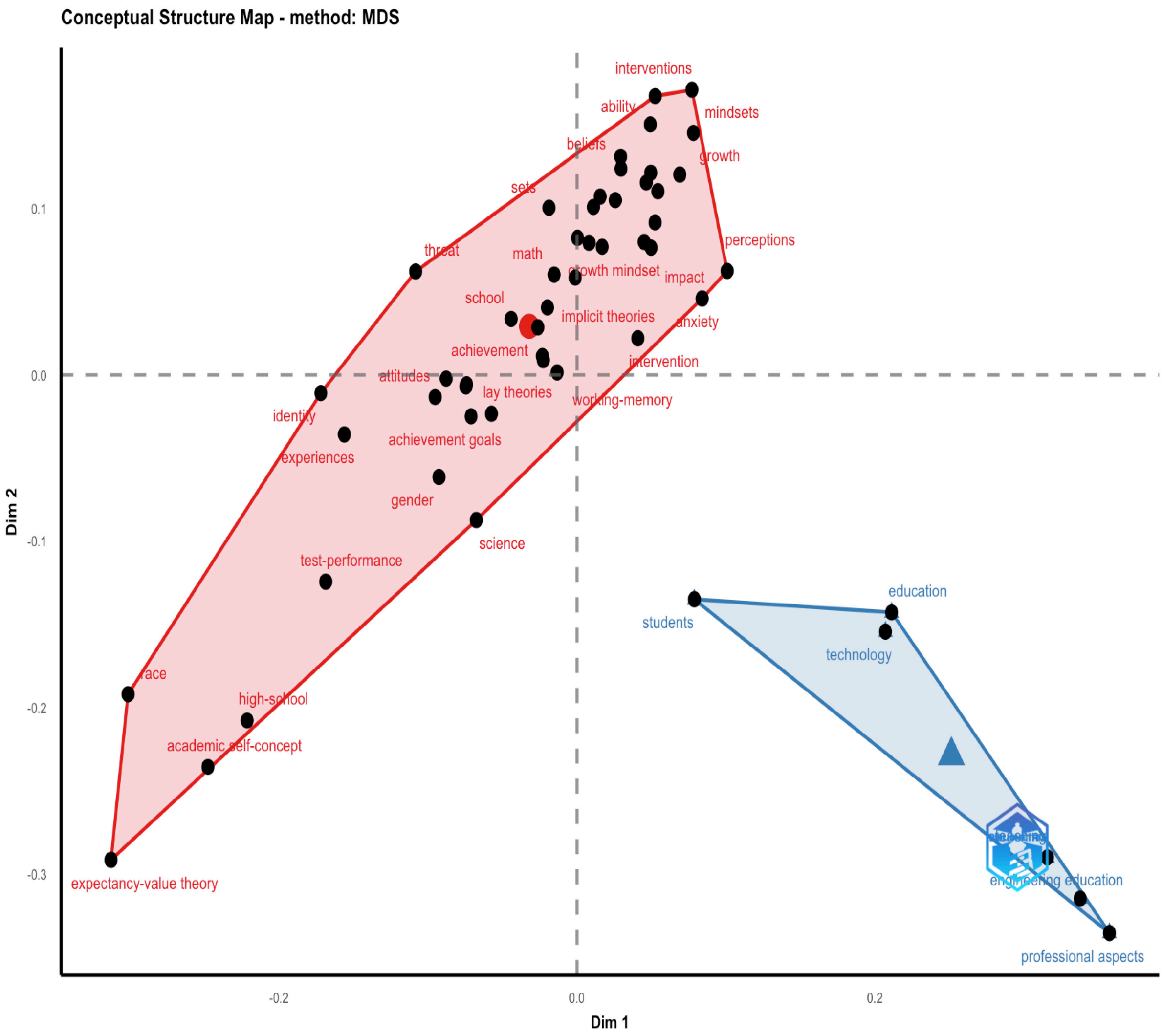


FIGURE 8
The conceptual structure of factor analysis.

most articles discussing the growth mindset. The reason is that the concept originated in this country. Dweck (1999) made a very substantial contribution when she introduced the idea of a growth mindset in 1999. The definition has since been refined and added to. By 2015, Dweck proposed that competencies can be developed through dedication and hard work, and it is this perspective that creates the love of learning and the resilience that is essential to outstanding achievement (Dweck, 2015). Articles from the USA also targeted the broadest set of top keywords compared to articles from other countries. However, it is also noted that Korea, China, and Singapore have also made substantial contributions (see **Figures 3**, **4**). This is an encouraging sign that not only Western but also Asian academics are gradually developing and investigating the growth mindset in their own educational systems. However, does the study conducted in each country have a unique interpretation of the development mindset? Are there any differences or special features in the development of a mathematical growth mindset across different regions? The answers to these questions deserve further deep exploration.

Considering the concept of the growth mindset, which originated in the USA, Western researchers have developed a deeper conceptual understanding of the topic and produced more research output than others. Understanding the influence of the growth mindset, particularly the intervention of changing the fixed mindset, is meaningful because it may be used in countries with diverse student populations, such as the USA—combatting stereotype threats successfully could have a significant impact on student achievement (Altbach, 2004). Around the world, nations are becoming more and more diverse and multicultural, and immigrants may enter nations via land, sea, and air. Regarding the complexity of cultural context for

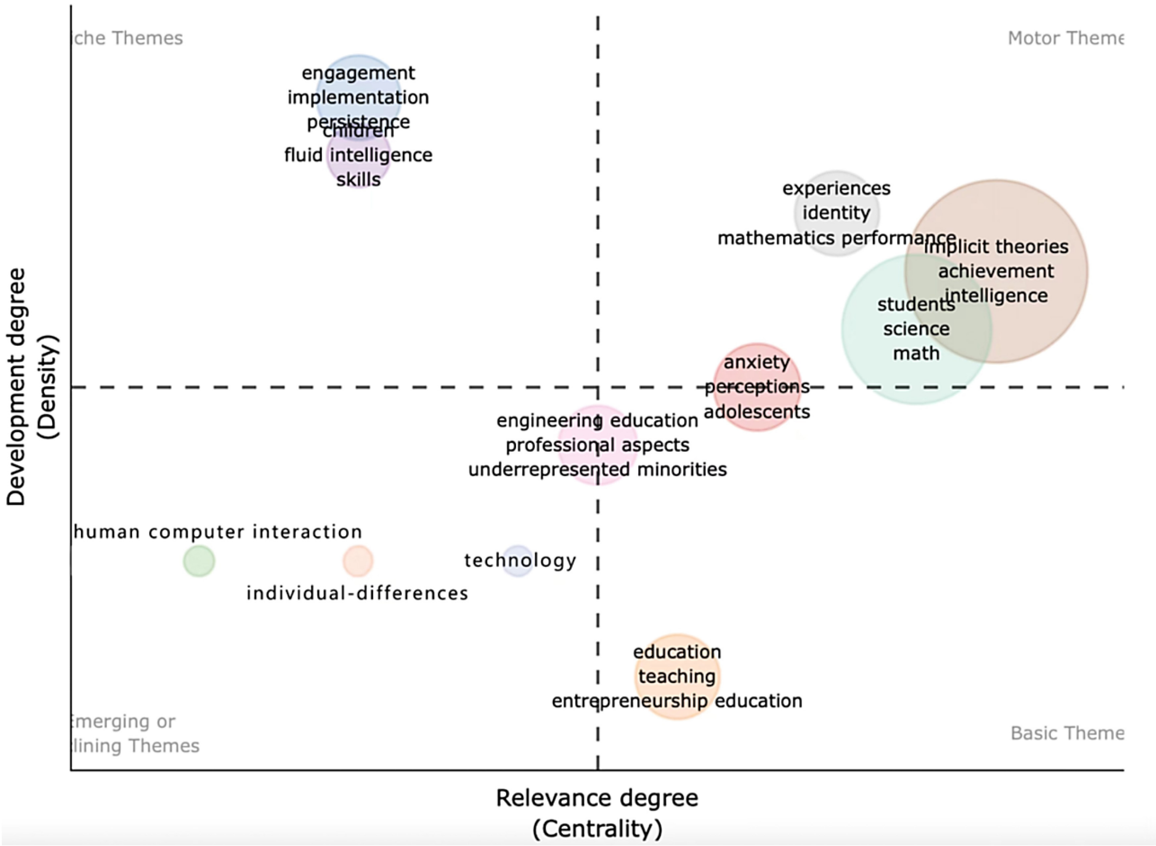


FIGURE 9
Bubble chart mapping thematic networks.

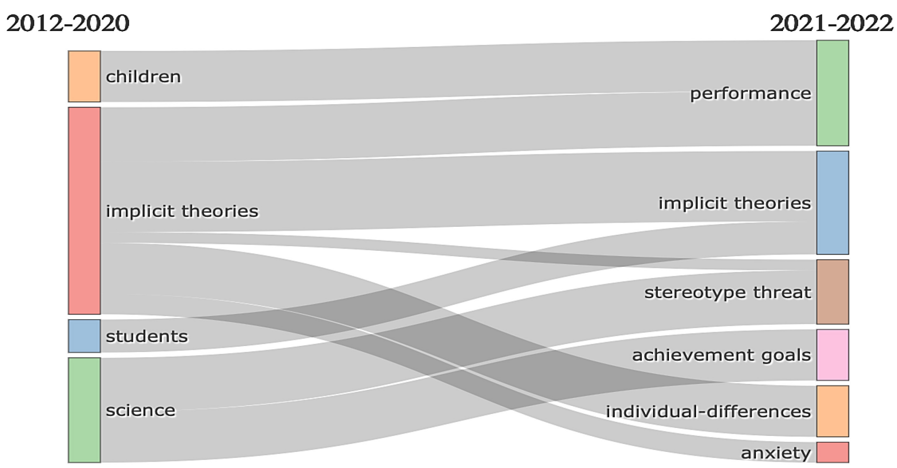


FIGURE 10
Flow diagram of longitudinal thematic evolution.

different students, a culture-specific growth mindset (e.g., Zeeb et al., 2020) deserves future research attention. Moreover, the focus on mathematics itself is a culturally transmitted body of knowledge (Stigler and Baranes, 1988). For instance, the system of Hindu–Arabic numerals that we typically use to represent numbers is applied in the majority of the world, particularly in schools, thus promoting a level of commonality across cultures in mathematical knowledge. Nonetheless, if we combine a growth mindset with different cultures of mathematics learning, does this concept show commonality? What will transpire? Will this be modified or accepted? At this point, it is worthwhile to investigate the applicability of growth mindset theory to other geographic regions, such as Asia, which requires further exploration that considers different cultural backgrounds and beliefs about mathematics learning. Nelson et al. (1993) emphasized the importance of culture in the classroom for the mental health of students living in multicultural neighborhoods. They gave examples of some mathematics topics, showing different approaches to these topics developed in different cultures. Students have a chance to appreciate that there are many different ways of arriving at the same answer. In this setting, the use of multiculturalism is important and necessary in this context in order to foster better self-development and intercultural understanding. Therefore, future research should consider more diverse cultural backgrounds and perceptions of growth mindsets in mathematics learning.

Growth mindset research in students' learning

Our findings identified four groups of important terms in the dataset, namely, author keywords, "Keyword Plus" phrases, title words, and abstract words. Excluding search terms, the most frequent keywords were "motivation", "STEM", "implicit theories", "intelligence", "achievement", "beliefs", "anxiety", and "intervention". In conclusion, mathematical mindsets or implicit theories ("implicit theories") include students' self-concept and self-efficacy beliefs ("beliefs), as well as their attitude about failure, which includes math anxiety ("anxiety"), all of which determine their willingness to learn mathematics (Meece et al., 1990; Panaoura et al., 2009). Additionally, students' mathematical beliefs are often cyclically related to their academic achievement ("achievement"); in turn, positive feedback on academic achievement provides students with positive motivation to learn (Ross et al., 2012). Therefore, students' academic achievement ("achievement") and motivation ("motivation") are closely related to the growth mindset (Zhao et al., 2018). On the other hand, a growth mindset sees intelligence ("intelligence") as a moldable trait that is not fixed and could be improved through effort and intervention ("intervention") (Macnamara and Rupani, 2017; Zhu et al., 2019). Additionally, the theories and intervention methods related to mathematical beliefs could also be applied to other STEM-related ("STEM") disciplines (Boaler et al., 2021).

When we examine each keyword thoroughly, it is not difficult to conclude that there are insufficient growth-mindset-related keywords. The available analysis is very simplistic, and no additional keywords interacted. Perhaps the lack of growth-mindset-related research publications is the reason for this study constraint. However, if we examine the field of mathematics education, it is widely considered that the addition of the growth mindset theory to future study will yield greater opportunities. Instead of just examining and comprehending this theory, we expect that in the future, the direction and content of the analysis will be clarified.

Moving on to focusing on the keywords and determining their commonality between the research areas, the conceptual structure map, as represented in **Figure 8**, provides a high-level overview of the keyword clusters analyzed from the bibliometric information. Two clusters were formed using multivariate correspondence analysis by determining their commonality. Surprisingly, segment topics that have gained traction are "academic self-concept", "test-performance", "lay theories", and "expectancy-value theory". It seems that researchers are attempting to explain and describe the growth mindset through various theories and students' academic performance. Regarding the studies by Sarrasin et al. (2018) and Yeager et al. (2019), the growth mindset had positive effects on student motivation and academic performance. However, the findings of the keyword analysis in growth mindset revealed no topics related to specific subfields in mathematics, such as algebra, calculus, or geometry. Learning mathematics should help students acquire knowledge and skills in arithmetic, algebra, geometry, statistics, space, and structure logically and systematically (Vinner, 2002). As for the teaching of mathematics, different parts of mathematics may require different pedagogies to inspire students' mathematical thinking. Hence, researchers may need to incorporate mathematical content in the field of growth mindset research in the future, to foster the growth mindset among students in mathematics learning. The previous literature review indicated that learning with a growth mindset could be considered a very promising approach for students in learning mathematics. However, in the results of the keyword analysis, we did not see any keywords related to students' learning, such as classroom learning or classroom activity.

Another gap in the research of the analysis of the keywords was that it did not address the topic of teachers' mindsets and parents' mindsets. The influence of teachers and parents on students' mathematics learning has been well-investigated in the literature. It seems necessary to investigate the impact of teachers' mindsets or parents' mindsets on students' growth mindsets as well. Analysis in such research could be applied to the teacher's mindset of their past learning experience and associated with teaching practices. For instance, if a teacher does not have a growth mindset in mathematics teaching and learning, can the teacher still teach mathematics effectively? Can a teacher's mindset be changed when the teacher wants to change their students' fixed mindset? What would be the differences between teachers with and without a growth mindset, when they aim to promote their students' growth mindset? Few studies mentioned the relationship between the characteristics of mathematics teachers and their growth mindset (Shoshani, 2021). Those questions related to teachers are possible inquiries for further study.

To conclude, research associated with more keywords needs to be discussed, including comparative studies of mathematics learning at different grade levels, student learning habits, self-regulation, self-efficacy, mathematical thinking, and even teachers' mindsets. In addition to considering the topic matter itself, researchers should also consider the learning experiences of students. The development of a growth mindset that enables learners to focus their attention is necessary for learners to raise their interest and motivation in mathematics learning; these strategies should be discussed in future research. Finally, in practical terms, future educators should investigate how to develop growth mindsets and skills progressively among their students, implementing their insights into teaching designs.

Growth mindset research in different themes

When facing the future challenges of a complex and uncertain world, school education is undergoing a competency-based curriculum reform (Organisation for Economic Co-operation and Development [OECD], 2018). Students who are best prepared for the future are change agents. They need a broad set of knowledge, skills, attitudes, and values in action, including broad and specialized knowledge, cognitive and meta-cognitive skills, social and emotional skills, and practical and physical skills. The use of this broader range of knowledge and skills will be mediated by attitudes and values (Organisation for Economic Co-operation and Development [OECD], 2018). To process and apply their knowledge and skills in unknown and evolving circumstances, mathematics can play a crucial role. Mathematics and its applications permeate lots of facets of contemporary life. However, mathematics, the significance of which we feel affects every area of our lives, is not sufficiently learned by many people for various reasons. This might be due to the methods and tactics for learning mathematics, and also may relate to students' mathematics learning difficulty (MLD) (Von Aster and Shalev, 2007; Hartmann, 2013), or mathematics anxiety (Huang et al., 2019; Samuel and Warner, 2021). This may further explain why the thematic evolution in **Figure 10** is trending toward concretization, from student themes to an emphasis on student anxiety and individual differences in mathematics learning, for example. However, if we further examine the thematic network, it is evident that various themes are involved, and their connections appear to be very scattered and macroscopic, such as gender and intelligence. Obviously, it is evident that those studies are still shallow (as in **Figures 8**, **9**). Surprisingly, **Figure 9** involves two variables: entrepreneurial education and technology. It can be concluded that even though the concept of the growth mindset is being investigated in depth, the related connections still need to be explored more actively. Such as how to use this concept of a growth mindset sufficiently to link technology to game-based learning in the classroom. Can entrepreneurial education contribute to the growth of students' leadership capabilities? These associations are not seen in the figure, and these factors are not adequately studied in the mathematics discipline. No factors such as the growth mindset versus geometry or algebra were addressed in mathematics. Researchers are still regarding the discipline of mathematics as a whole, and lack consideration of different learning areas in mathematics, which may make a difference in the development of a mathematical mindset. Some specific questions can be further explored, such as how can growth mindsets can be used to facilitate student learning in the domain of numbers.

Conclusion

This review focused on research about the growth mindset in mathematics, drawing on article metadata from two different databases. Since researchers have begun exploring the growth mindset in mathematics education, an increasing trend in research outputs has been observed since 2012, with achievement and academic success becoming popular topics in education research. Given that no prior study has used quantitative analysis and statistics to investigate pattern relationships in the field of the growth mindset as a research topic, the current study adopted the bibliometric package in RStudio to analyze 85 studies published from 2012 to 2022, revealing notable trends and hidden relationships in growth mindset research.

The findings of this study address prevalent subject areas and find new networks of research topics for the growth mindset in mathematics education. They may help mathematics educators gain a deeper, more diverse understanding of current research on the theme, which can then help them design or explore possible effective strategies for the development of students' growth mindset. However, we acknowledge that this review remains limited as it only analyzed limited journal articles published within the past decade. Different types of documents such as research reports or book chapters from more databases can be considered. Nevertheless, based on the results of this review, we make several recommendations for future practice and research. Our findings suggest that additional aspects should be considered in research on the growth mindset in mathematics teaching and learning, including mathematical knowledge, cultural differences, and learner characteristics. To sum up, this review contributes to the understanding of the primary topics in the research on the growth mindset in mathematics, the concept of the growth mindset, and possible directions for further research on the growth mindset in mathematics education.

Author contributions

XX, QZ, and JS: conceptualization, writing, review, and editing the manuscript. XX, YW, and QZ: methodology. XX and YW: formal analysis. XX and QZ: writing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Seed Funding Grant, the Education University of Hong Kong under Project R4248; and partially supported by Hong Kong Research Grants Council under Grant ECS-28609121.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Altbach, P. (2004). Higher education crosses borders: Can the United States remain the top destination for foreign students? *Change Mag. High. Learn.* 36, 18–25. doi: 10.1080/00091380409604964

Aria, M., and Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* 11, 959–975. doi: 10.1016/j.joi.2017.08.007

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Dev.* 78, 246–263. doi: 10.1111/j.1467-8624.2007.00995.x

Boaler, J. (2015). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching*. San Francisco, CA: Jossey-Bass.

Boaler, J., Dieckmann, J. A., LaMar, T., Leshin, M., Selbach-Allen, M., and Pérez-Núñez, G. (2021). The transformative impact of a mathematical mindset experience taught at scale. *Front. Educ.* 6:784393. doi: 10.3389/feduc.2021.784393

Burnette, J. L., O'boyle, E. H., VanEpps, E. M., Pollack, J. M., and Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychol. Bull.* 139, 655–701. doi: 10.1037/a0029531

Chen, C., Zhang, Z., and Jia, M. (2020). Effect of stretch goals on work–family conflict: Role of resource scarcity and employee paradox mindset. *Chin. Manag. Stud.* 14, 737–749. doi: 10.1108/CMS-06-2019-0240

Chen, X., Hao, J., Chen, J., Hua, S., and Hao, T. (2018). "A bibliometric analysis of the research status of the technology enhanced language learning," in *Lecture notes in computer science*, Vol. 11284, eds T. Hao, W. Chen, H. Xie, W. Nadee, and R. Lau (Cham: Springer), doi: 10.1007/978-3-030-03580-8_18

Chew, S. L. (2014). "Helping students to get the most out of studying," in *Applying the science of learning in education: Infusing psychological science into the curriculum*, eds V. A. Benassi, C. E. Overson, and C. M. Hakala (Washington, DC: Society for the Teaching of Psychology), 215–223.

Daly, I., Bourgaize, J., and Vernitski, A. (2019). Mathematical mindsets increase student motivation: Evidence from the EEG. *Trends Neurosci. Educ.* 15, 18–28. doi: 10.1016/j.tine.2019.02.005

Dweck, C. S. (2006). *Mindset: The new psychology of success (Updated edition)*. Morrisville, NC: Lulu Press, Inc.

Dweck, C. S. (2008). *Mindsets and math/science achievement*. New York, NY: Carnegie Corporation of New York.

Dweck, C. S. (2015). Carol Dweck revisits the growth mindset. *Educ. Week* 35, 20–24.

Dweck, C. S. (2017a). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychol. Rev.* 124, 689–719. doi: 10.1037/rev0000082

Dweck, C. S. (2017b). *Mindset: Changing the way you think to fulfil your potential*. New York, NY: Random House.

Dweck, C. S., Chiu, C. Y., and Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychol. Inquiry* 6, 267–285. doi: 10.1207/s15327965pli0604_1

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press. doi: 10.4324/9781315783048

Echchakoui, S. (2020). Why and how to merge scopus and web of science during bibliometric analysis: The case of sales force literature from 1912 to 2019. *J. Mark. Anal.* 8, 165–184. doi: 10.1057/s41270-020-00081-9

Eichhorn, M. S. (2016). How can 5+ 6= 7? Exploring the use of a screening tool to investigate students' mathematical thinking in class two in Kolkata, India. *Asia Pac. J. Dev. Differ.* 3, 254–286. doi: 10.3850/S2345734116000302

Elliot, A. J., and Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *J. Pers. Soc. Psychol.* 70, 461–475. doi: 10.1037/0022-3514.70.3.461

Garfield, E., Sher, I. H., and Torpie, R. J. (1964). *The use of citation data in writing the history of science*. Philadelphia PA: Institute for Scientific Information Inc. doi: 10.21236/AD0466578

Guidera, I. A. (2014). *Principals implementing growth mindset norms: Insights on school culture reform*. Los Angeles, CA: University of California.

Hakim, L. L., and Nurlaelah, E. (2018). Mathematical mindsets: The abstraction in mathematical problem solving. *J. Phys. Conf. Ser.* 1132:012048. doi: 10.1088/1742-6596/1132/1/012048

Hartmann, G. M. (2013). *The relationship between mindset and students with specific learning disabilities (Master Thesis)*. Arcata, CA: Humboldt State University.

Hawkins, D. (1981). Testing for equality of variance of correlated normal variables. *Stat. Neerl.* 35, 39–47. doi: 10.1111/j.1467-9574.1981.tb00709.x

Huang, X., Zhang, J., and Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: The gender moderating effect. *Eur. J. Psychol. Educ.* 34, 621–640. doi: 10.1007/s10212-018-0403-z

Kamins, M. L., and Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Dev. Psychol.* 35, 835–847. doi: 10.1037/0012-1649.35.3.835

Luo, F., Li, R. Y. M., Crabbe, M. J. C., and Pu, R. (2022). Economic development and construction safety research: A bibliometrics approach. *Saf. Sci.* 145:105519. doi: 10.1016/j.ssci.2021.105519

Macnamara, B. N., and Rupani, N. S. (2017). The relationship between intelligence and mindset. *Intelligence* 64, 52–59. doi: 10.1016/j.intell.2017.07.003

Meece, J. L., Wigfield, A., and Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *J. Educ. Psychol.* 82, 60–70. doi: 10.1037/0022-0663.82.1.60

National Council of Teachers of Mathematics (2014). *Principles to actions: Ensuring mathematics success for all*. Reston, VA: National Council of Teachers of Mathematics.

Nelson, D., Joseph, G. G., and Williams, J. (1993). *Multicultural mathematics: Teaching mathematics from a global perspective*. Oxford: Oxford University Press.

Ng, B. (2018). The neuroscience of growth mindset and intrinsic motivation. *Brain Sci.* 8:20. doi: 10.3390/brainsci8020020

Organisation for Economic Co-operation and Development [OECD] (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: PISA, OECD Publishing, doi: 10.1787/9789264190511-en

Organisation for Economic Co-Operation and Development [OECD] (2018). *The future of education and skills: Education 2030. OECD education working papers*. Available online at: http://www.oecd.org/education/2030/oecd-education-2030-position-paper.pdf

Ostroff, W. L. (2016). *Cultivating curiosity in K-12 classrooms: How to promote and sustain deep learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Panaoura, A., Gagatsis, A., Deliyianni, E., and Elia, I. (2009). The structure of students' beliefs about the use of representations and their performance on the learning of fractions. *Educ. Psychol.* 29, 713–728. doi: 10.1080/01443410903229437

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *J. Doc.* 25, 348–349.

Ronkainen, R., Kuusisto, E., and Tirri, K. (2019). Growth mindset in teaching: A case study of a finnish elementary school teacher. *Int. J. Learn. Teach. Educ. Res.* 18, 141–154. doi: 10.26803/ijlter.18.8.9

Ross, J. A., Scott, G., and Bruce, C. D. (2012). The gender confidence gap in fractions knowledge: Gender differences in student belief–achievement relationships. *Sch. Sci. Math.* 112, 278–288. doi: 10.1111/j.1949-8594.2012.00144.x

Samuel, T. S., and Warner, J. (2021). "I Can Math!": Reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Community Coll. J. Res. Pract.* 45, 205–222. doi: 10.1080/10668926.2019.1666063

Sarrasin, J. B., Nenciovici, L., Foisy, L. M. B., Allaire-Duquette, G., Riopel, M., and Masson, S. (2018). Effects of teaching the concept of neuroplasticity to induce a growth mindset on motivation, achievement, and brain activity: A meta-analysis. *Trends Neurosci. Educ.* 12, 22–31. doi: 10.1016/j.tine.2018.07.003

Shoshani, A. (2021). Growth mindset in the maths classroom: A key to teachers' well-being and effectiveness. *Teach. Teach.* 27, 730–752. doi: 10.1080/13540602.2021.2007370

Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., and Macnamara, B. N. (2018). To what extent and under which circumstances are growth mindsets important to academic achievement? Two meta-analyses. *Psychol. Sci.* 29, 549–571. doi: 10.1177/0956797617739704

Smiley, P. A., Buttitta, K. V., Chung, S. Y., Dubon, V. X., and Chang, L. K. (2016). Mediation models of implicit theories and achievement goals predict planning and withdrawal after failure. *Motiv. Emot.* 40, 878–894. doi: 10.1007/s11031-016-9575-5

Stigler, J., and Baranes, R. (1988). "Culture and mathematics learning," in *Review of research in education*, ed. E. Rothkopf (Washington, DC: American Educational Research Association), 253–306. doi: 10.3102/0091732X015001253

Stipek, D., and Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *J. Educ. Psychol.* 88, 397–407. doi: 10.1037/0022-0663.88.3.397

Trilling, B., and Fadel, C. (2009). *21st century skills: Learning for life in our times*. San Francisco, CA: John Wiley & Sons.

Vinner, S. (2002). "The role of definitions in the teaching and learning of mathematics," in *Advanced mathematical thinking*, ed. D. Tall (Dordrecht: Springer), 65–81. doi: 10.1007/0-306-47203-1_5

Voogt, J., and Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *J. Curric. Stud.* 44, 299–321. doi: 10.1080/00220272.2012.668938

Vongkulluksn, V. W., Matewos, A. M., and Sinatra, G. M. (2021). Growth mindset development in design-based makerspace: A longitudinal study. *J. Educ. Res.* 114, 139–154. doi: 10.1080/00220671.2021.1872473

Von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Yeager, D. S., and Dweck, C. S. (2020). What can be learned from growth mindset controversies? *Am. Psychol.* 75:1269. doi: 10.1037/amp0000794

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature* 573, 364–369. doi: 10.1038/s41586-019-1466-y

Zeeb, H., Ostertag, J., and Renkl, A. (2020). Towards a growth mindset culture in the classroom: Implementation of a lesson-integrated mindset training. *Educ. Res. Int.* 2020, 1–13. doi: 10.1155/2020/8067619

Zhao, Y., Niu, G., Hou, H., Zeng, G., Xu, L., Peng, K., et al. (2018). From growth mindset to grit in Chinese schools: The mediating roles of learning motivations. *Front. Psychol.* 9:2007. doi: 10.3389/fpsyg.2018.02007

Zhu, P., Garcia, I., Boxer, K., Wadhera, S., and Alonzo, E. (2019). *Using a growth mindset intervention to help ninth-graders: An independent evaluation of the national study of learning mindsets*. New York, NY: MDRC.

frontiers | Frontiers in Psychology

TYPE Methods
PUBLISHED 12 December 2022
DOI 10.3389/fpsyg.2022.1049472

Check for updates

OPEN ACCESS

EDITED BY
Shuhua An,
California State University, Long Beach, United States

REVIEWED BY
Eric Lacourse,
Université de Montréal, Canada
Seock-Ho Kim,
University of Georgia, United States

*CORRESPONDENCE
Tao Yang
yangtao@bnu.edu.cn
Tao Xin
xintao@bnu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 20 September 2022
ACCEPTED 21 November 2022
PUBLISHED 12 December 2022

CITATION
Meng X, Yang T, Shi N and Xin T (2022) Full-information item bifactor model for mathematical ability assessment in Chinese compulsory education quality monitoring.
Front. Psychol. 13:1049472.
doi: 10.3389/fpsyg.2022.1049472

Full-information item bifactor model for mathematical ability assessment in Chinese compulsory education quality monitoring

Xiangbin Meng[1], Tao Yang[2]*, Ningzhong Shi[1] and Tao Xin[2]*

[1]School of Mathematics and Statistics, KLAS, Northeast Normal University, Changchun, China,
[2]Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, China

This study focuses on the measurement of mathematical ability in the Chinese Compulsory Education Qualification Monitoring (CCEQM) framework using bifactor theory. First, we propose a full-information item bifactor (FIBF) model for the measurement of mathematical ability. Second, the performance of the FIBF model is empirically studied using a data set from three representative provinces were selected from CCEQM 2015–2017. Finally, Monte Carlo simulations are conducted to demonstrate the accuracy of the model evaluation indices and parameter estimation methods used in the empirical study. The obtained results are as follows: (1) The results for the four used model selection indices (AIC, SABIC, HQ, BIC) consistently showed that the fit of the FIBF model is better than that of the UIRT; (2) All of the estimated general and domain-specific abilities of the FIBF model have reasonable interpretations; (3) The model evaluation indices and parameter estimation methods exhibit excellent accuracy, indicating that the application of the FIBF model is technically feasible in large-scale testing projects.

KEYWORDS

full-information bifactor item factor model, mathematical ability, item response theory, China compulsory education quality monitoring, large scale testing

1. Introduction

The Chinese Compulsory Education Qualification Monitoring (CCEQM) project (The National Assessment Center for Education Quality, 2018), which is organized by the Basic Education Quality Monitoring Centre (BEQMC) under the Ministry of Education of the People's Republic of China (MOE), is the largest student assessment project in China. CCEQM applies to mathematics, Chinese reading, science, moral education, art, and physical education. Each subject is monitored every 3 years, with a focus on two subjects per year (Jiang et al., 2019). The first assessment cycle ran from 2015 to 2017, with a total of 572,314 fourth-grade and eighth-grade students from 32 Chinese provinces, municipalities, and autonomous regions participating in the assessment (Yin, 2021). In July 2018, the first CCEQM Report was released, attracting considerable attention in China. Furthermore, CCEQM 2015–2017 evaluated students' academic achievements based on the concept of core literacy. Therefore, the work conducted toward CCEQM 2015–2017 provides valuable experience for educational evaluation reform in China.

Mathematics is one of the most important basic subjects in the compulsory education stage of China, and mathematical literacy is the core content assessed by CCEQM 2015–2017. On the basis of the "10 core concepts" put forward by the "Mathematics Curriculum Standard of Compulsory Education (2011 Edition)," and referring to the experience of international large-scale assessment projects such as the Programme for International Student Assessment (PISA) or the Trends in Mathematics and Science Study (TIMSS), a mathematics literacy assessment framework was developed by the experts of BEQMC for CCEQM 2015–2017. Specifically, in the mathematics literacy framework, mathematical ability—as a general concept—includes the five domains of "mathematical computation," "space imagination," "data analysis," "logical reasoning," and "problem solving" (The National Assessment Center for Education Quality, 2018; Jiang et al., 2019). The first four are the same as those in the mathematics literacy framework developed by the "Mathematics Curriculum Standards of Senior High School (2017 Edition)" and the "Mathematics Curriculum Standards of Compulsory Education (2022 Edition)." The domain of "problem solving" was defined by reference to PISA 2012 (OECD, 2014), and covers the ability to discover, analyze, and solve problems. The definitions of mathematical ability, as well as the five domains, are not the focus of this study, so they are not described in detail here.

In CCEQM 2015–2017, the subscores on the five cognitive domains for mathematics are estimated using between-item multidimensional IRT (MIRT) models, and the unidimensional IRT (UIRT) model is used to estimate the overall score for mathematics (Jiang et al., 2019). Note that the UIRT model is also the main measurement model in PISA (OECD, 2014). There are some issues that need attention. First, the subscores on the five domains must be strongly correlated, because all five domains share common cognitive and intelligence influences. The common element is not captured in the between-item MIRT models, which is likely to result in a false interpretation of the subscores. However, the one-factor structure of the UIRT model does not match the five-dimensional assessment framework, in which only the common element is considered, and so the idiosyncratic nature of each domain cannot be explained. Furthermore, as discussed by Jiang et al. (2019), the subscores are hardly comparable with the overall score, as they are obtained from different models. Therefore, it is desirable to develop a more powerful and reasonable measurement model for the evaluation of mathematical ability in large-scale assessment projects.

Bifactor models are a powerful approach for representing a general construct comprised of several highly correlated domains (Chen et al., 2006; Bornovalova et al., 2020), in which the common and unique elements of all domains are modeled separately. The bifactor theory and model were originally proposed by Holzinger and Swineford (1937) to rectify the problem of adequately separating a single general factor of intelligence (Spearman, 1904) from domain factors. To analyze item response data using a bifactor structure, Gibbons and Hedeker (1992) and Gibbons et al. (2007) generalized the work of Holzinger and Swineford (1937) to derive full-information item bifactor (FIBF) models for dichotomous and polytomous response data, respectively. Cai et al. (2011) further extended the FIBF framework to a multiple-group model that supports a variety of MIRT models for an arbitrary mixture of dichotomous, ordinal, and nominal items. After years of relative neglect, bifactor analysis has become an important statistical method for handling multidimensional concepts. The bifactor model has been used primarily in studying intelligence and personality (Gault, 1954; Acton and Schroeder, 2001; Rushton and Irwing, 2009a,b; Watkins, 2010; Martel et al., 2011; McAbee et al., 2014; Watkins and Beaujean, 2014; Cucina and Byle, 2017; Moshagen et al., 2018). Recently, it has become increasingly popular across a broad range of research fields such as depression and anxiety (Simms et al., 2008; Gomez and McLaren, 2015; Kim and Eaton, 2015; Olatunji et al., 2017; Snyder et al., 2017; Jorge-Botana et al., 2019; Heinrich et al., 2020; Waldman et al., 2020; Arens et al., 2021; Caiado et al., 2022), health outcomes (Reise et al., 2007; Leue and Beauducel, 2011; Shevlin et al., 2016; Monteiro et al., 2021), emotion expression (Caiado et al., 2022), and cognitive abilities (McFarland, 2013, 2016; Beaujean et al., 2014; Valerius and Sparfeldt, 2014; Foorman et al., 2015). In addition, as a special case of confirmatory MIRT modeling, FIBF models have been used to address some important problems in psychological and educational measurement. For instance, modeling test response data and identifying the local dependence of item responses (DeMars, 2006; Liu and Thissen, 2012), assessing the dimension of test scales (Immekus and Imbrie, 2008), and equating and vertical scaling of test scores (Li and Lissitz, 2012; Kim and Cho, 2020). It is apparent that the advantages and values of bifactor analysis have been largely verified and are widely recognized.

Motivated by previous studies, this article proposes a bifactor structure for mathematical ability, and applies a mixed FIBF model to assess mathematical achievement. Furthermore, an empirical study is conducted based on data from three representative provinces sampled for the CCEQM 2015–2017 survey. The main tasks of this empirical study are to verify the advantages of the FIBF model over the traditional models of between-item MIRT and UIRT, and to interpret the bifactor scores of mathematical ability. Furthermore, to ensure the accuracy of the empirical analysis, a Monte Carlo simulation study is conducted to investigate the performance of the statistical analysis methods used in the empirical study. Finally, we summarize the research conclusions and review the main results of this study, while identifying its limitations, and discuss some future research issues.

2. FIBF model for mathematical ability in CCEQM 2015–2017

The Mathematics Curriculum Standards of Senior High School (2017 Edition) stated that "...different domains of mathematical literacy are not absolutely different, but integrated with each other..." So the five domains ("mathematical computation," "space imagination," "data analysis," "logical reasoning," and "problem solving") are different and represent different aspects of mathematical ability, but they must have a overlap or common part. From this point of view, we consider that the bifactor structure is suitable for modeling the assessment of mathematical ability.

The item bifactor measurement structure for mathematical ability is illustrated in Figure 1A, where the observed categorical responses are indicated by squares, the latent factors are represented by circles. All items load on a general or common factor, although each item loads on only one group or domain-specific factor. The general factor, which represents the common element of all aspects of mathematical ability, is interpreted as the general mathematical ability, which is a broadly defined concept. The five group factors represent the unique elements of the five domains, and can be interpreted as domain-specific mathematical abilities that are conceptually more narrowly defined mathematical facets. The particularity and commonality of mathematical ability are represented together in the bifactor structure.

When there is no group factor, the bifactor structure is reduced to a one-factor structure; when there is no general factor, it is a five-factor structure. From this point of view, the bifactor model can be thought of as a combination of one-factor and five-factor structures. Because the commonality of the five domains of mathematical ability is explained by the general factor, the general and group factors are assumed to be orthogonal in the bifactor analysis. The orthogonality of general and specific factors is beneficial for evaluating the relative contribution of each factor to the overall test performance. In the following, an FIBF model is proposed for the math test item response data under the bifactor structure of mathematical ability.

Consider $j = 1, ..., M$ items in the total item pool, each scored in $K_j \geq 2$ categories, where $K_j = 2$ indicates that the item is scored dichotomously. Let there be $i = 1, ..., N$ independent students, and let X_{ij} (with x_{ij} representing one observation) denote the response variable from person i to item j. Without loss of generality, we assume that X_{ij} takes integer values from $\{0, 1, ..., K_j - 1\}$. Let $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i}..., \theta_{5i})$ denote the vector of the latent mathematical ability of student i, wherein θ_{0i} denotes the general mathematical ability and $(\theta_{1i}..., \theta_{5i})$ denote the five domain-specific mathematical abilities. The mathematics testing instrument consists of two types of items: dichotomously scored items and graded scored items. Thus, the FIBF model must be a mixed dichotomous and polytomous model. To represent the dichotomous item response data, the bifactor extension of the 2-parameter logistic (2PL) model is used. The 2PL is one of the most important dichotomous IRT models, and is widely used in practice. When item j is a dichotomous item, $X_{ij} \in \{0, 1\}$; $X_{ij} = 1$ denotes the correct response, and $X_{ij} = 0$ otherwise. The conditional probability of the correct response given θ_{0i} and θ_{vi} is formulated as

$$P(X_{ij} = 1|\theta_{0i}, \theta_{vi}) = \frac{\exp(a_{0j}\theta_{0i} + a_{vj}\theta_{vi} + b_j)}{1 + \exp(a_{0j}\theta_{0i} + a_{vj}\theta_{vi} + b_j)}, \quad (1)$$

where a_{0j} and a_{vj} are the item slopes, which are analogous to the factor loading parameters on the general factor and the domain-specific factor, and b_j is the item intercept parameter.

The bifactor extension of the logistic version of GRM is used to represent the graded item response data. When item j is graded, $X_{ij} \in \{0, 1, ..., K_j - 1\}$ and $K_j > 2$. The conditional probability of response category k given θ_{0i} and θ_{vi} is formulated as

$$P(X_{ij} = k|\theta_{0i}, \theta_{vi}) = P(X_{ij} \geq k|\theta_{0i}, \theta_{vi}) - P(X_{ij} \geq k + 1|\theta_{0i}, \theta_{vi}), \quad (2)$$

and

$$P(X_{ij} \geq 0|\theta_{0i}, \theta_{vi}) = 1, \quad (3)$$

$$P(X_{ij} \geq k|\theta_{0i}, \theta_{vi}) = \frac{\exp(a_{0j}\theta_{0i} + a_{vj}\theta_{vi} + b_{kj})}{1 + \exp(a_{0j}\theta_{0i} + a_{vj}\theta_{vi} + b_{kj})}, \quad k = 1, .., K_j - 1, \quad (4)$$

$$P(X_{ij} \geq K_j|\theta_{0i}, \theta_{vi}) = 0, \quad (5)$$

where $b_{1j}, ..., b_{(K_j-1)j}$ are the set of $K_j - 1$ (strictly ordered) intercepts. As before, a_{0j} and a_{vj} are the item slope parameters.

Importantly, the statistical inference (such as parameter estimation and model fit evaluation) using the FIBF models is mature, and a number of software packages have been developed. At present, the application of the FIBF model is technically feasible in modeling the incomplete mixed item response data that often occur in large-scale testing projects. Zhan et al. (2019) proposed a third-order DINA model, which is a cognitive diagnosis model, for assessing scientific literacy in PISA 2015. As verified by Zhan et al. (2019), this third-order Deterministic Inputs Noisy "And" gate model (DINA) model has some advantages over the UIRT model for modeling scientific test data. However, the DINA model is dichotomous, and cannot model polytomous response data, which greatly limits its application in large-scale assessment projects. From this point of view, practicality and feasibility are important benefits of using the FIBF model to measure mathematical ability.

The second-order factor model is an alternative method for representing general constructs consisting of multiple highly related domains (Chen et al., 2006). The second-order factor structure for mathematical ability is illustrated in Figure 1B, in which the five domains of mathematical ability are explained by defining five first-order factors, and correlations among the five domains are identified by stipulating a single second-order factor. In the second-order model, general mathematical ability is conceptualized in terms of a second-order factor. Different from the bifactor model, the effects of the second-order factor on the item responses are mediated by the five first-order factors. Consequently, the first-order factors reflect two sources of variance (general and group), while the group factor in the bifactor model only reflects group effects. The bifactor model directly separates the unique contributions to the item responses of the general and group factors. Compared with the second-order model, the bifactor model makes evaluating theoretical hypotheses about general and group factors clearer and more interpretable. Furthermore, the second-order model is simply a more constrained version of the bifactor model. The second-order structure can be derived from the bifactor structure by constraining the ratio of the weights between any given specific factors and keeping the general factor constant (Reise, 2012). Overall, in contrast to the second-order model, the bifactor model makes theoretical hypotheses more interpretable, and has more degrees of freedom with which to fit the data. In addition, several studies have verified that bifactor models can produce a better fit than second-order models (Morgan et al., 2015; Cucina and Byle, 2017; Bornovalova et al., 2020).

3. Empirical study

3.1. Data description

Data from three representative provinces were selected from the CCEQM 2015–2017 survey, in which 2,017 fourth-grade students (53% males) and 1,404 eighth-grade students (54% males) participated. To ensure the representativeness of the data, three provinces were selected from different geographical regions (east, middle, and west), economic development levels (developed, moderately developed, and less developed), and mathematical academic achievement ranking (high, middle, and low).

Let us introduce the design of the mathematical assessment instrument in CCEQM. To ensure broad content coverage while avoiding an excessive testing burden, the partial balanced incomplete block design was employed to administer the math tests. Specifically, for the fourth-grade students, 59 items were allocated to six test booklets, each booklet consisting of 10 dichotomous items and 8 polytomous items; for the eighth-grade students, a total of 60 items were grouped into six test booklets, each booklet consisting of 12 dichotomous items and 8 polytomous items. Each student was assessed with only one booklet and each booklet was completed by several of the students. In this way, the testing time was held to <2 h, and all five domains of mathematical literacy could be adequately covered. However, the incomplete test administration design resulted in incomplete test data, which increases the difficulty of data analysis. In this empirical study, the R package "mirt" was used to conduct the statistical analysis. This package allows for statistical inferences on multidimensional item response models under incomplete test administration; additionally, it is open source and can be easily obtained.

Three competing models were fitted to the data, namely the FIBF (bifactor structure), between-item MIRT (correlated-factor structure), and UIRT (one-factor structure) models. The MIRT was estimated using the Metropolis-Hastings Robbins-Monro (MH-RM) algorithm of Cai (2010), whereas the FIBF and UIRT were estimated using the expectation-maximization algorithm, with the mathematical abilities of students estimated using the expectation *a posteriori* (EAP) estimation. These estimation methods are commonly used in practice and are known to be powerful. Bifactor models are more general than one-factor, correlated-factor, and second-order factor models, and are thus more prone to overfitting (DeMars, 2006; Murray and Johnson, 2013; Rodriguez et al., 2016; Bonifay and Cai, 2017; Greene et al., 2019; Sellbom and Tellegen, 2019), that is, a bifactor model is inappropriately favored by model selection indices. Therefore, overfitting is an important issue in the use of bifactor models. To avoid unreasonable model fitting, the four commonly used model selection indices of Akaike's information criterion (AIC; Akaike, 1987), Bayesian information criterion (BIC or Schwarz criterion; Schwarz, 1978), Hanna–Quinn index (HQ; Hannan and Quinn, 1979), and sample size-adjusted BIC (SABIC; Sclove, 1987) were computed to compare the model fitting.

3.2. Results

3.2.1. Comparison of models

First, the obtained values of the four model selection indices (AIC, SABIC, HQ and BIC) for the three competing models are reported in Table 1. The four model selection indices of the FIBF model are consistently smaller than those of the other two models, with the largest values given by the between-item MIRT model. These results consistently support the FIBF model as the best for fitting this empirical data. The fit of the UIRT model is better than that of the MIRT model.

Further, the estimated correlation coefficients of the five latent abilities of the between-item MIRT model are given in Figure 2. All values are larger than 0.7, and most

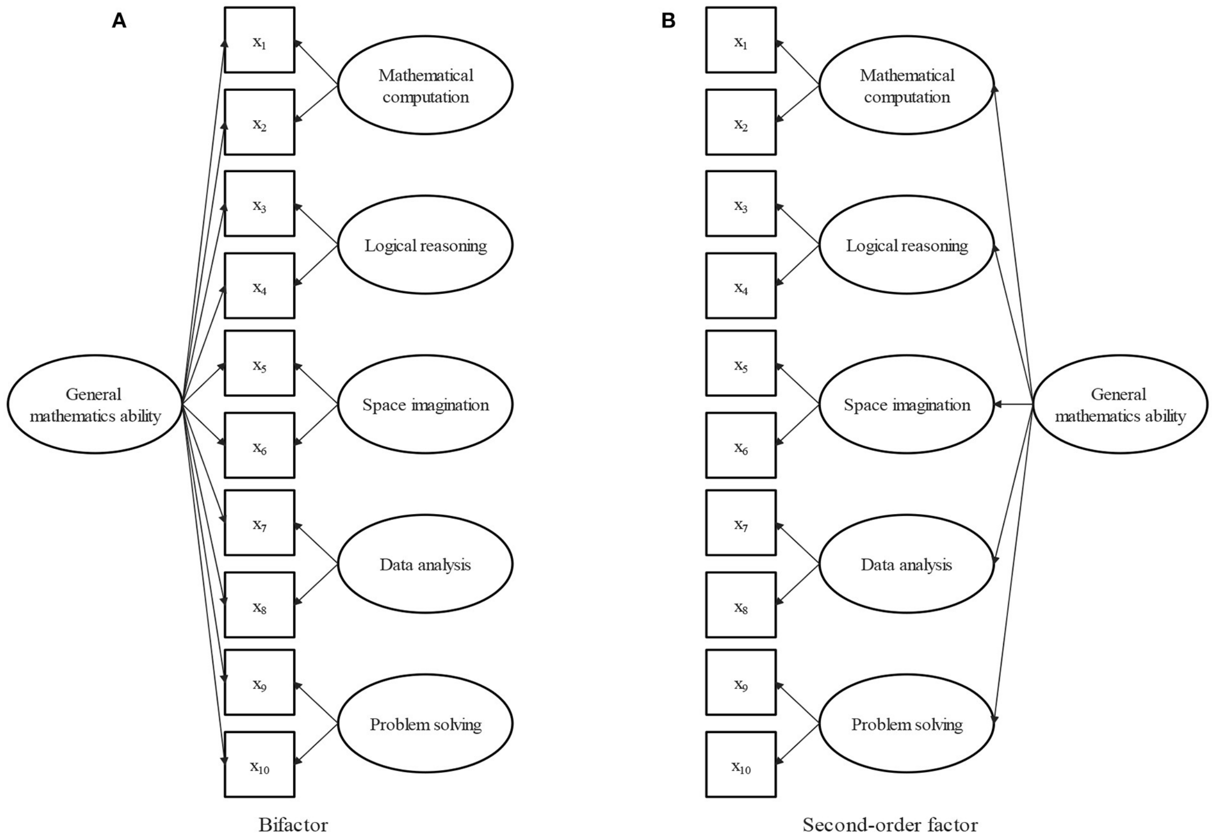


FIGURE 1
(A) Bifactor structure of mathematical ability. **(B)** Second-order factor structure of mathematical ability.

TABLE 1 Four model selection indices (AIC, SABIC, HQ, and BIC) of the three competing models (FIBF, MIRT, and UIRT) for fitting the mathematics test data of the fourth and eighth grades in CCEQM 2015–2017.

		AIC	SABIC	HQ	BIC
Fourth grade	FIBF	**45916**	**46422**	**46342**	**47029**
	MIRT	46199	46586	46525	47088
	UIRT	46166	46528	46471	47068
Eighth grade	FIBF	**29213**	**29640**	**29617**	**30294**
	MIRT	30363	30684	30667	31176
	UIRT	29613	29916	29802	30380

The smallest values of the four model selection indices are bold.

of them are above 0.8, indicating that the five domains of mathematical ability are highly related. The strong correlations among the five domains of mathematical ability once again support the assertion that the bifactor structure is suitable for representing mathematical ability. Based on these results, it is not surprising that the four model selection indices consistently demonstrate that UIRT is superior to between-item MIRT, because a between-item MIRT model with a high related factor structure is close to the UIRT model, and the model evaluation indices prefer simpler models.

3.2.2. Correlation analysis of the factor scores

To investigate the performance of the FIBF model, the relationships between the mathematical ability scores from the FIBF model and those from the between-item MIRT and UIRT models were analyzed. First, the correlation coefficients

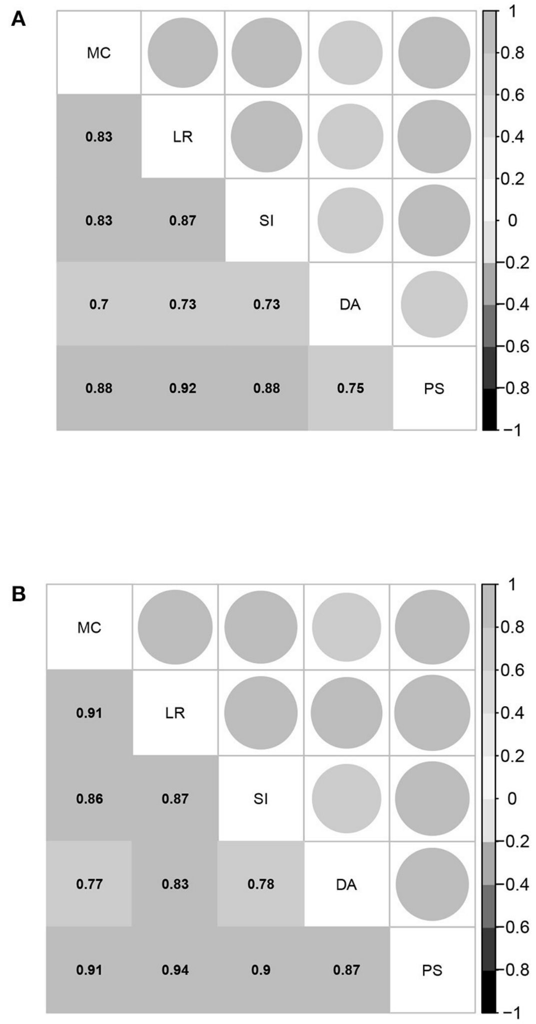


FIGURE 2
Estimated correlation coefficients between the five domain-specific abilities of the between-item MIRT model for students in the fourth and eighth grades. MC, mathematical computation; LR, logical reasoning; SI, space imagination; DA, data analysis; PS, problem solving. **(A)** Fourth grade. **(B)** Eighth grade.

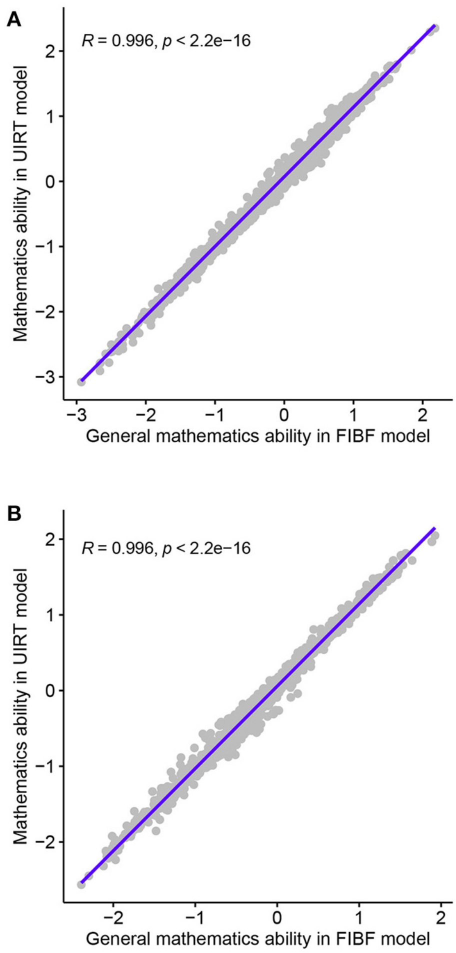


FIGURE 3
Correlation coefficients between the general abilities of the FIBF model and the abilities of the UIRT model for students in the fourth and the eighth grades. **(A)** Fourth grade. **(B)** Eighth grade.

between the general ability of the FIBF and the ability parameter of the UIRT model (denoted as *R*) were computed; the results are shown in Figure 3. Almost all points fall on the diagonal, and the correlation coefficients for both the fourth- and eighth-grade students are $R \cong 0.996$, that is, they are approximately equal to 1.00. This indicates that the general mathematical ability reflected by the FIBF model is nearly equal to the ability suggested by the UIRT model. This phenomenon supports the assertion that general factor of the FIBF model can be interpreted as the general mathematical ability and represents the common element of the five domains of mathematical ability.

Second, the correlation coefficients between the five domain-specific mathematical abilities of the FIBF model and those of the MIRT model were computed; the results are presented in Figure 4. For the fourth-grade students, the correlation coefficients of the same domain-specific mathematical ability from the two models are between 0.3 and 0.6, and exhibit moderate positive correlations; for the eighth-grade students, these correlation coefficients are slightly smaller. However, for both grades, the correlation coefficients between different domain-specific mathematical abilities from the two models are close to 0.0. Based on these results, it can be concluded that the group factors of the FIBF model represent the unique elements of the five specific domains. Additionally, we conducted a correlation analysis of the estimated latent factors in the FIBF model, which reflects whether the latent factors are orthogonal; the results are presented in Figure 5. All of the correlation

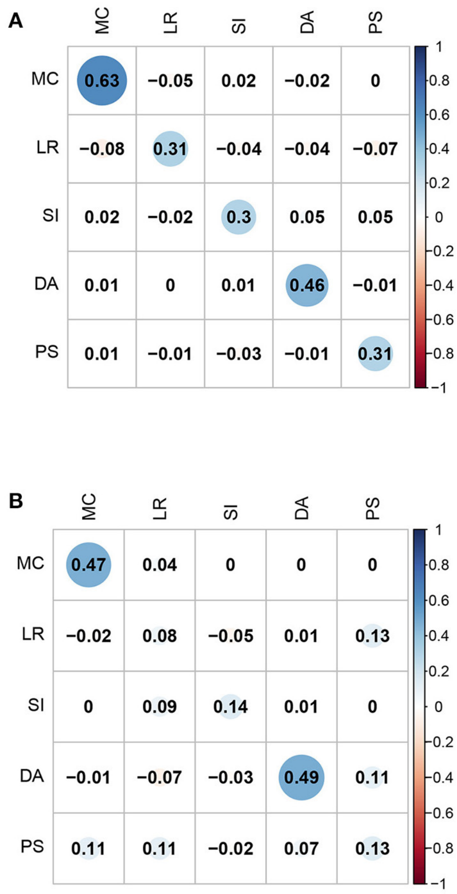


FIGURE 4
Correlation coefficients between the five domain-specific abilities of the FIBF model and those of the between-item MIRT model. MC, mathematical computation; LR, logical reasoning; SI, space imagination; DA, data analysis; PS, problem solving. **(A)** Fourth grade. **(B)** Eighth grade.

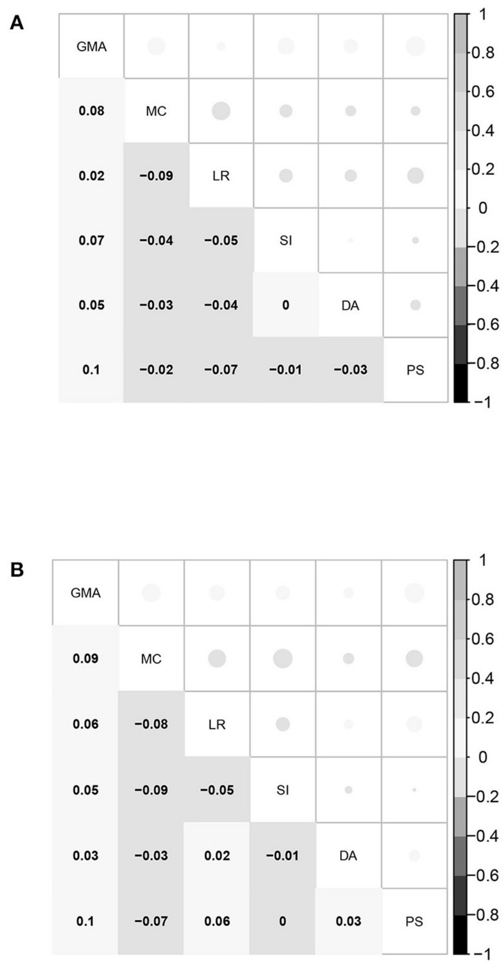


FIGURE 5
Correlation coefficients between the six abilities (the general ability and the five domain-specific abilities) of the FIBF model for students in the fourth and eighth grades. GMA, general mathematical ability; MC, mathematical computation; LR, logical reasoning; SI, space imagination; DA, data analysis; PS, problem solving. **(A)** Fourth grade. **(B)** Eighth grade.

coefficients are very close to 0.0, which strongly indicates that the general and domain-specific latent abilities are independent. In addition, due to the independence of general and specific factors, the relative contribution of each to the overall test performance can be evaluated more readily.

Overall, the results obtained in this empirical study demonstrate that the bifactor model is a powerful approach for representing the intercorrelations among the five domains of mathematical ability. The FIBF model provides a better fit to the empirical data and a cleaner interpretation of mathematical ability than the MIRT and UIRT models.

4. Monte Carlo simulation study

In this section, a Monte Carlo simulation is used to illustrate the performance of the four model selection indices and the parameter estimation methods used in the empirical study.

4.1. Design

As in the mathematics test in CCEQM 2015–2017, there were six booklets in this simulated test, each including 20 items (12 dichotomous and 8 three-category items). Furthermore, each booklet had items in common with two other booklets; for instance, of the 20 items in “Booklet A,” 10 items were the same as those in “Booklet B,” and 10 items were the same as those in “Booklet C.” Thus, there were $M = 60$ items in total. In addition, as for CCEQM 2015–2017, the 60 items were divided into five dimensions. The sample size of the test takers was $N = 2,000$, similar to the sample size of students in the empirical study. Each booklet was answered by ≥ 300 test takers, and each item was included in two booklets. Thus, each item was answered by

TABLE 2 Four model selection indices (AIC, SABIC, HQ, and BIC) of the three competing models (FIBF, MIRT, and UIRT) under the condition that the generating model is FIBF.

	AIC	SABIC	HQ	BIC		AIC	SABIC	HQ	BIC
FIBF	**49737**	**50210**	**50151**	**50858**	FIBF	**49543**	**50016**	**49957**	**50,664**
MIRT	50175	50532	50487	51021	MIRT	49990	50343	50302	50836
UIRT	50652	50986	50944	51443	UIRT	50423	50757	50716	51215
FIBF	**50122**	**50479**	**50434**	**50968**	FIBF	**49773**	**50245**	**50186**	**50894**
MIRT	50652	50986	50944	51443	MIRT	50271	50628	50583	51117
UIRT	50755	51088	51047	51546	UIRT	50755	51088	51047	51546
FIBF	**49750**	**50223**	**50163**	**50871**	FIBF	**49629**	**50102**	**50043**	**50751**
MIRT	50174	50535	50491	51025	MIRT	50110	50467	50423	50957
UIRT	50561	50895	50853	51353	UIRT	50634	50968	50926	51426
FIBF	**49773**	**50246**	**50187**	**50894**	FIBF	**49855**	**50328**	**50269**	**50976**
MIRT	50300	50657	50612	51146	MIRT	50318	50675	50630	51164
UIRT	50749	51083	51041	51540	UIRT	50768	51102	51060	51559
FIBF	**49839**	**50312**	**50253**	**50961**	FIBF	**49920**	**50393**	**50334**	**51041**
MIRT	50222	50579	50534	51068	MIRT	50344	50701	50656	51190
UIRT	50696	51030	50988	51488	UIRT	50765	51099	51057	51557
FIBF	**49863**	**50336**	**50277**	**50984**	FIBF	**49661**	**50134**	**50075**	**50783**
MIRT	50230	50587	50542	51076	MIRT	50181	50538	50493	51027
UIRT	50736	51070	51029	51528	UIRT	50684	51018	50976	51475
FIBF	**49941**	**50414**	**50355**	**51062**	FIBF	**49740**	**50213**	**50154**	**50862**
MIRT	50375	50732	50688	51221	MIRT	50230	50587	50543	51077
UIRT	50836	51170	51128	51627	UIRT	50646	50980	50938	51437
FIBF	**49741**	**50214**	**50155**	**50862**	FIBF	**49947**	**50420**	**50361**	**51068**
MIRT	50172	50529	50485	51019	MIRT	50381	50738	50693	51227
UIRT	50565	50899	50857	51356	UIRT	50834	51168	51127	51626
FIBF	**49516**	**49989**	**49930**	**50638**	FIBF	**49829**	**50302**	**50243**	**50950**
MIRT	50021	50378	50334	50868	MIRT	50335	50692	50647	51181
UIRT	50435	50769	50727	51226	UIRT	50752	51085	51044	51543
FIBF	**49618**	**50091**	**50031**	**50739**	FIBF	**49916**	**50389**	**50329**	**51037**
MIRT	50114	50471	50426	50960	MIRT	50375	50732	50687	51221
UIRT	50525	50859	50818	51317	UIRT	50850	51184	51142	51641

The smallest values of the four model selection indices are bold.

$\geq$ 600 test takers. Overall, the design of this simulation mimics the real situation of CCEQM 2015–2017 as much as possible.

To guarantee that the superior fit of the FIBF model to the empirical data is not due to overfitting, the performance of the four model fit assessment indices (AIC, BIC, SABIC, and HQ) is investigated. In this simulation study, the FIBF, MIRT, and UIRT models were used as the generating models respectively to generate item response data. Let $U(l_1, l_2)$ denote the uniform distribution with a range of $[l_1, l_2]$, and $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2. Let $MVN(\mathbf{\Lambda}, \Sigma)$ denote the multivariate normal distribution with mean vector $\mathbf{\Lambda}$ and covariance matrix Σ. Based on the results of empirical data analysis, the true values of the item parameters of the three models are selected as follows.

a) *FIBF generating model*

The slope parameters a_{0j} and a_{vj} are sampled from

$$a_{0j} \sim U(0.5, 2.5) \tag{6}$$

TABLE 3 Four model selection indices (AIC, SABIC, HQ, and BIC) of the three competing models (FIBF, MIRT, and UIRT) under the condition that the generating model is between-item MIRT.

	AIC	SABIC	HQ	BIC		AIC	SABIC	HQ	BIC
MIRT	**54761**	**55118**	**55073**	**55607**	MIRT	**54904**	**55261**	**55216**	**55750**
FIBF	54868	55341	55282	55989	FIBF	54993	55466	55407	56114
UIRT	55139	55473	55431	55930	UIRT	55378	55712	55670	56109
MIRT	**54676**	**55033**	**54988**	**55522**	MIRT	**54772**	**55129**	**55085**	**55619**
FIBF	54780	55253	55194	55901	FIBF	54936	55409	55350	56057
UIRT	55074	55408	55366	55865	UIRT	55188	55522	55480	55980
MIRT	**54796**	**55153**	**55109**	**55642**	MIRT	**54853**	**55210**	**55165**	**55699**
FIBF	54925	55398	55339	56046	FIBF	54985	55458	55399	56106
UIRT	55085	55419	55377	55876	UIRT	55284	55618	55576	56076
MIRT	**54866**	**55223**	**55178**	**55712**	MIRT	**54638**	**54995**	**54951**	**55485**
FIBF	55007	55480	55421	56129	FIBF	54770	55243	55184	55891
UIRT	55283	55617	55575	56075	UIRT	55084	55417	55376	55875
MIRT	**55003**	**55360**	**55316**	**55849**	MIRT	**54932**	**55289**	**55244**	**55778**
FIBF	55122	55595	55536	56243	FIBF	55043	55516	55456	56164
UIRT	55417	55751	55710	56209	UIRT	55315	55649	55607	56106
MIRT	**54919**	**55276**	**55231**	**55765**	MIRT	**54758**	**55115**	**55071**	**55605**
FIBF	55046	55519	55459	56167	FIBF	54879	55352	55293	56000
UIRT	55286	55619	55578	56077	UIRT	55180	55514	55472	55971
MIRT	**54722**	**55079**	**55034**	**55568**	MIRT	**54637**	**54994**	**54950**	55484
FIBF	54847	55320	55261	55968	FIBF	54739	55212	55152	55860
UIRT	55138	55472	55430	55929	UIRT	54988	55321	55280	55779
MIRT	**54808**	**55165**	**55121**	**55655**	MIRT	**54993**	**55351**	**55306**	**55840**
FIBF	54919	55392	55333	56040	FIBF	55122	55595	55536	56243
UIRT	55079	55412	55371	55870	UIRT	55419	55753	55711	56211
FIBF	**54756**	**55113**	**55069**	**55602**	MIRT	**54721**	**55078**	**55033**	**55567**
MIRT	54869	55342	55283	55990	FIBF	54858	55331	55272	55979
UIRT	55130	55464	55422	55921	UIRT	55156	55490	55448	55947
MIRT	**54824**	**55181**	**55137**	**55671**	MIRT	**54796**	**55153**	**55108**	**55642**
FIBF	54935	55408	55348	56056	FIBF	54906	55379	55320	56027
UIRT	55222	55555	55514	56013	UIRT	55178	55512	55470	55970

The smallest values of the four model selection indices are bold.

and

$$a_{vj} \sim U(0, 1.5) \quad (7)$$

for $v = 1, ..., 5$ and $j = 1, .., M$.

The simulated test is a combination of dichotomous and three-category items, and the intercept parameters of the two types of item response models are different. For dichotomous items, the intercept parameter b_j is sampled from

$$b_j \sim N(0.0, 1.0), \quad (8)$$

while for three-category items, the intercept parameters $\boldsymbol{b}_j = (b_{1j}, b_{2j})$ are randomly sampled from

$$b_{1j} \sim U(-2.0, 0.0), b_{2j} \,|b_{1j} \sim U(b_{1j}, b_{1j} + 2.0) \quad (9)$$

for $j = 1, ..., M$.

The latent abilities of the test takers are orthogonal in the FIBF model, so the true values of $\boldsymbol{\theta}_i$ are randomly generated from

TABLE 4 Four model selection indices (AIC, SABIC, HQ, and BIC) of the three competing models (FIBF, MIRT, and UIRT) under the condition that the generating model is UIRT.

	AIC	SABIC	HQ	BIC		AIC	SABIC	HQ	BIC
UIRT	**45512**	**45845**	**45804**	**46303**	UIRT	**45457**	**45789**	**45747**	**46243**
FIBF	45548	46021	45962	46669	FIBF	45500	45971	45912	46615
MIRT	45633	45990	45945	46479	MIRT	45553	45908	45863	46394
UIRT	**45106**	**45440**	**45398**	**45897**	UIRT	**45406**	**45738**	**45696**	**46192**
FIBF	45150	45623	45564	46271	FIBF	45425	45896	45837	46541
MIRT	45201	45535	45487	46017	MIRT	45507	45862	45817	46348
UIRT	**45397**	**45728**	**45687**	**46182**	UIRT	**45430**	**45764**	**45722**	**46221**
FIBF	45440	45911	45852	46556	FIBF	45478	45951	45892	46599
MIRT	45493	45848	45804	46334	MIRT	45539	45896	45851	46385
UIRT	**45799**	**46133**	**46091**	**46590**	UIRT	**45430**	**45763**	**45722**	**46221**
FIBF	45856	46329	46270	46977	FIBF	45468	45941	45881	46589
MIRT	45915	46272	46228	46761	MIRT	45501	45858	45813	46347
UIRT	**45574**	**45906**	**45864**	**46360**	UIRT	**45511**	**45843**	**45801**	**46297**
FIBF	45622	46093	46034	46738	FIBF	45561	46031	45973	46676
MIRT	45672	46027	45983	46513	MIRT	45613	45968	45923	46454
UIRT	**45030**	**45364**	**45322**	**45821**	UIRT	**45536**	**45870**	**45828**	**46327**
FIBF	45084	45557	45498	46205	FIBF	45572	46045	45986	46693
MIRT	45130	45487	45442	45976	MIRT	45653	46010	45965	46499
UIRT	**45632**	**45989**	**45944**	**46478**	UIRT	**45327**	**45658**	**45617**	**46113**
FIBF	45737	46207	46149	46852	FIBF	45374	45844	45785	46489
MIRT	45737	46068	46027	46523	MIRT	45437	45792	45748	46278
UIRT	**45545**	**45879**	**45837**	**46337**	UIRT	**45710**	**46042**	**46001**	**46496**
FIBF	45592	46065	46006	46713	FIBF	45757	46228	46169	46873
MIRT	45640	45996	45952	46486	MIRT	45805	46159	46115	46646
UIRT	**45200**	**45532**	**45490**	**45986**	UIRT	**45353**	**45684**	**45643**	**46138**
FIBF	45237	45707	45648	46352	FIBF	45385	45855	45797	46500
MIRT	45317	45671	45627	46157	MIRT	45469	45824	45780	46310
UIRT	**44973**	**45307**	**45265**	**45764**	UIRT	**45261**	**45595**	**45553**	**46052**
FIBF	45004	45477	45418	46125	FIBF	45288	45761	45701	46409
MIRT	45084	45441	45396	45930	MIRT	45364	45721	45676	46210

The smallest values of the four model selection indices are bold.

$$\boldsymbol{\theta}_i \sim MVN(\mathbf{0}_{6\times1}, \mathbf{I}_6) \quad (10)$$

for $i = 1, ..., N$; here, $\mathbf{0}_{6\times1}$ is a 6×1 vector in which all elements are 0 and $\mathbf{I}_6$ is the five-dimensional identity matrix.

b) *Between-item MIRT generating model*

The between-item MIRT model can be derived from the FIBF model with the constraint that $a_{0j} = 0$ for all items, that is, only the domain-specific slope parameters $a_{vj}(v = 1, ..., 5)$ need to be generated. The true value of a_{vj} is sampled from,

$$a_{vj} \sim U(0.5, 2.5) \quad (11)$$

for $v = 1, ..., 5$ and $j = 1, ..., M$.

The generation of the true values of the intercept parameters is the same as for the FIBF model, that is, b_j (dichotomous items) and $\boldsymbol{b}_j$ (three-category items)

TABLE 5 ARMSE (ACor) values for the estimation of the item slope and intercept parameters in the three models: FIBF, between-item MIRT, and UIRT.

	General-factor slope	Specific-factor slope					Intercept
	a_0	a_1	a_2	a_3	a_4	a_5	b
FIBF	0.19 (0.95)	0.27 (0.95)	0.29 (0.94)	0.27 (0.97)	0.29 (0.95)	0.28 (0.94)	0.15 (0.98)
MIRT	–	0.24 (0.98)	0.28 (0.97)	0.32 (0.99)	0.25 (0.99)	0.26 (0.99)	0.15 (0.99)
UIRT	0.19 (0.97)	–	–	–	–	–	0.15 (0.99)

are sampled from the distributions in Equations (8) and (9).

The latent ability factors of the between-item MIRT model, $\boldsymbol{\theta}_i = (\theta_{i1}, ..., \theta_{i5})'$, are correlated, and the true values of $\boldsymbol{\theta}_i$ are sampled from

$$\boldsymbol{\theta}_i \sim MVN(\mathbf{0}_{5\times 1}, \Sigma_{\boldsymbol{\theta}}), \tag{12}$$

where $\Sigma_{\boldsymbol{\theta}}$ is a 5×5 covariance matrix, for $i = 1, ..., N$. In this simulation, the main diagonal elements of $\Sigma_{\boldsymbol{\theta}}$ are fixed to 1, and the remaining elements are covariance parameters that are randomly drawn from a uniform distribution over the range $[0.4, 0.8]$.

c) *UIRT generating model*

The UIRT model can be derived from the FIBF model with the constraint that $a_{vj} = 0$ for $v = 1, ..., 5$. There is only one general ability θ_{0i} in the UIRT case. The true values of a_{0j}, b_j, and $\boldsymbol{b}_j$ are drawn from the distributions in Equations (6), (8), and (9) for $j = 1, ..., M$; θ_{0i} is randomly drawn from $N(0, 1)$ for $i = 1, ..., N$.

In this simulation, 20 replications were performed under each simulation condition, and each simulated dataset was fitted by the three models: FIBF, between-item MIRT, and UIRT. All simulations were conducted using the R software, and the four model evaluation indices (AIC, BIC, SABIC, and HQ), as well as the model estimations, were computed based on the "mirt" R package. Note that, if you need the R code, you can contact the authors.

4.2. Results

4.2.1. Behavior of model selection indices

The values of AIC, SABIC, HQ, and BIC for the FIBF, MIRT, and UIRT models are reported in Tables 2–4.

Table 2 presents the results obtained under the condition that the FIBF is the true model. The four model selection indices of the FIBF model are consistently the smallest, which suggests that the FIBF is the best model under this simulation condition. Furthermore, across the 20 replications, the four model evaluation methods consistently suggest that the MIRT model is better than the UIRT model. Because the generating model is FIBF, which is a multidimensional structure, it is correct that the model selection indices support the fit of the MIRT model being better than that of the UIRT model.

Table 3 presents the results obtained under the condition that the between-item MIRT is the true model. All four indices consistently suggest that MIRT, which is the generating model, is the best model. Furthermore, the values of AIC, SABIC, and HQ for the FIBF model are smaller than those for UIRT, whereas the opposite is true for BIC. That is, in comparison with the other model selection methods, the BIC prefers simpler models.

Table 4 presents the results obtained under the condition that the UIRT is the true model. As for the above two simulation conditions, the four model evaluation indices consistently indicate that the generating model is the best model. Except for AIC, the evaluation indices indicate that MIRT is better than FIBF. We believe that MIRT and FIBF should be very close in fitting the unidimensional test data, but as the MIRT model is simpler, it is preferred by the SABIC, HQ, and BIC indices.

Summarizing these results, the four model selection indices (AIC, BIC, SABIC, and HQ) provide excellent accuracy in assessing the three models used in the empirical study. First, they successfully identified the model used to generate the item response data across all simulation conditions. Second, when the FIBF was not the generating model, some model selection indices such as BIC did not support the FIBF model being better than the simpler models. This indicates that overfitting of the bifactor model does not occur for the FIBF model under the incomplete block design.

4.2.2. Recovery of item parameters

The recovery of the three models (FIBF, MIRT, and UIRT) using the "mirt" package was checked in this simulation. The estimation accuracy was assessed by computing the average root mean squire error (ARMSE) of each parameter over all items and the average of the correlation (ACor) between the estimates and the true parameter across the 20 replications.

These metrics were calculated as

$$\text{ARMSE}_{\hat{\delta}} = \frac{1}{M}\sum_{j=1}^{M}\sqrt{20^{-1}\sum_{g=1}^{20}\left(\hat{\delta}_{jg} - \delta_j\right)^2}$$

and

$$\text{ACor}_{\hat{\delta}} = \frac{1}{20}\sum_{g=1}^{20}\frac{\sum_{j=1}^{M}\hat{\delta}_{jg}\delta_j - \frac{1}{M}\sum_{j}^{M}\hat{\delta}_{jg}\sum_{j}^{M}\delta_j}{\sqrt{(\sum_{j=1}^{M}\hat{\delta}_{jg}^2 - \frac{1}{M}(\sum_{j=1}^{M}\hat{\delta}_{jg})^2)(\sum_{j=1}^{M}\delta_j^2 - \frac{1}{M}(\sum_{j=1}^{M}\delta_j)^2)}},$$

where δ_j denotes any one of the parameters of item j and $\hat{\delta}_{jg}$ denotes the corresponding estimate obtained with the g-th simulated data.

The obtained results are presented in Table 5. Use of the "mirt" package allows the item parameters of the three models to be recovered satisfactorily under simulated conditions that are similar to the math tests in CCEQM 2015–2017. First, for all three models, the ARMSE values of the slope parameters do not exceed 0.3, and the corresponding correlations range from 0.94 to 0.99. These results indicate that the slope parameters were well-recovered. Second, the ARMSEs of the intercept parameters of the three models are 0.15, and the values of ACor are >0.98. The recovery for the intercept parameters is excellent, and is slightly better than that for the slope parameters. Furthermore, the estimation accuracy of FIBF is slightly poorer than that of MIRT and UIRT. As stated above, FIBF is the most complex of the three models, so it is normal that the accuracy of its parameter estimation is slightly poorer. Finally, each item was answered by no more than 700 test takers in this simulation. It is likely that the recovery accuracy of the three models will improve as the sample size increases.

5. Conclusion and further issues

The main contribution of this study is to propose a bifactor structure for modeling mathematical ability. On this basis, a mixed FIBF model has been developed for the measurement of mathematical ability in CCEQM 2015–2017. Within the discipline of core literacy theory, mathematical ability is defined as a construct that consists of several domains. These domains differ in their specific concept, but they have common elements, and are thus highly correlated because they all belong to the broader concept of mathematical ability. Bifactor models are a powerful approach for describing such constructs, in which the common element of the five domains is represented by a general factor that is considered as general mathematical ability, and the uniqueness of each domain is represented by a group or domain-specific factor. From the view of bifactor theory, this study has proposed a mixed FIBF model that is a bifactor extension of the mixed IRT model for the measurement of mathematical ability in CCEQM 2015–2017. Furthermore, an important advantage of FIBF analysis is its wide practical adaptability. FIBF models can be applied to various test situations, and numerous related computing tools have been developed. These provide strong support for the application of FIBF to actual large-scale tests. Taken together, not only is the bifactor structure a reasonable approach for representing mathematical ability, but FIBF models are also highly feasible in practice.

The second contribution of this study comes from the empirical study conducted using data from CCEQM 2018 to verify the performance of the FIBF model. The results for the four model selection indices (AIC, BIC, SABIC, and HQ) consistently showed that the fit of the FIBF model is better than that of the UIRT and MIRT models, and the ability scores from the FIBF model had a more reasonable interpretation. The advantages of the FIBF model are fully verified by this empirical study. One important problem with the application of bifactor models is their ease of overfitting. To ensure that no overfitting occurred in the empirical study, a Monte Carlo simulation was constructed to investigate the performance of the four model selection indices under a test design similar to that of CCEQM 2015–2017. The results indicate that the four indices consistently select the generating or true model as the best model across all simulation conditions, and when the FIBF is not the true model, it is not supported. Therefore, the model selection results in the empirical study provide strong evidence that, for fitting the math test data, the FIBF model is superior to the MIRT and UIRT approaches. Furthermore, the simulation results demonstrate that the estimations of the FIBF model have high recovery accuracy under the incomplete test design with mixed item types. Overall, the simulation results show that the existing methods and technologies can support the application of the FIBF model in large-scale testing projects.

There are several considerations for the current study that warrant mention. (1) The estimations of domain-specific factors cannot be directly regarded as scale scores of domain-specific abilities, because they only represent the uniqueness of domain-specific abilities. The scale score of each domain should be a combination of the general and domain-specific factors. Therefore, computing the scale score of domain-specific ability is an important issue that requires further study. (2) Heterogeneity is an important issue in large-scale assessments, such as measurement invariance, heterogeneity of residual variance, and whether the distribution of latent ability is multimodal or skewed. These factors inevitably result in serious unfairness in testing. Thus, to ensure fairness in testing, the development and application of a heterogeneity FIBF model should be further studied. (3) In this study, the bifactor structure of mathematical ability was verified by only one empirical study, and so the obtained conclusion has certain limitations. The bifactor assumption of mathematical ability should be discussed based on more empirical data from large-scale assessment projects. (4) The application of bifactor theory to a measurement model for other disciplines is also worthy of further study.

Data availability statement

The original contributions presented in the study are included in the article, further inquires can be directed to the corresponding authors.

Author contributions

XM contributed to implementing the studies and writing the initial draft. TY and TX contributed to providing the data, key technical support, and manuscript revision. NS contributed to conceptualizing ideas, key technical support and providing a few suggestions on the focus, and direction of the research. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (Grant nos. 11501094 and 11571069) and the National Educational Science Planning Project of China (Grant no. BHA220143).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acton, G. S., and Schroeder, D. H. (2001). Sensory discrimination as related to general intelligence. *Intelligence* 29, 263–271. doi: 10.1016/S0160-2896(01)00066-6

Akaike, H. (1987). Factor analysis and the AIC. *Psychometrika* 52, 317–332. doi: 10.1007/BF02294359

Arens, A. K., Jansen, M., Preckel, F., Schmidt, I., and Brunner, M. (2021). The structure of academic self-concept: a methodological review and empirical illustration of central models. *Rev. Educ. Res.* 91, 34–72. doi: 10.3102/0034654320972186

Beaujean, A. A., Parkin, J., and Parker, S. (2014). Comparing Cattell-Horn-Carroll factor models: differences between bifactor and higher order factor models in predicting language achievement. *Psychol. Assess.* 26, 789–805. doi: 10.1037/a0036745

Bonifay, W., and Cai, L. (2017). On the complexity of item response theory models. *Multivar. Behav. Res.* 52, 465–484. doi: 10.1080/00273171.2017.1309262

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., and Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: appreciating benefits and limitations. *Biol. Psychiatry* 88, 18–27. doi: 10.1016/j.biopsych.2020.01.013

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika* 75, 33–57. doi: 10.1007/s11336-009-9136-x

Cai, L., Yang, J. S., and Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychol. Methods* 16, 221–248. doi: 10.1037/a0023350

Caiado, B., Canavarro, M. C., and Moreira, H. (2022). The bifactor structure of the emotion expression scale for children in a sample of school-aged Portuguese children. *Assessment* 1–15. doi: 10.1177/10731911221082038. [Epub ahead of print].

Chen, F. F., West, S. G., and Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivar. Behav. Res.* 41, 189–225. doi: 10.1207/s15327906mbr4102_5

Cucina, J., and Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *J. Intell.* 5, 1–27. doi: 10.3390/jintelligence5030027

DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *J. Educ. Measure.* 43, 145–168. doi: 10.1111/j.1745-3984.2006.00010.x

Foorman, B. R., Koon, S., Petscher, Y., Mitchell, A., and Truckenmiller, A. (2015). Examining general and specific factors in the dimensionality of oral language and reading in 4th-10th grades. *J. Educ. Psychol.* 107:884. doi: 10.1037/edu0000026

Gault, U. (1954). Factorial patterns of the Wechsler intelligence scales. *Austr. J. Psychol.* 6, 85–89. doi: 10.1080/00049535408256079

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., et al. (2007). Full-information item bifactor analysis of graded response data. *Appl. Psychol. Measure.* 31, 4–19. doi: 10.1177/0146621606289485

Gibbons, R. D., and Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika* 57, 423–436. doi: 10.1007/BF02295430

Gomez, R., and McLaren, S. (2015). The center for epidemiologic studies depression scale: support for a bifactor model with a dominant general factor and a specific factor for positive affect. *Assessment* 22, 351–360. doi: 10.1177/1073191114545357

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., et al. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *J. Abnorm. Psychol.* 128:740. doi: 10.1037/abn0000434

Hannan, E. J., and Quinn, B. G. (1979). The determination of the order of an autoregression. *J. R. Stat. Soc. Ser. B Methodol.* 41, 190–195. doi: 10.1111/j.2517-6161.1979.tb01072.x

Heinrich, M., Zagorscak, P., Eid, M., and Knaevelsrud, C. (2020). Giving G a meaning: an application of the bifactor-(S-1) approach to realize a more symptom-oriented modeling of the Beck depression inventory-II. *Assessment* 27, 1429–1447. doi: 10.1177/1073191118803738

Holzinger, K. J., and Swineford, F. (1937). The bi-factor method. *Psychometrika* 2, 41–54. doi: 10.1007/BF02287965

Immekus, J. C., and Imbrie, P. (2008). Dimensionality assessment using the full-information item bifactor analysis for graded response data: an illustration with the state metacognitive inventory. *Educ. Psychol. Measure*. 68, 695–709. doi: 10.1177/0013164407313366

Jiang, Y., Zhang, J., and Xin, T. (2019). Toward education quality improvement in China: a brief overview of the national assessment of education quality. *J. Educ. Behav. Stat*. 44, 733–751. doi: 10.3102/1076998618809677

Jorge-Botana, G., Olmos, R., and Luzón, J. M. (2019). Could LSA become a "bifactor" model? Towards a model with general and group factors. *Expert Syst. Appl*. 131, 71–80. doi: 10.1016/j.eswa.2019.04.055

Kim, H., and Eaton, N. R. (2015). The hierarchical structure of common mental disorders: connecting multiple levels of comorbidity, bifactor models, and predictive validity. *J. Abnorm. Psychol*. 124:1064. doi: 10.1037/abn0000113

Kim, K. Y., and Cho, U. H. (2020). Approximating bifactor IRT true-score equating with a projective item response model. *Appl. Psychol. Measure*. 44, 215–218. doi: 10.1177/0146621619885903

Leue, A., and Beauducel, A. (2011). The PANAS structure revisited: on the validity of a bifactor model in community and forensic samples. *Psychol. Assess*. 23:215. doi: 10.1037/a0021400

Li, Y., and Lissitz, R. W. (2012). Exploring the full-information bifactor model in vertical scaling with construct shift. *Appl. Psychol. Measure*. 36, 3–20. doi: 10.1177/0146621611432864

Liu, Y., and Thissen, D. (2012). Identifying local dependence with a score test statistic based on the bifactor logistic model. *Appl. Psychol. Measure*. 36, 670–688. doi: 10.1177/0146621612458174

Martel, M. M., Roberts, B., Gremillion, M., Von Eye, A., and Nigg, J. T. (2011). External validation of bifactor model of ADHD: explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. *J. Abnorm. Child Psychol*. 39, 1111–1123. doi: 10.1007/s10802-011-9538-y

McAbee, S. T., Oswald, F. L., and Connelly, B. S. (2014). Bifactor models of personality and college student performance: a broad versus narrow view. *Eur. J. Pers*. 28, 604–619. doi: 10.1002/per.1975

McFarland, D. J. (2013). Modeling individual subtests of the WAIS IV with multiple latent factors. *PLoS ONE* 8:e74980. doi: 10.1371/journal.pone.0074980

McFarland, D. J. (2016). Modeling general and specific abilities: evaluation of bifactor models for the WJ-III. *Assessment* 23, 698–706. doi: 10.1177/1073191115595070

Monteiro, F., Fonseca, A., Pereira, M., and Canavarro, M. C. (2021). Measuring positive mental health in the postpartum period: the bifactor structure of the mental health continuum-short form in Portuguese women. *Assessment* 28, 1434–1444. doi: 10.1177/1073191120910247

Morgan, G. B., Hodge, K. J., Wells, K. E., and Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: a comparison of fit in correlated factors, higher-order, and bi-factor models *via* Monte Carlo simulations. *J. Intell*. 3, 2–20. doi: 10.3390/jintelligence3010002

Moshagen, M., Hilbig, B. E., and Zettler, I. (2018). The dark core of personality. *Psychol. Rev*. 125:656. doi: 10.1037/rev0000111

Murray, A. L., and Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence* 41, 407–422. doi: 10.1016/j.intell.2013.06.004

OECD (2014). *PISA 2012 Technical Report*. OECD.

Olatunji, B. O., Ebesutani, C., and Abramowitz, J. S. (2017). Examination of a bifactor model of obsessive-compulsive symptom dimensions. *Assessment* 24, 45–59. doi: 10.1177/1073191115601207

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivar. Behav. Res*. 47, 667–696. doi: 10.1080/00273171.2012.715555

Reise, S. P., Morizot, J., and Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual. Life Res*. 16, 19–31. doi: 10.1007/s11136-007-9183-7

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychol. Methods* 21:137. doi: 10.1037/met0000045

Rushton, J. P., and Irwing, P. (2009a). A general factor of personality (GFP) from the multidimensional personality questionnaire. *Pers. Individ. Differ*. 47, 571–576. doi: 10.1016/j.paid.2009.05.011

Rushton, J. P., and Irwing, P. (2009b). A general factor of personality in 16 sets of the Big Five, the Guilford-Zimmerman Temperament Survey, the California Psychological Inventory, and the Temperament and Character Inventory. *Pers. Individ. Differ*. 47, 558–564. doi: 10.1016/j.paid.2009.05.009

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat*. 6, 461–464. doi: 10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika* 52, 333–343. doi: 10.1007/BF02294360

Sellbom, M., and Tellegen, A. (2019). Factor analysis in psychological assessment research: common pitfalls and recommendations. *Psychol. Assess*. 31:1428. doi: 10.1037/pas0000623

Shevlin, M., McElroy, E., Bentall, R. P., Reininghaus, U., and Murphy, J. (2016). The psychosis continuum: testing a bifactor model of psychosis in a general population sample. *Schizophr. Bull*. 43, 133–141. doi: 10.1093/schbul/sbw067

Simms, L. J., Grös, D. F., Watson, D., and O'Hara, M. W. (2008). Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depress. Anxiety* 25, E34–E46. doi: 10.1002/da.20432

Snyder, H. R., Hankin, B. L., Sandman, C. A., Head, K., and Davis, E. P. (2017). Distinct patterns of reduced prefrontal and limbic gray matter volume in childhood general and internalizing psychopathology. *Clin. Psychol. Sci*. 5, 1001–1013. doi: 10.1177/2167702617714563

Spearman, C. (1904). General ability, objectively determined and measured. *Am. J. Psychol*. 15:201. doi: 10.2307/1412107

The National Assessment Center for Education Quality (2018). *The China Compulsory Education Quality Oversight Report [in Chinese*]. Available online at: http://www.moe.gov.cn/jyb_xwfb/moe_1946/fj_2018/201807/P020180724685827455405.pdf

Valerius, S., and Sparfeldt, J. R. (2014). Consistent g- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory factor analyses using three test batteries. *Intelligence* 44, 120–133. doi: 10.1016/j.intell.2014.04.003

Waldman, I. D., Poore, H. E., Luningham, J. M., and Yang, J. (2020). Testing structural models of psychopathology at the genomic level. *World Psychiatry* 19, 350–359. doi: 10.1002/wps.20772

Watkins, M. W. (2010). Structure of the Wechsler intelligence scale for children-Fourth edition among a national sample of referred students. *Psychol. Assess*. 22:782. doi: 10.1037/a0020043

Watkins, M. W., and Beaujean, A. A. (2014). Bifactor structure of the Wechsler preschool and primary scale of intelligence-fourth edition. *School Psychol. Q*. 29:52. doi: 10.1037/spq0000038

Yin, D. (2021). Education quality assessment in China: what we learned from official reports released in 2018 and 2019. *ECNU Rev. Educ*. 4, 396–409. doi: 10.1177/2096531120944522

Zhan, P., Yu, Z., Li, F., and Wang, L. (2019). Using a multi-order cognitive diagnosis model to assess scientific literacy. *Acta Psychol. Sin*. 51:734. doi: 10.3724/SP.J.1041.2019.00734

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 06 January 2023
DOI 10.3389/fpsyg.2022.927757

OPEN ACCESS

EDITED BY
Shuhua An,
California State University, Long Beach, United States

REVIEWED BY
Nurul Hafizah Azizan,
Universiti Teknologi Mara, Malaysia
Kazuo Mori,
Tokyo University of Agriculture and Technology, Japan
Antonio Luque,
University of Almería, Spain

*CORRESPONDENCE
Yassin Mohammed Yesuf
✉ yasspsycho@gmail.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 24 April 2022
ACCEPTED 05 December 2022
PUBLISHED 06 January 2023

CITATION
Yesuf YM, Kebede SA, Zewdu A and Gebru DM (2023) Predictors of high school students' mathematics self-efficacy in Addis Ababa: The importance of educational expectations. *Front. Psychol.* 13:927757. doi: 10.3389/fpsyg.2022.927757

Predictors of high school students' mathematics self-efficacy in Addis Ababa: The importance of educational expectations

Yassin Mohammed Yesuf[1]*, Sebsibew Atikaw Kebede[2], Atinkut Zewdu[3] and Dawit Mekonnen Gebru[4]

[1]Department of Psychology, College of Social Sciences and Humanities, University of Gondar, Gondar, Ethiopia, [2]Department of Mathematics, College of Natural and Computational Sciences, University of Gondar, Gondar, Ethiopia, [3]Department of Psychology, Institute of Education and Behavioral Sciences, Ambo University, Ambo, Ethiopia, [4]School of Psychology, College of Education and Behavioral Studies, Addis Ababa University, Addis Ababa, Ethiopia

In Ethiopia studies on high school students' mathematics self-efficacy and associated factors are scarce. The present study examined students' mathematics self-efficacy and associated predictors among high school students in Addis Ababa. Data were collected using adapted questionnaire from 120 students (9th and 10th graders) recruited *via* multi-stage sampling. Descriptive statistics, independent sample *t*-test, ANOVA, Chi-square and logistic regressions were utilized to analyze the collected data. In the study it was found that students have more than average mathematics self-efficacy even though significant numbers of students (44.2%) have low mathematics self-efficacy. It was also revealed that differences in grade level [$t(118) = 2.545$, $p < 0.05$] and students' expected grade in the upcoming national exam [$F(3,116) = 5.553$, $p < 0.05$] were statistically significant. Living arrangements (AOR = 6.704, 95% CI = 1.598–28.118), expected grade in the upcoming national exam (AOR = 5.808, 95% CI = 1.804–18.696) and expected marks in the semester (AOR = 1.126, 95% CI = 1.055–1.202) are significant predictors of students' mathematics self-efficacy. Generally educational expectations are important predictors of students' mathematics self-efficacy. Therefore, researchers and organizations need to gear their attention towards improving students' mathematics self-efficacy.

KEYWORDS

mathematics self-efficacy, educational expectations, high school students, Addis Ababa, Ethiopia

10.3389/fpsyg.2022.927757

1. Introduction

The social cognitive theory developed by Albert Bandura is the most prominent learning theory, and self-efficacy is an important component of the theory (Liu and Koirala, 2009). As described by Bandura, self-efficacy is "beliefs in one's capability to organize and execute the courses of action required to manage prospective situations" (Bandura, 1997). Bandura also described self-efficacy as beliefs about one's capabilities to learn or perform behaviors at designated levels (Reynolds and Miller, 2003). Heslin and Klehe (2006) argued that self-efficacy is "one of the most powerful motivational predictors of how well a person will perform at almost any endeavor". A person's self-efficacy is a strong determinant of their effort, persistence, and strategizing, as well as their subsequent training and job performance (Heslin and Klehe, 2006).

Self-efficacy influences the choices people make and the courses of action they pursue. Most people engage in tasks in which they feel competent and confident and avoid those in which they do not. Beliefs in personal competence also help to determine how much effort people will devote to an activity, how long they will persevere when confronting obstacles, and how resilient they will prove in the face of adverse situations; the higher the sense of efficacy, the greater the effort, persistence, and resilience (Bandura, 1997). Self-efficacy influences the choice we make, the effort we put forth, and how long we persist (Tait–McCutcheon, 2008).

Generally speaking, self-efficacy influences not only an individual's performance but also the choices they make. The influence of self-efficacy on an individual's performance and choices is also applicable to students' performance in school settings. Self-efficacy in an academic setting includes students' confidence in their cognitive skills to perform the academic task and influences their choice of tasks, persistence, effort, and achievement (Reynolds and Miller, 2003).

Prior national and international studies consistently found that students with high self-efficacy outperform students with low self-efficacy (Huang, 2013; Shaine, 2015; Tizazu and Ambaye, 2017). Moreover, studies depict that self-efficacy determines students' career decision-making (Ogutu et al., 2017; Akter et al., 2018).

At present, the focus of both educators and policymakers has shifted to science education since a nation's success is dependent on scientific innovations and advances in technology (Kahveci, 2010). The Ethiopian Ministry of Education, as ratified in Education Sector Development Program V (2015/16–2019/20) has given high priority to science education (Ministry of Education, 2015).

For this to be a reality, students need to join the science and technology fields. But the question here is, "what are the factors that determine students' future career choices?" Studies conducted to understand what really determines students' future career choices in science, technology, engineering, and mathematics (STEM) fields have revealed that "high school mathematics achievement, exposure to mathematics and science courses, and mathematics self-efficacy beliefs all affect students' intent to major in STEM fields, which in turn influences entrance into STEM majors" (Wang, 2013). Mathematics knowledge and skills are requirements not only for tertiary education but also for further studies (Schulz, 2005). Mathematics is a prerequisite for pursuing higher education in most professions, including all exact sciences, finance, programming, and so on. It allows students to choose from a wide range of vocations with high prospects of academic acceptance, mostly in engineering, the natural sciences, and technology, as well as a significant portion of the social sciences (Davidovitch and Yavich, 2018). This calls for a closer look at teaching and learning of mathematics.

In this regard, mathematics self-efficacy, defined as "a situational assessment of an individual's confidence in her or his ability to successfully perform or accomplish a particular mathematical task or problem", has become a prominent construct for research (Kiamanesh et al., 2004).

Mathematics self-efficacy (MSE) is one of the crucial factors in students' mathematics learning (Roslan and Maat, 2019). It is a stronger predictor of math performance than math anxiety (Pajares and Miller, 1994), previous math experience (Pajares and Kranzler, 1995b), mathematics self-concept, perceived usefulness of mathematics, prior experience with mathematics, or gender (Pajares and Miller, 1994) and influences math performance as strongly as overall mental ability (Pajares and Kranzler, 1995a).

In practice, the roles of students' mathematics self-efficacy on their achievements have been reported in studies conducted at all education levels. For example, in a study among 5th and 6th graders in Spain, students' mathematics self-efficacy significantly predicted their mathematics achievement (Rodríguez et al., 2020). Mathematics self-efficacy significantly predicted mathematics achievement in 7th graders in Turkey (Recber et al., 2018). Mathematics self-efficacy significantly predicted mathematics achievement in 10th graders in Bhutan (Norbu and Dukpa, 2021).

In a systematic review of studies, it was found that MSE is an important predictor of high school students' mathematics achievement (Roslan and Maat, 2019). In two studies in Greece among students from grades 7–11 mathematics self-efficacy significantly predicted mathematics achievement using data from the Program for International Student Assessment (PISA) (Cheema, 2018; Hiller et al., 2021). In a study that included sample students from the United States and China aged 15 years, using data from PISA, mathematics self-efficacy significantly predicted mathematics achievement (Wu, 2016). In a study among Bahirdar University students, MSE was a significant predictor of mathematics performance (Getahun et al., 2016).

Alongside these findings, students' mathematics self-efficacy is found to be a strong predictor of their choice of mathematics-related courses and majors (Zarch and Kadivar, 2006). Besides, studies among college students consistently revealed that MSE is an important predictor of both their performance and major choice (Lent et al., 2008; Lin et al., 2018).

Hence, assessing students' mathematics self-efficacy would inform us not only about students' future performance but also about their future career choice. Such assessments need to be conducted at the high school level. This is particularly important in Ethiopia since students are given the opportunity to choose between social and natural science streams after the completion of their high school years (grades 9 and 10). The present study, therefore, aimed to examine high school students' mathematics self-efficacy in Addis Ababa.

Cognizant of examining students' mathematics self-efficacy, there is a need to examine the demographic variability among students. Theoretically, the social cognitive theory argues that demographic variability shapes people's self-efficacy beliefs (Lin et al., 2018). In practice, the effects of demographic variables on students' mathematics self-efficacy are still inconclusive. For instance, gender was found to be associated with students' mathematics self-efficacy, favoring males (Lloyd et al., 2005; Wu, 2016; Recber et al., 2018; Rodríguez et al., 2020; Zander et al., 2020). In a meta-analytic review of studies, males were found to have a higher MSE than their female counterparts (Huang, 2013). On the contrary, no gender differences in MSE were observed in other studies (Ayotola and Adedeji, 2009; Clutts, 2010; Turgut, 2013; Davidovitch and Yavich, 2018; Probst, 2019).

School type is another factor associated with students' mathematics self-efficacy with variable results. For example, in a study conducted by Zedan and Bitar (2014) in Israel, it was found that the dimensions of the classroom environment explain 50% of the variations in high school students' mathematics self-efficacy. Likewise, in a study among high school students in Greece using data from PISA, students from private schools tend to have higher mathematics self-efficacy than students from public schools (Cheema, 2018). In contrast, in a study in Turkey, there is no difference in MSE based on school type (Recber et al., 2018).

These and other research findings include extensive examinations of variables associated with students' mathematics self-efficacy. In Ethiopia, studies that assess the mathematics self-efficacy of high school students are scarce, and the ones that are available (mainly student works) come up with contradictory findings. For example, the study by Abebe (2001) found gender differences favoring males, while the studies by Wubalem (2006) and Ayele and Dadi (2016) found no gender differences in students' mathematics self-efficacy. Contrary to these and many other findings described earlier, a study in Tigray, Ethiopia depicted that female grade 9 students have higher mathematics self-efficacy than their male counterparts (Tekola et al., 2020).

The present study, therefore, tried to examine students' mathematics self-efficacy and associated predictors among high school students in Addis Ababa. Based on the literature alluded to, we first hypothesized that the students would have a medium level of MSE. It is also hypothesized that there will be variations in students' MSE based on their demographic characteristics. Besides, students' demographic characteristics will be important predictors of their MSE. In a country that aspires for attracting a huge amount of students to science and technology fields, a closer inquiry into students' mathematics self-efficacy would give them the opportunity to make informed decisions. Moreover, the findings of the present study will add new knowledge from a resource poor setting to the inconclusive global findings on the predictive roles of students' demographic characteristics on their MSE.

2. Materials and methods

2.1. Research design

The purpose of the present study was to examine high school students' mathematics self-efficacy and associated factors in Addis Ababa, Ethiopia. For this purpose to be achieved, a quantitative, descriptive, cross-sectional, and explanatory study design was used. The present study is quantitative in terms of the type of data collected. With regard to the timing of the data collection, it is cross-sectional. The study is both descriptive and explanatory in terms of the statistical computations employed to analyze the data.

2.2. Sampling technique and sample characteristics

Participants in the study are 120 students selected from two schools in Addis Ababa, Ethiopia. Based on the findings from earlier studies, gender (Rodríguez et al., 2020; Zander et al., 2020), school type (Özgen and Bindak, 2011), and grade level (Wubalem, 2006; Özgen and Bindak, 2011) are important variables in students' mathematics self-efficacy. As such, an equal number of students from both genders, from both government and private high schools, and from both grades 9 and 10 (60 from each group) were included in the study.

Multistage sampling was used to recruit equal numbers of students from government and private schools. In doing so, first one private and one government school were randomly selected. At the school level, one class from grade 9 and another class from grade 10 were selected again using random sampling. Following that, 15 male and 15 female students were randomly selected from a class (i.e., a total of 30 students were selected from a class). Needless to say, an equal number of students from both genders (60 male and 60 female) and grade levels (60 9th graders and 60 10th graders) were included in the study.

Additional characteristics of the respondents were collected, and summaries of these characteristics are presented in **Table 1**.

As shown in **Table 1**, the mean age of the respondents is 16.12 (SD = 1.063), where the minimum and maximum ages are 14 and 19, respectively. The mean expected mark at the end of the semester is 82.02 (SD = 11.514), where 45 and 100 are the minimum and maximum expected marks, respectively. Of all the respondents, 97 (80.8%) of the respondents live with both parents. With regard to the respondents' expected grade in the upcoming national exam, 58 (48.3%) of the respondents expect B grade, 52 (43.3%) of them expect A grade, 6 (5.0%) of them expect F grade, and 4 of them (3.3%) expect C grade. Of all the respondents, 45.8% of the students received tutorials weekly, 30.0% of them (36 in number) never received tutorials, 20.0% of them (24 in number) received tutorials daily, and the remaining 4.2% of them (5 in number) seldom received tutorials. In terms of their plan to join a college/university, 93 (77.5%) of the respondents have a definite plan, 15 (12.5%) of them have a tentative plan, and the remaining 12 (10.0%) of them have no plan to attend a college/university in the future. **Table 1** also depicts that 63.3% of the students aspire to join STEM, while 5.8% of them aspire to join social science professions. The remaining 30.8% of the respondents planned to join other professions.

TABLE 1 Demographic characteristics of the respondents.

Variables	Category	Mean (SD) or *N* (%)	Min (max)
Age, mean (SD)		16.12 (1.063)	14 (19)
Expected mark, mean (SD)		82.02 (11.514)	45 (100)
Living with, number (%)	One parent	23 (19.2%)	
	Both parents	97 (80.8%)	
Expected grade	A	52 (43.3%)	
	B	58 (48.3%)	
	C	4 (3.3%)	
	F	6 (5.0%)	
Receiving tutorial	Daily	24 (20.0%)	
	Weekly	55 (45.8%)	
	Seldom	5 (4.2%)	
	Never	36 (30.0%)	
Plan to attend college/university	Definitely	93 (77.5%)	
	Maybe	15 (12.5%)	
	No	12 (10.0%)	
Profession	Social science	7 (5.8%)	
	STEM	76 (63.3%)	
	Other	37 (30.8%)	

2.3. Instrument

A questionnaire was used to assess students' mathematics self-efficacy. Ultimately, the questionnaire has two sections, where the first section collects data on students' demographic characteristics. This includes school type, age, gender, grade level, living conditions with parents, plans to attend college/university, expected grade in the upcoming national examination, marks expected in the semester, receiving tutorial, and professions students aspire to join. The second section of the questionnaire is adapted from an instrument developed by Pajares (1996) and later modified by Johnson (2008) to be used for assessing high school students' mathematics self-efficacy. The original tool has 39 items. Before collecting the final data, the adapted tool was translated into Amharic, and a pilot study was conducted on 30 students. In the pilot study, the reliability of the tool was found to be 0.919. The deletion of an item increased the reliability of the tool to 0.924. Therefore, the item "I have never been very excited about mathematics" was deleted from the tool. The final data was thus collected with 38 items measuring students' mathematics self-efficacy. The replies for the items are based on a five-point Likert scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Undecided, 4 = Agree, and 5 = Strongly Agree. During analysis, 13 items (Pajares and Miller, 1994; Pajares and Kranzler, 1995a; Wubalem, 2006; Zarch and Kadivar, 2006; Lent et al., 2008; Clutts, 2010; Zedan and Bitar, 2014; Ministry of Education, 2015; Wu, 2016; Cheema, 2018; Lin et al., 2018; Probst, 2019; Norbu and Dukpa, 2021) were reverse coded. The highest score, then, represents high mathematics self-efficacy.

2.4. Procedures

Permission to conduct the study was secured from Addis Ababa University, School of Psychology's Ethical Review Committee. Then a formal letter was written from the school directed to the schools, requesting cooperation with the researchers. Then the researchers explained the scope and purpose of the study to the school directors, thereby securing their permission. At the individual level, purpose and scope of the study were communicated with participants, and they were assured that all the information that they would give would be kept confidential. Data are collected after each participant signs the consent form. Participation in this study was totally voluntary, and no compensation was offered.

2.5. Methods of data analysis

Descriptive statistics, including percentage, number of cases, mean, scores below, and above the mean and standard deviation, were utilized to describe students' mathematics self-efficacy and demographic variables collected from the students. In addition, an independent sample *t*-test, a one-way ANOVA, and a chi-square test of significance were utilized to check for differences in mathematics self-efficacy based on associated variables. Simple and multiple binary logistic regressions were computed to assess the effects of the predictor variables over the criterion variable (mathematics self-efficacy). All data were analyzed using the Statistical Package for Social Sciences (SPSS), Windows version 23.

3. Results

3.1. Students' mathematics self-efficacy

To describe students' mathematics self-efficacy, descriptive statistical tools including mean, standard deviation, and minimum and maximum scores were computed. As shown in Table 2, the mean score of the respondents' mathematics self-efficacy is 129.9 (SD = 22.893), where the minimum and maximum values are 52 and 170, respectively. On a scale of five, the expected mean from the 38-item tool is 114 (3 × 38). Therefore, the mean score found here implies that students have higher than average mathematics self-efficacy. In addition, the standard deviation found here tells us that there is a high dispersion among students' mathematics self-efficacy.

Furthermore, students were categorized as having high or low mathematics self-efficacy using the mean score as a cutoff point. Table 2 also shows that 67 (55.8%) of the respondents scored above the mean value, whereas 53 (44.2%) of them scored below the mean value. This informs us that more than half of the respondents have high mathematics self-efficacy.

3.2. Predictors of students' mathematics self-efficacy

Exploring factors associated with students' mathematics self-efficacy is one of the main objectives of the present study. To achieve this objective, first of all, mean differences in students' mathematics self-efficacy based on demographic variables were examined using independent sample *t*-tests and ANOVA tests.

TABLE 2 Description of students' mathematics self-efficacy.

Variable	*M* (SD)	>mean		<mean		Min (max)
		N	%	*N*	%	
Mathematics self-efficacy	129.95 (22.893)	67	55.8	53	44.2	52 (170)

TABLE 3 Mean difference in students' mathematics self-efficacy based on school type, gender, grade level, and living arrangements.

Variables	Category	Mean	SD	*t*-values
School type	Government	127.75	19.198	−1.053
	Private	132.15	26.051	
Gender	Male	130.32	24.696	0.175
	Female	129.58	21.139	
Grade level	9	135.15	15.428	2.545*
	10	124.75	27.641	
Living with	Single parent	133.91	14.529	0.923
	Both parents	129.01	24.426	

*Significant at the 0.05 level.

Independent sample *t*-tests were used to look into the mean difference in students' mathematics self-efficacy based on school type, gender, grade level, and living conditions with parents, and the results of the analysis are summarized in Table 3.

Table 3 shows us that the mean mathematics self-efficacy of respondents from government schools (*M* = 127.75, SD = 19.198) is lesser than their counterparts from private schools (*M* = 132.15, SD = 26.051), but the mean difference is not statistically significant ($t = -1.053$, $p = 0.294$).

Moreover, Table 3 also shows us that the mean mathematics self-efficacy of male respondents (*M* = 130.32, SD = 24.696) is higher than female respondents (*M* = 129.58, SD = 21.139) but not statistically significant ($t = 0.175$, $p = 0.862$). The table also informs us that the mean mathematics self-efficacy of students who are living with single parents (*M* = 133.91, SD = 14.529) is higher than the mean mathematics self-efficacy of students who are living with both parents (*M* = 129.01, SD = 24.426), although the mean difference is not statistically significant ($t = 0.923$, $p = 0.358$).

On the contrary, the mean mathematics self-efficacy of 9th graders (*M* = 135.15, SD = 15.428) is higher than 10th graders (*M* = 124.75 SD = 27.641), and the mean difference is statistically significant ($t = 2.545$, $p < 0.05$).

One-way ANOVA was employed to look into the mean difference in students' mathematics self-efficacy based on their expected grades in the upcoming national examination, the number of tutorials they received, their plan to join college/university, and the profession they aspire to join. The summaries of the findings are presented in Table 4.

As shown in Table 4, students' expected grades in the upcoming national examination have a statistically significant effect [$F(3,116) = 5.553$, $p < 0.05$] on students' mathematics self-efficacy.

Tukey's HSD *post hoc* test for differences in mathematics self-efficacy indicates that there are statistically significant

TABLE 4 One-way ANOVA.

Variables	Category	Mean	SD	*F*-values
Plan to join college/university	Definitely	132.63	23.893	2.955
	May be	119.93	15.026	
	No	121.67	18.307	
Receive tutorial	Daily	134.08	35.716	0.734
	Weekly	130.69	18.128	
	Seldom	120.00	27.414	
	Never	127.44	17.819	
Expected grade	A	125.21	25.657	5.553*
	B	136.93	17.683	
	C	100.00	1.155	
	F	123.50	25.034	
Profession	Social science	122.43	7.569	1.149
	STEM	132.25	25.748	
	Other	126.65	17.558	

*Significant at the 0.05 level.

differences between students who expect an A in the upcoming national examination and students who expect a B ($p < 0.05$) and between students who expect a B and students who expect a C in the upcoming national examination ($p < 0.01$).

Specifically, students who expect a B in the upcoming national examination (M = 136.93, SD = 17.683) have higher mathematics self-efficacy than students who expect an A (M = 125.21, SD = 25.657). Likewise, students who expect B (M = 136.93, SD = 17.683) have higher mathematics self-efficacy than students who expect C in the upcoming national examination (M = 100.00, SD = 1.155).

On the contrary, the amount of tutorial received [$F(3,116) = 0.734$, $p = 0.534$], students' plan to join college/university [$F(2,117) = 2.955$, $p = 0.056$], and students' professional aspiration [$F(2,117) = 1.149$, $p = 0.321$] do not have a statistically significant effect on students' mathematics self-efficacy.

In addition to all these computations, chi-square tests were used to check the proportion of respondents with high and low mathematics self-efficacy in each category of the demographic variables. The results of the computations are presented in Table 5.

Table 5 informs us that the number of 9th graders who score above the mean (58.2%) is higher than the number of 10th graders who score above the mean (41.8%). Table 5 also tells us that the number of 9th graders who score below the mean (39.6%) is lesser than the number of 10th graders who score below the mean (60.4%), and these differences in proportions are statistically significant ($\chi^2 = 4.089$, $df = 1$, $p < 0.05$).

Besides, the table informs us that the differences in the proportion of students' mathematics self-efficacy based on their living arrangements are statistically significant ($\chi^2 = 8.272$, $df = 1$, $p < 0.01$). As such, a higher number of students with single parents scored above the mean (82.6%), while a higher number of students who are living with both parents scored below the mean (50.5%).

Finally, Table 5 reveals that differences in mathematics self-efficacy based on students' expected grades in the upcoming national examination are statistically significant ($\chi^2 = 8.738$, $df = 2$, $p < 0.05$). Of the students who expect a B in the upcoming national exam, the majority of them scored above the mean (69.0%). Among students who expect an A or other grades, the majority of them scored below the mean (53.8 and 70%, respectively).

For the sake of exploring the factors associated with students' mathematics self-efficacy, multiple regression was computed. The result of the computation revealed that the variables added explained 13.6% of the variations in students' mathematics self-efficacy scores, but the model is statistically insignificant [$F(10, 109) = 1.722$, $p = 0.0.085$, $R^2 = 0.136$].

TABLE 5 Proportions of students with high and low mathematics self-efficacy.

Variables	Category	<mean		>mean		χ^2
		N	%	*N*	%	
Gender	Male	24	40.0	36	60.0	0.845
	Female	29	48.3	31	51.7	
Grade level	9	21	35.0	39	65.0	4.089*
	10	32	53.3	28	46.7	
School type	Government	29	48.3	31	51.7	0.845
	Private	24	40.0	36	60.0	
Living with	Single parent	4	17.4	19	82.6	8.272**
	Both parents	49	50.5	48	49.5	
Receive tutorial	Daily	7	29.2	17	70.8	4.967
	Weekly	25	45.5	30	54.5	
	Seldom	4	80.0	1	20.0	
	Never	17	47.2	19	52.8	
Plan to join college/university	Definitely	36	38.7	57	61.3	5.179
	May be	10	66.7	5	33.3	
	No	7	58.3	5	41.7	
Profession	Social science	5	71.4	2	28.6	5.230
	STEM	28	36.8	48	63.2	
	Other	20	54.1	17	45.9	
Expected grade	A	28	53.8	24	46.2	8.738*
	B	18	31.0	40	69.0	
	Other^	7	70.0	3	30.0	

**Significant at the 0.01 level. *Significant at the 0.05 level. ^Since some counts were below 1, this category is formed from students who expect C and F grades.

TABLE 6 Bivariate and multivariate logistic regression results of predictor variables.

Variable	Category	Efficacy		OR (CI)	*p*	AOR (CI)	*p*
		Low	High				
Age				0.624	0.013	0.629	0.112
Expected mark				1.043	0.016	1.126	0.000
Gender	Male	24	36	1.403	0.359		
	Female	29	31	1			
School type	Government	29	31	0.713	0.359		
	Private	24	36	1			
Grade level	9	21	39	2.122	0.044	1.113	0.831
	10	32	28	1			
Living with	Single parent	4	19	4.849	0.007	6.704	0.009
	Both parents	49	48	1			
Expected grade	A	28	24	1			
	B	18	40	2.593	0.017	5.808	0.003
	C	4	0	0.000	0.999	0.000	0.999
	F	3	3	1.167	0.858	9.132	0.136
Receive tutorial	Daily	7	17	2.173	0.165		
	Weekly	25	30	1.074	0.869		
	Seldom	4	1	0.224	0.199		
	Never	17	19	1			
Profession	Social science	5	2	0.471	0.402		
	STEM	28	48	2.017	0.085		
	Other	20	17	1			
Plan to join college/ university	Definitely	36	57	1			
	May be	10	5	0.316	0.050	1.527	0.641
	No	7	5	0.451	0.201	3.579	0.306

As described in the "3.1 Students' mathematics self-efficacy" section, the mean score was used, and students were categorized as having high or low mathematics self-efficacy. Therefore, bivariate logistic regression models were computed to explore the variables associated with students' mathematics self-efficacy. Summaries of the results of the computations are presented in Table 6.

In the simple logistic regression analysis, age, grade level, living arrangements, expected grades in the upcoming national exam, expected marks from mathematics, and aspiration to join university were significant predictors.

In the multiple logistic regression analysis, only living arrangements, expected grades in the upcoming national exam, and expected marks in the semester from mathematics remain significant. Specifically, it was found that a one-point increase in the marks students expect to score in the semester is associated with a 1.13-point increase in their mathematics self-efficacy (AOR = 1.126, 95% CI = 1.055–1.202). Compared to students who expect A in the upcoming national examination, the odds of having high self-efficacy among students who expect B are five times higher (AOR = 5.808, 95% CI = 1.804–18.696). Students from single parents are six times more likely than students from both parents to have high self-efficacy (AOR = 6.704, 95% CI = 1.598–28.118).

4. Discussion

The main purpose of the present study was to examine mathematics self-efficacy and associated predictors among high school students in Addis Ababa. In this study, it was found that the overall mathematics self-efficacy of students is higher than average, and more than half of them have high mathematics self-efficacy, implying that our first hypothesis is confirmed. This finding is consistent with findings from other studies elsewhere. For example, in a study in the Philippines among secondary school students, the mean mathematics self-efficacy of the students was reported to be moderately high with a mean score of 3.25 (Laranang and Bondoc, 2020). Similarly, in a study of 10th graders in Bhutan, the students' MSE score is reported to be average with a score of 3.25 (Norbu and Dukpa, 2021). Likewise, Perez (2013), who conducted a study in Bangkok, Thailand revealed that more than half of the students showed a self-efficacy score higher than the mean score.

Coupled with this, a study among 12th graders in Bale Zone, Ethiopia assessed students' belief in mathematics education. The study found a mean score of 3.2 and argued that students have a medium level of belief in mathematics education (Demeke et al., 2020).

On the contrary, in a study in Uganda among high school students, it was found that the mean MSE of students was high with a rate of 4.35 on a scale of five (Batiibwe et al., 2020). The difference could be attributed to methodological variations, where the study in Uganda collected data from only 60 students, using 14 self-opinion items to measure mathematics self-efficacy.

So far as our second hypothesis is concerned, the independent sample *t*-tests and the ANOVA tests confirmed that there are variations in students' MSE based on grade level and their expected grades in the upcoming national exam. Besides, variations in the proportions of students with high and low MSE are observed based on grade level, living conditions with parents, and expected grades in the upcoming national exam.

With regard to predictors associated with students' MSE, i.e., our third hypothesis, the present study depicts that students' expected marks in the semester, their expected grades in the upcoming national exam, and living arrangements with parents are important factors.

In our study, a statistically significant difference was found between 9th graders and 10th graders, where the former had higher mathematics self-efficacy than the latter. This finding is similar to the findings of the study by Özgen and Bindak (2011). The study confirmed that self-efficacy beliefs in Mathematics Literacy vary based on grade, with 9th grade students having the highest ML self-efficacy beliefs and 12th grade students having the lowest. Likewise, in a study among elementary school students (grades 4–6), 4th graders were found to have higher mathematics self-efficacy than their 7th grade counterparts (Lloyd et al., 2005).

On the contrary, some studies found that the mathematics self-efficacy beliefs of high school students do not vary based on grade and that grade does not have an effect in explaining the variance in students' mathematics self-efficacy (Özyürek, 2010; Batiibwe et al., 2020). The difference in the findings can be explained through methodological differences. The study in Uganda found no difference based on grade level, but it was conducted among 60 students, making group comparisons difficult (Batiibwe et al., 2020). Similarly, the focus of the study by Özyürek (2010) is mainly on sources of self-efficacy scale, not actually on self-efficacy *per se* like our study. Another possible explanation for the differences in findings could be attributed to the fact that we include only two grade levels.

An important finding of the present study is that educational expectations (expressed in the form of anticipating high marks in the semester and good grades in the upcoming national examination) are important predictors of students' mathematics self-efficacy. Students who expect high grades in the upcoming exam and higher marks in the semester were found to have significantly higher mathematics self-efficacy than their counterparts. According to the findings of the present study, self-educational aspiration has been reported to have a positive relationship with students' MSE in a study among 5th and 6th graders in China (Liu et al., 2020). Similar findings were reported from a study in South Africa among 9th graders (Fadiji and Reddy, 2021).

According to Bandura's theory, students with high self-efficacy expect success in future tasks. Within the theory, it is claimed that it is individuals' efficacy that determines their goal-setting (Bandura, 1997). Thus, our findings substantiate Bandura's claim that self-efficacy and future expectations are related.

One of the interesting findings of this study is the effect of living arrangements with parents on students' mathematics self-efficacy. In the chi-square test, it is depicted that the majority of the students who are living with single parents have high MSE. Besides, the regression analysis has shown that the odds of students with a single parent having high MSE are six times higher than those of students who are living with both parents.

This finding is interesting in that living with single parents seems to be a contributing factor to students' mathematics self-efficacy. Given the relatively small numbers of students from single parents in the study (19.2%), this finding needs to be interpreted carefully.

Actually, other studies have indicated other parent-related factors in students' mathematics self-efficacy. For instance, in a study in the Philippines among high school students, mothers' occupation, fathers' occupation, and parental monthly income were related to students' MSE scores (Laranang and Bondoc, 2020).

Likewise, in a study in Greece using data from PISA among 7–11 graders socioeconomic status (measured by parental education, parental occupation, and an index of home possessions) was associated with students' MSE (Hiller et al., 2021). In another study that used data from PISA and included sample students from the United States and China, socioeconomic status (measured by parental education, parental occupation, and an index of home possessions) was associated with students' MSE (Wu, 2016).

Besides, the roles of parental involvement in students' MSE are reported in prior studies (Howard, 2015; Rodríguez et al., 2017). Although not on MSE, in a study among university students in Ethiopia, parenting style was found to be an important predictor of students' academic self-efficacy (Gota, 2012). Evenly speaking, the findings of the present study shed light on the role of a new parent-related factor (i.e., living with single parents *per se*) in shaping students' mathematics self-efficacy.

5. Conclusion, recommendations, and limitations

In the present study, it is evident that high school students in Addis Ababa have average mathematics self-efficacy. Given the critical role of students' mathematics self-efficacy in their career choice and success, this could go against the country's aspirations of attracting a large number of students to STEM fields.

The fact that educational expectations are important predictors of students' MSE in Addis Ababa implied the ecological validity of Bandura's social cognitive theory. Besides, the present study pinpointed the role of living arrangements with parents in predicting students' mathematics self-efficacy, which calls for further investigations, preferably longitudinal with a large sample size and a wider scope, including additional variables (e.g., parenting style). Furthermore, researchers and organizations that are working with and/or concerned about students' achievement, including but not limited to the Ministry of Education, need to gear their attention toward improving students' mathematics self-efficacy and, thus, their future academic success and career choice.

Finally, the fact that this study includes a small sample size from only two schools, is conducted with only two grade levels, and is cross-sectional are the limitations of the study.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Addis Ababa University, School of Psychology review committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

YY and SK conceived the study, developed the tool, and collected the data. YY analyzed the data. YY, AZ, and DG discussed the findings and developed the manuscript. All the authors read and approved the final manuscript.

Acknowledgments

The authors would like to thank the teachers in the two schools who helped in collecting the data. They also want to thank the research participants.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.927757/full#supplementary-material

References

Abebe, E. (2001). *Gender Differences in Mathematics Achievement as a Function of Math Self-Efficacy and Math Anxiety Among Bale Zone High School Students.* Addis Ababa: Addis Ababa University, College of Education.

Akter, S., Jabbar, M., and Khatun, M. (2018). Factors affecting career choice of the female secondary students in Khulna district of Bangladesh. *Khulna Univ. Stud.* 15, 91–103. doi: 10.53808/KUS.2018.15.1and2.1706-S

Ayele, M., and Dadi, T. (2016). Students' beliefs about mathematics learning and problem solving: The case of grade eleven students in West Arsi Zone, Ethiopia. *Educ. J.* 5, 62–70. doi: 10.11648/j.edu.20160504.14

Ayotola, A., and Adedeji, T. (2009). The relationship between mathematics self-efficacy and achievement in mathematics. *Proc. Soc. Behav. Sci.* 1, 953–957. doi: 10.1016/j.sbspro.2009.01.169

Bandura, A. (1997). *Self-efficacy: The Exercise of Control.* New York, NY: Freeman.

Batiibwe, M., Nannyonga, B., Taliba, C., Nalule, R., and Puglia, C. (2020). Investigating Math self-efficacy and math anxiety regarding gender, a-level math entry grade and mathematics achievement. *J. Educ. Pract.* 11, 41–53.

Cheema, J. (2018). Effect of math-specific self-efficacy on math literacy: Evidence from a Greek survey. *Res. Educ.* 102, 13–36. doi: 10.1177/0034523717741914

Clutts, D. (2010). *Mathematics Self-Efficacy of Community College Students in Developmental Mathematics Courses.* Lynchburg, VA: Liberty University.

Davidovitch, N., and Yavich, R. (2018). The effect of motivation and self-efficacy on math studies in the israeli ministry of education's program "give five.". *Int. J. High. Educ.* 7, 63–77. doi: 10.5430/ijhe.v7n6p63

Demeke, A., Belachew, T., and Negassa, G. (2020). Status of beliefs in mathematics education among grade twelve students_ in Bale Zone, Ethiopia. *J. Educ. Pract.* 11, 21–28.

Fadiji, A., and Reddy, V. (2021). School and Individual predictors of mathematics achievement in south africa: The mediating role of learner aspirations. *Afr. J. Res. Math. Sci. Technol. Educ.* 25, 65–76. doi: 10.1080/18117295.2021.1874687

Getahun, D., Adamu, G., Andargie, A., and Mebrat, J. (2016). Predicting mathematics performance from anxiety, enjoyment, value, and self-efficacy beliefs towards mathematics among engineering majors. *Bahir Dar J. Educ.* 16, 16–33.

Gota, A. (2012). *Effects of Parenting Styles, Academic Self-Efficacy, and Achievement Motivation on the Academic Achievement of University Students in Ethiopia.* Joondalup, WA: Edith Cowan University.

Heslin, P., and Klehe, U. (2006). "Self-Efficacy," in *Encyclopedia of Industrial/Organizational Psychology*, ed. S. G. Rogelberg (Thousand Oaks, CA: Sage), 705–708.

Hiller, S., Kitsantas, A., Cheema, J., and Poulou, M. (2021). Mathematics anxiety and self-efficacy as predictors of mathematics literacy. *Int. J. Math. Educ. Sci. Technol.* 53, 1–19. doi: 10.1080/0020739X.2020.1868589

Howard, N. (2015). *The Influences of Mathematics Self-Efficacy, Identity, Interest, and Parental Involvement on STEM Achievement in Algebra for Female High School Students. Dissertation*. Orange, CA: Chapman University.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *Eur. J. Psychol. Educ.* 28, 1–35. doi: 10.1007/s10212-011-0097-y

Johnson, C. (2008). *The Link Between Teacher Practices and High School Students' Mathematics Self-Efficacy: A Multilevel Analysis*. Dissertation. Cleveland, OH: Cleveland State University.

Kahveci, A. (2010). Quantitative analysis of science and chemistry textbooks for indicators of reform: A complementary perspective. *Int. J. Sci. Educ.* 32, 1495–1519. doi: 10.1080/09500690903127649

Kiamanesh, A., Hejazi, E., and Esfahani, Z. (2004). "The role of math self-efficacy, math self-concept, perceived usefulness of mathematics and math anxiety in math achievement," in *Proceedings of the 3rd International Biennial SELF Research Conference, Self-Concept, Motivation and Identity* (Sydney, NS: SELF Research Centre).

Laranang, J., and Bondoc, J. (2020). Attitudes and self- efficacy of students toward mathematics. *Int. J. Engl. Lit. Soc. Sci.* 5, 1393–1423. doi: 10.22161/ijels.55.11

Lent, R., Sheu, H., Singley, D., Schmidt, J., Schmidt, L., and Gloster, C. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *J. Vocat. Behav.* 73, 328–335. doi: 10.1016/j.jvb.2008.07.005

Lin, L., Lee, T., and Snyder, L. (2018). Math self-efficacy and STEM intentions: A person-centered approach. *Front. Psychol.* 9:2033. doi: 10.3389/fpsyg.2018.02033

Liu, T., Chen, X., Liu, M., Zhang, Y., Xin, T., and Wang, Y. (2020). The effects of childrens self-educational aspiration and self-efficacy on mathematics achievement: A moderated chained mediation model. *Ann. Psychol.* 36, 262–270. doi: 10.6018/analesps.366621

Liu, X., and Koirala, H. (2009). "The effect of mathematics self-efficacy on mathematics achievement of high school students," in *Proceedings of the Northeastern Educational Research Association (NERA) Annual Conference* (Philadelphia, PA: Northeastern Educational Research Association).

Lloyd, J., Walsh, J., and Yailagh, M. (2005). Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If i'm so smart, why do't i know it? *Can. J. Educ.* 28, 384–408. doi: 10.2307/4126476

Ministry of Education (2015). *Education Sector Development Programme V (ESDP V) 2008 - 2012 E.C. 2015/16 - 2019/20 G.C Programme Action Plan*. Addis Ababa: Federal Ministry of Education.

Norbu, P., and Dukpa, P. (2021). Investigating the level of the tenth-grade student's self-efficacy in mathematics and its impact on their performance: A study in Pemagatshel district. *Asian Res. J. Math.* 17, 34–46. doi: 10.9734/arjom/2021/v17i230272

Ogutu, J., Odera, P., and Maragia, S. (2017). The effectiveness of peer pressure on students' career decision making in secondary schools in Kenya. *Int. J. Humanit. Soc. Stud.* 3:8113. doi: 10.20908/ijarsijhss.v3i3.8113

Özgen, K., and Bindak, R. (2011). Determination of self-efficacy beliefs of high school students towards math literacy. *Educ. Sci. Theory Pract.* 11, 1085–1089.

Özyürek, R. (2010). The reliability and validity of the mathematics self-efficacy informative sources scale. *Educ. Sci. Theory Pract.* 10, 439–447.

Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemp. Educ. Psychol.* 21, 325–344. doi: 10.1006/ceps.1996.0025

Pajares, F., and Kranzler, J. (1995b). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemp. Educ. Psychol.* 20, 426–443. doi: 10.1006/ceps.1995.1029

Pajares, F., and Kranzler, J. (1995a). "Role of self-efficacy and general mental ability in mathematical problem-solving: A path analysis," in *Paper Presented at the Annual Meeting of the American Educational Research Association* (San Francisco, CA: American Educational Research Association).

Pajares, F., and Miller, M. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *J Educ Psychol.* 86, 193–203. doi: 10.1037/0022-0663.86.2.193

Perez, E. (2013). The relationship between mathematics self-efficacy and mathematics achievement of Mathayomsuksa students in the English program of St. Joseph Bangna school. *Hum. Sci.* 5:82.

Probst, J. A. (2019). *Causal-Comparative Analysis of Mathematics Self-Efficacy Based on Gender and Math Acceleration*. Dissertation. Lynchburg, VA: Liberty University.

Recber, S., Isiksal, M., and Koc, Y. (2018). Investigating self-efficacy, anxiety, attitudes and mathematics achievement regarding gender and school type. *Ann. Psychol.* 34, 41–51. doi: 10.6018/analesps.34.1.229571

Reynolds, M., and Miller, G. (2003). "Current perspectives in educational psychology," in *Handbook of Psychology: Educational Psychology*, eds W. M. Reynolds, and G. E. Miller (Hoboken, NJ: Wiley), 3–20. doi: 10.1002/0471264385.wei0701

Rodríguez, S., Piñeiro, I., Gómez Taibo, M., Regueiro, B., Estévez, I., and Valle, A. (2017). An explanatory model of maths achievement: Perceived parental involvement and academic motivation. *Psicothema* 29, 184–190.

Rodríguez, S., Regueiro, B., Piñeiro, I., Valle, A., Sánchez, B., Vieites, T., et al. (2020). Success in mathematics and academic wellbeing in primary-school students. *Sustainability* 12:3796. doi: 10.3390/su12093796

Roslan, N., and Maat, S. (2019). Systematic literature review on secondary school students mathematics self-efficacy. *Int. J. Acad. Res. Bus. Soc. Sci.* 8, 975–987. doi: 10.6007/IJARPED/v8-i4/6817

Schulz, W. (2005). "Mathematics Self-efficacy and student expectations. results from PISA 2003," in *Proceedings of the Paper prepared for the Annual Meetings of the American Educational Research Association in Montreal* (Montreal, OC: American Educational Research Association).

Shaine, M. (2015). The effect of self-regulated learning strategies and self-efficacy on academic achievement of primary school students. *Psychol. Behav. Sci.* 4:107. doi: 10.11648/j.pbs.20150403.14

Tait–McCutcheon, S. (2008). "Self-efficacy in mathematics: Affective, cognitive, and conative domains of functioning," in *Proceedings of the 31st annual conference of the Mathematics Education Group of Australasia* (Brisbane, QL: MERGA), 507.

Tekola, N., Getahun, D., and Hagos, H. (2020). Gender similarities in high school mathematics: Affective and cognitive aspects. *Bahir Dar J. Educ.* 9, 105–124.

Tizazu, S., and Ambaye, D. (2017). Assessment on Students academic self-efficacy, self-concept, goal orientation and academic achievement: The case of Debre Markos higher education preparatory school, Ethiopia. *Eur. J. Educ. Stud.* 3, 778–792.

Turgut, M. (2013). Academic self-efficacy beliefs of undergraduate mathematics education students. *Acta Didact. Napoc.* 6, 33–40.

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. *Am. Educ. Res. J.* 50, 1081–1121. doi: 10.3102/0002831213488622

Wu, Y. S. (2016). Universal beliefs and specific practices?: Students math self-efficacy and related factors in the United States and China. *Int. Educ. Stud.* 9, 61–74. doi: 10.5539/ies.v9n12p61

Wubalem, A. (2006). *Relationship of Gender & Mathematics Self-Efficacy to Mathematics Acheivement Among Grade 5 and 9 Students*. M.A. Thesis. Addis Ababa: Addis Ababa University.

Zander, L., Höhne, E., Harms, S., Pfost, M., and Hornsey, M. (2020). When grades are high but self-efficacy is low: Unpacking the confidence gap between girls and boys in mathematics. *Front. Psychol.* 11:552355. doi: 10.3389/fpsyg.2020.552355

Zarch, M., and Kadivar, P. (2006). "The role of mathematics self-efficacy and Mathematics ability in the structural model of Mathematics performance," in *Proceedings of the 9th WSEAS International Conference on Applied Mathematics*, Istanbul, 242–249.

Zedan, R., and Bitar, J. (2014). Environment learning as a predictor of mathematics self - efficacy and math achievement. *Am. Int. J. Soc. Sci.* 3, 85–97.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fpsyg.2022.1044580

Check for updates

OPEN ACCESS

EDITED BY
Shuhua An,
California State University,
United States

REVIEWED BY
Jaroslav Říčan,
Jan Evangelista Purkyně University in Ústí nad Labem, Czechia
Limei Wang,
South west University,
China
Mohd Effendi Ewan Mohd Matore,
National University of Malaysia, Malaysia

*CORRESPONDENCE
Jian Liu
✉ professorliu9506@126.com
Zhiyong Xie
✉ 503214139@qq.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 14 September 2022
ACCEPTED 22 December 2022
PUBLISHED 13 January 2023

CITATION
Wang T, Zhang L, Xie Z and Liu J (2023) How does mathematical modeling competency affect the creativity of middle school students? The roles of curiosity and guided inquiry teaching.
Front. Psychol. 13:1044580.
doi: 10.3389/fpsyg.2022.1044580

How does mathematical modeling competency affect the creativity of middle school students? The roles of curiosity and guided inquiry teaching

Tian Wang[1,2], Libin Zhang[1], Zhiyong Xie[3]* and Jian Liu[1,2]*

[1]Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University, Beijing, China, [2]China Education Innovation Institute, Beijing Normal University, Zhuhai, China, [3]College of Teacher Education, South China Normal University, Guangzhou, China

Introduction: Mathematical modeling has become a crucial competence in mathematics education in many countries and regions due to the increasingly complex real-world problems that students face in the 21st century. Previous research has shown that mathematical modeling contributes to the development of students' creativity, particularly with respect to stimulating and protecting the curiosity of children. However, previous studies have not explored or examined the relationships among middle school students' mathematical modeling competency, curiosity, and creativity based on data drawn from large-scale assessments and have not investigated the influence of teachers' teaching methods in this context.

Methods: This study used convenience sampling to select 4,531 seventh-grade students from eastern and western, urban and rural areas in China. Online tests and questionnaires were used to measure their mathematical modeling competency, curiosity, creativity and guided inquiry teaching, and a moderated mediation model was used to analyze the effect of mathematical modeling competency on creativity.

Results: The results showed the following. (1) There are statistically significant differences between boys and girls in terms of their mathematical modeling competency, curiosity, and creativity. Specifically, boys score significantly higher than girls on these variables. (2) Creativity exhibits a statistically significant positive correlation with mathematical modeling competency, curiosity, and guided inquiry teaching. (3) Curiosity mediates the relationship between mathematical modeling competency and creativity, and guided inquiry teaching moderates the influence of curiosity. In high-level guided inquiry teaching classes, curiosity has a stronger influence on creativity, and it mediates the relationship between mathematical modeling competency and creativity more strongly.

Discussion: This study empirically verified the influence of mathematical modeling competency on creativity and provided a possible way to cultivate children's creativity. Future research should use longitudinal analysis to verify the causal relationship between mathematical modeling competency and creativity and to systematically explore the possible path by which mathematical modeling competency affects creativity.

KEYWORDS

mathematical modeling competency, creativity, curiosity, guided inquiry teaching, middle school students

1. Introduction

Due to the continuous development of science and technology, mathematical modeling has come to play an increasingly important role in promoting social development and people's ability to adapt to life, and mathematical modeling competency has become a key competency of future citizens (Kaiser, 2007). Since mathematics curriculum reform was initiated at the beginning of this century, many countries and regions have incorporated the cultivation of mathematical modeling competency or related modeling ideas and applications into mathematics curricula or teaching practices as an important goal of mathematics education (Blum et al., 2007; Cai, 2017) and have considered this skill to represent a necessary key competency for students in the 21st century (Cai and Xu, 2016). In the United States, the *Common Core State Standards for mathematics* list mathematical modeling as one of the eight Standards for Mathematical Practice so that students at all stages of learning can come to understand that mathematics can be used to solve problems in the real world (Common core state standards initiative, 2010). The description and requirements of mathematical modeling in China's Mathematics Curriculum Standard undergo constant improvement (Huang et al., 2019). Mathematical modeling is a key component of literacy at the primary and middle school stages, and it guides students to learn from an early age that "mathematical models can be used to solve a class of problems and are the basic way to apply mathematics," according to the *Mathematics Curriculum Standard for Compulsory Education* (The Ministry of Education of the People's Republic of China, 2022).

Since mathematical modeling usually involves the task of solving unconventional and open problems in the real world, a task which requires creativity from the modeler to understand the real situation and propose new solutions (Niss and Blum, 2020), mathematical modeling is closely related to students' creativity. In the face of the increasingly complex living and working environment of the 21st century, creativity has become an indispensable ability that allows people to cope with these new challenges and problems (OECD, 2010), and it has promoted the development of all aspects of society (Hennessey and Amabile, 2010). The China Education Innovation Institute of Beijing Normal University and the Twenty-First Century Learning (P21) of the United States incorporated "Creativity Competence" into *the 5Cs Framework for Twenty-first Century Key Competences* and claimed that new knowledge, new technology, new crafts and new values can be achieved *via* creativity (Gan et al., 2020). This concept can replace traditional resources, energy and capital as the driving force of sustainable economic development, and it emphasizes the fact that problem-solving based on real situations can enhance students' creativity (Gan et al., 2020). In recent years, various international mathematical modeling activities have attached great importance to students' creativity. The topics associated with modeling tasks cover cutting-edge fields such as "global warming," "renewable energy" and "self-driving vehicles," and mathematical modeling is used to stimulate and cultivate students' creativity (Mei, 2018). Especially for primary and middle school students, the openness and uncertainty associated with mathematical modeling tasks can facilitate their development of creativity because they are full of curiosity and accustomed to creating (COMAP and SIAM, 2016).

Although correlations may exist between mathematical modeling and creativity, only a few studies have focused on this relationship (Wessels, 2014; Lu and Kaiser, 2022a), and there is a lack of large-scale evaluations based on empirical studies to verify these studies. This study aims to explore the path by which mathematical modeling competency influences the development of creativity by reference to large-scale evaluation data as well as to investigate the stimulation and cultivation of creativity in middle school students.

1.1. Mathematical modeling competency and creativity

Mathematical modeling is a cyclic process by which mathematics can be used to solve real problems, and it thus facilitates a two-way transformation between the mathematical world and the real world (Niss et al., 2007; Blum and Ferri, 2016). Mathematical modeling competency refers to the ability of a person to perform the required operations in a modeling environment to promote modeling (Niss et al., 2007) and is composed of the sub-competencies that are necessary to complete each step of the modeling cycle (Kaiser, 2007). A widely accepted model of the sub-competencies of mathematical modeling mainly includes five sub-competencies: simplifying, mathematising, working mathematically, interpreting and validating. Simplifying is the competency to understand real-world problems and develop real-world models; mathematising is the competency to establish mathematical models based on real-world models; working mathematically is the competency to solve mathematical problems in mathematical models; interpreting is the ability to interpret mathematical results in real-world models or situations; and validating is the competency to challenge the solution thus developed and to implement the modeling process again, if necessary (Maaß, 2006).

In a broad sense, creativity focuses mainly on everyday creativity, which refers the creative thinking in which everyone can engage in daily life and which can be improved *via* education and practice (OECD, 2019). Creativity denotes the ability of an individual to use relevant information and resources to produce novel and valuable ideas, programs, and products. It mainly includes three elements: creative personality, creative thinking and creative task engagement (Gan et al., 2020). Creative personality comprises the characteristics of curiosity, open mindedness, the courage to take on challenge and risks and independent self-confidence. Creative thinking comprises divergent thinking, convergent thinking and restructuring thinking, which are helpful when engaging in innovative activities. Creative task engagement involves participating and investing in practices that aim to produce novel and valuable results.

Mathematical modeling can effectively promote the development of students' creativity and their mastery of mathematical knowledge and skills (Wessels, 2014). The development of creativity relies on mathematical tasks associated with higher levels of cognitive activity that are intended to stimulate students' high-level cognitive processes (Leikin and Elgrably, 2020), and the characteristics of higher cognitive requirements for mathematical modeling tasks may help improve students' creativity (Lu and Kaiser, 2022a). Lu and Kaiser (2022a) examined the process of mathematical modeling from the perspective of creativity and proposed a model of mathematical modeling cycle theory that includes creativity on the basis of the mathematical modeling process model (see Figure 1). According to this model, many processes involve creativity: in understanding and simplifying, the modeler analyzes a real situation from various perspectives, thereby generating various models of reality; in mathematical working, the modeler obtains the results of the mathematical model using various methods; in interpreting, the modeler interprets the mathematical results as results in the real world; and in validating, the modeler employs a variety of approaches to test the correctness of the results in a real-world situation. This model was applied to high school students, preservice mathematics teachers and in-service mathematics teachers. The results showed that mathematical modeling competency is significantly correlated with creativity (Wessels, 2014; Suh et al., 2017; Lu and Kaiser, 2022b). For high school students, the difficulty of modeling tasks may affect the relationship between mathematical modeling competency and creativity (Lu and Kaiser, 2022a). Wessels (2014) conducted a two-year longitudinal study to investigate preservice mathematics teachers. By analyzing all the materials (drafts, charts, formulas, etc.) involved in the process of solving mathematical modeling tasks, he found that the mathematical modeling process can effectively improve the individual's level of creativity. Suh et al. (2017) selected two primary school teachers and their students in two classes as research objects. By reference to interviews with teachers, classroom observation and analysis of students' works, these authors found that mathematical modeling can effectively improve students' creativity, critical thinking, communication and collaboration.

Previous studies have provided evidence that creativity is involved in the mathematical modeling process (Lu and Kaiser, 2022a) and obtained certain empirical evidence in the context of teaching (Wessels, 2014; Suh et al., 2017). However, the potential relationship between mathematical modeling competency and creativity has rarely been explored by reference to large-scale assessments. It is necessary to verify the effectiveness of mathematical modeling for creativity cultivation based on empirical results. Simultaneously, the cultivation of children's creativity has always received a great deal of attention in the field of educational psychology (Camp, 1994; Cheung et al., 2004; Hu et al., 2011). Researchers have found that the development of creativity in upper elementary school shows an upward trend;

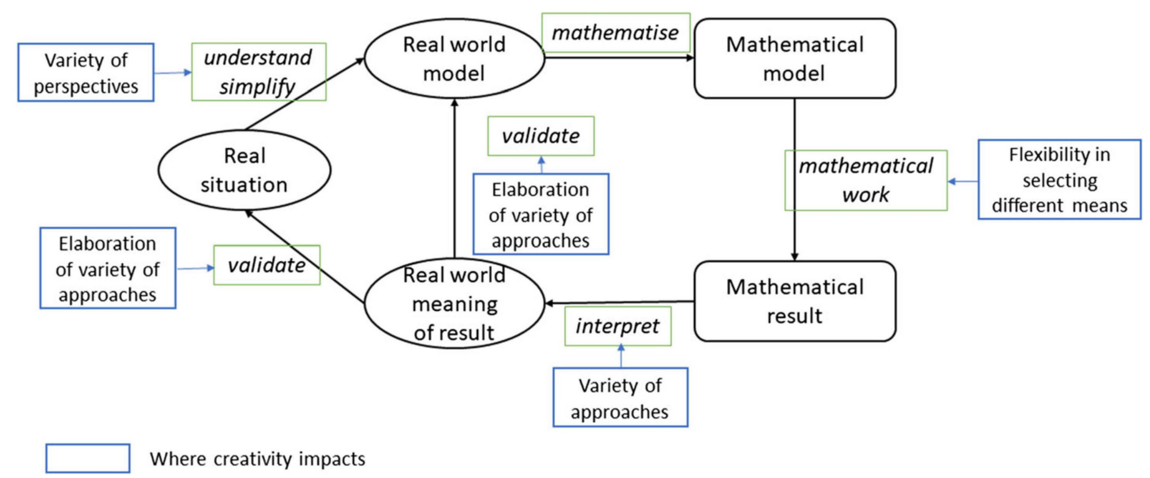


FIGURE 1
Modeling cycle enriched by aspects of creativity (Lu and Kaiser, 2022a) (CC-BY) (http://creativecommons.org/licenses/by/4.0/).

based on these findings, Smith and Carlsson proposed that the development of creativity may originate from high grades in primary school (Smith and Carlsson, 1990; Camp, 1994). Therefore, the impact of mathematical modeling on the creativity of primary and middle school students deserves further attention. In summary, **Research Hypothesis 1** of this study proposes that there is a significant positive correlation between the mathematical modeling competency of middle school students and their creativity.

1.2. Curiosity, mathematical modeling competency, and creativity

Curiosity is a ubiquitous psychological trait among humans. This term refers to one's desire for learning even when the application of the knowledge is not readily apparent (Facione et al., 1994). Studies have shown that curiosity has a positive effect on individual creativity: the stronger an individual's curiosity is, the greater that individual's creativity (Celik et al., 2016; Hardy et al., 2017). The results of a meta-analysis of research regarding the relationship between curiosity and creativity indicated a moderate positive correlation between curiosity and creativity (Schutte and Malouff, 2020). According to the creative process model proposed by Mumford and McIntosh (2017), problem definition and information collection represent the initial steps of this process, and curiosity helps individuals collect information and define the problem that is to be solved (Schutte and Malouff, 2020). Hardy et al. (2017) analyzed the relationships among various types of curiosity, creative performance and creative problem-solving by reference to 122 college students and found that diversive curiosity (e.g., I find it interesting to learn new information) has a positive effect on creative performance *via* the complete mediation of the information gathering behavior associated with creative problem-solving. Simultaneously, as a positive emotion that provides motivation, curiosity may encourage new exploratory ideas, and behaviors (Fredrickson and Joiner, 2018), thereby enhancing individual creativity.

Mathematical modeling plays an important role in stimulating curiosity. From the perspective of cognitive neuroscience, curiosity is the result of situation-based prediction errors and information-based prediction errors (Gruber and Ranganath, 2019). On the one hand, when an individual is faced with a new or changing situation, a gap emerges between the prediction generated by his or her hippocampus and the actual situation at hand, which leads to exploratory behavior to address the associated uncertainty (O'Keefe and Nadel, 1979). On the other hand, when the knowledge that the individual wants to obtain is beyond that individual's current level of knowledge or does not conform to the individual's prior knowledge, an information gap is generated, thus triggering the individual's curiosity (Litman, 2005; Gottlieb et al., 2013). By reference to classroom observations and interviews with teachers and students, Geiger et al. (2022) found that the openness and uncertainty exhibited by the real situation represent essential features of mathematical modeling tasks. Students may understand and analyze the situation they face in modeling tasks based on different sorts of previous experience and knowledge. The authenticity and diversity of the situations that are faced by students are quite different from those associated with ordinary mathematics problems, resulting in the occurrence of situational prediction errors that stimulate students' curiosity. Simultaneously, mathematical modeling tasks are more challenging than traditional mathematics questions and thus require a higher level of cognition (English, 2021), so students are often unable to solve modeling tasks through simple memorization or the repetition of prior knowledge. Instead, it is necessary to integrate that prior knowledge with flexible applications based on *a priori* knowledge, thereby generating prediction errors regarding the information at hand and generating curiosity in students.

There is a strong positive correlation between individual curiosity and creativity (Celik et al., 2016; Hardy et al., 2017; Schutte and Malouff, 2020). The openness of mathematical modeling tasks and the high cognitive level required in this context may stimulate students' curiosity (Litman, 2005; COMAP and SIAM, 2016; English, 2021), which may allow mathematical modeling to promote the development of creativity by enhancing individual curiosity. Further exploration of this path of influence may improve the theoretical model of creativity cultivation and promote the connection between mathematics learning and creativity cultivation. Since students' curiosity may be stimulated during mathematical modeling processes and curiosity is an important predictor of creativity, curiosity may mediate the relationship between mathematical modeling competency and creativity. In summary, **Research Hypothesis 2** of this study proposes that curiosity mediates the relationship between middle school students' mathematical modeling competency and their creativity.

1.3. Guided inquiry teaching, curiosity, and creativity

Guided inquiry teaching is a type of inquiry-based teaching. Inquiry-based teaching employs a student-centered approach in which teachers pose particular questions, such as open-ended or divergent questions, which allow students to respond in different ways (Oliveira, 2010). Inquiry-based teaching mainly includes the actions of making observations; asking questions; examining known information; planning surveys; reviewing known information based on experimental evidence; using tools to collect, analyze and interpret data; proposing answers; explaining and predicting; and exchanging results (National Research Council, 1996). In guided inquiry teaching, the source of the tasks or questions under consideration is the teacher, and the data collection methods and the interpretation of the results are designed and completed by the students; that is, the teacher provides the students with the questions that are to be investigated and any necessary information, while the students are required to design the inquiry program and develop a plan to solve and answer the problem by themselves (Blanchard et al., 2010; Chen et al., 2021). Studies have shown that inquiry-based teaching can effectively improve students' academic performance (Schroeder et al., 2007; Minner et al., 2010) and has a positive impact on students' learning attitudes and interests (Jiang and McComas, 2015).

Inquiry-based teaching may affect students' curiosity and creativity as well as the relationship between those two factors. Erbas and Yenmez (2011) found that inquiry-based learning is positively correlated with curiosity, and Schijndel et al. (2018) found that students in an inquiry-based teaching experimental group exhibited greater curiosity than students in the control group. Rodríguez et al. (2019) found that inquiry-based learning can effectively promote the development of students' creativity in controlled experiments. Guided inquiry teaching can stimulate students' diversity of problem definition and information collection during the creative process, thus enabling students to generate new ideas. Studies have found that teaching methods or activities such as brainstorming and idea linking can affect the relationship between curiosity and creativity and that different teaching methods or activities may have different degrees of impact on the relationship between the two factors. Idea linking activities can promote students' creativity more effectively than can brainstorming because they can establish connections with previous ideas (Hagtvedt et al., 2019). One characteristic of inquiry-based teaching is the generation of uncertainty (Schijndel et al., 2018), which is embodied in the possibility of collecting different information, using different methods, and obtaining different results. The uncertainty may affect the process by which the individual's curiosity leads to the generation of new ideas, which may in turn affect the individual's creativity. Loewenstein (1994), Choi et al. (2015), Yang et al. (2016), and Gruber and Ranganath (2019).

As a teaching method that stimulates curiosity and promotes creativity, guided inquiry teaching may regulate the relationship between students' curiosity and creativity. Therefore, **Research Hypothesis 3** of this study proposes that guided inquiry teaching regulates the relationship between middle school students' curiosity and their creativity.

1.4. The current study

Based on previous studies, to explore the ways in which mathematical modeling competency influences creativity *via* curiosity as well as the influence of guided inquiry teaching on the relationship between curiosity and creativity, this study constructed a moderated mediation model of the influence of mathematical modeling competency on creativity (see Figure 2):

> ***Research Hypothesis 1***: The mathematical modeling competency of middle school students positively affects their creativity.
>
> ***Research Hypothesis 2***: Curiosity mediates the effect of mathematical modeling competency on the creativity of middle school students.

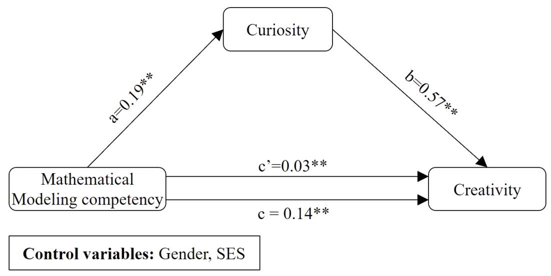


FIGURE 2
Hypothetical indirect pathways between mathematical modeling competency and creativity.

> ***Research Hypothesis 3***: Guided inquiry teaching regulates the relationship between middle school students' curiosity and their creativity.

2. Materials and methods

2.1. Participants and data collection

This study used convenience sampling to select students from 83 middle schools, with 44 middle schools (53.0%) located in eastern China and 39 (47.0%) located in western China. The participants were 4,620 seventh-grade students who were enrolled in 2021; participants ranged from 11 to 13 years old at the time of the study. After removing incomplete data, we obtained 4,531 valid subjects, including 2,377 boys (52.5%) and 2,154 girls (47.5%).

In this study, data were collected using online tests and online questionnaires. All students who participated in the test entered the computer classroom and utilized the test platform under the guidance of the teacher. After starting the test, the students completed an online test of mathematical modeling competency (with a time limit of 40 min) and an online questionnaire (which had no time limit). Subsequently, the students answered an ungraded warm-up question to familiarize themselves with the various operations and answer specifications of the platform. The online test of mathematical modeling competency was to measure students' mathematical modeling competency, and the online questionnaire was to measure students' performance on curiosity, creativity and guided inquiry teaching.

2.2. Measures

2.2.1. Online test of mathematical modeling competency

The online test of mathematical modeling competency included one warm-up question and two formal tasks, which were developed by the Beijing Normal University Regional Assessment of Education Quality (RAEQ). There were 15 items in the two formal tasks, which were choice items and open-ended

10.3389/fpsyg.2022.1044580

items (see Appendix A). The assessment framework was based on the five sub-competencies model of mathematical modeling proposed by Kaiser (2007). The model was simplified and revised by using the thinking aloud and interviews of Chinese primary and middle school students to make it more suitable for their process of mathematical modeling. Finally, an assessment framework of mathematical modeling competency was developed, which includes four sub-competencies: understanding information, making models, working mathematically, and interpreting and validating. There were four items each for understanding information, making models and interpreting and validating, and three items for working mathematically. The RAEQ developed two modeling tasks for this online test based on the assessment framework, and the task situations were derived from *Mathematical Modelling: A Guidebook for Teachers and Teams* Galbraith and Holton (2018) and *Mathematical Modeling Handbook II: The Assessments* (Fletcher et al., 2013). After this adaptation had been completed, a 6-person thinking aloud session and external reviews by experts were used to examine the content validity of the test, and the pretest was used to examine its construct validity. Each open-ended item was coded by two or more raters and the interrater reliability scores were all more than 0.9. The test uses the IRT method to combine students' scores on each item into an overall test score, which is standardized to produce a mathematical modeling competency score, in which context 500 is the average score and 100 is the standard deviation. The higher an individual's score is, the higher that individual's mathematical modeling competency. The confirmatory factor analysis results of the test were as follows: comparative fit index (CFI) = 0.931 > 0.90, Tucker-Lewis index (TLI) = 0.915 > 0.90, root mean square error of approximation (RMSEA) = 0.043 < 0.08, and standardized root mean square residual (SRMR) = 0.027 < 0.08 (Brown and Cudeck, 1993; Hox and Bechger, 1998; Kline, 2005). Cronbach's alpha coefficient was 0.758, and the reliability and validity were acceptable.

2.2.2. Creativity questionnaire

The creativity questionnaire was developed by the RAEQ and contains a total of 3 dimensions including 26 questions (see Appendix B). This questionnaire is based on the theoretical framework of creativity included in the 5Cs Framework for Twenty-First Century Key Competences and contains the three dimensions of creative personality, creative thinking, and creative task engagement (Gan et al., 2020). The questionnaire uses a five-point Likert scale to calculate the average score of each item as the creativity score (1 = Strongly disagree, 5 = Strongly agree). The higher this score is, the higher the individual's level of creativity. This questionnaire was developed by reference to expert interviews, teacher interviews, and student pretests to ensure its validity. The Cronbach's alpha coefficient of the creativity questionnaire was 0.937, which is acceptable. The results of the confirmatory factor analysis of the creativity questionnaire indicated that $x^2/df = 19.37$, CFI = 0.922, TLI = 0.912, RMSEA = 0.064, and SRMR = 0.055.

2.2.3. Curiosity questionnaire

The curiosity questionnaire is based on the adaptation of questions related to curiosity drawn from the California Critical Thinking Disposition Inventory (CCTDI) compiled by Facione et al. (1998). It contains a total of 5 items (see Appendix B). This questionnaire mainly examines respondents' attitudes toward researching new things and their expectation of facing challenges. The questionnaire uses a five-point Likert scale to calculate the average of the scores of each item as the curiosity score (1 = Strongly disagree, 5 = Strongly agree). The higher this score is, the stronger the curiosity of the individual. This questionnaire underwent expert interviews, teacher interviews, and student pretests to ensure its validity. The Cronbach's alpha coefficient of the curiosity questionnaire was 0.939, which is acceptable. The results of the confirmatory factor analysis of the curiosity questionnaire indicated that $x^2/df = 10.03$, CFI = 0.998, TLI = 0.995, RMSEA = 0.045, and SRMR = 0.006.

2.2.4. Guided inquiry teaching questionnaire

The guided inquiry teaching questionnaire was developed by the RAEQ and contains a total of 5 items (see Appendix B). The project team developed a questionnaire based on the characteristics of the guided inquiry teaching process proposed by Blanchard et al. (2010). The questionnaire used a five-point Likert scale to calculate the average of the scores of each item as the guided inquiry teaching score (1 = Strongly disagree, 5 = Strongly agree). The higher this score is, the more strongly the individual feels that the teacher used guided inquiry teaching. This questionnaire underwent expert interviews, teacher interviews, and student pretests to ensure its validity. The Cronbach's alpha coefficient of the curiosity questionnaire was 0.946, which is acceptable. The results of the confirmatory factor analysis of the guided inquiry teaching questionnaire indicated that $x^2/df = 7.37$, CFI = 0.999, TLI = 0.997, RMSEA = 0.037, and SRMR = 0.004.

2.2.5. Demographic variables

This study mainly investigated demographic variables such as the gender and family socioeconomic status (SES) of students. Among these variables, SES was assessed using the relevant part of the PISA 2012 technical report of OECD (2014), which is mainly divided into three parts: parents' level of education, highest occupational status, and family possessions. The level of education primarily refers to the highest level of education attained by the parents of the surveyed student, and the highest level of education attained by the parents is regarded as the "parents' level of education" of the student. The highest occupational status of the parents refers to the main job of the parents of the surveyed student, and the highest occupational status of the parents is regarded as the "parental highest occupational status" of the student. Family possessions are measured mainly in terms of four aspects: family wealth, cultural possessions, family education resources, and family books. Finally, the standardized scores of the three components of SES are used as composite scores of the students' SES.

3. Results

3.1. Common method biases analysis

The Harman single-factor test was used (Malhotra et al., 2006) to evaluate common method bias (Podsakoff et al., 2003). We conducted exploratory factor analysis to investigate all 36 items related to creativity, curiosity and guided inquiry teaching using SPSS 28 software (SPSS Inc., Chicago, Illinois, United States). The results showed that Bartlett = 113974.69, $df = 630$, $p < 0.01$, KMO = 0.963, communalities variance = 69.56%, the total variance explained by the first common factor was 19.16%, i.e., less than the critical value of 40% (Tang and Wen, 2020). Therefore, no common method bias affected the results of the current study.

3.2. Descriptive statistics and correlation analysis

Prior to the formal analysis, all main variables were tested for normality, and the results showed that all variables had kurtosis values between −0.339 and 0.618 and skewness values between −0.957 and 0.220, indicating that all variables followed a normal distribution. Table 1 shows the means (M), standard deviations (SDs), gender-based differences, and intercorrelations among the key variables. Children's mathematical modeling competency, curiosity, and guided inquiry teaching were significantly and positively related to their creativity ($r = 0.149 \sim 0.586$, $p_s < 0.01$). Both children's mathematical modeling competency and their curiosity were positively correlated with guided inquiry teaching ($r = 0.167$, 0.507, $p_s < 0.01$). The correlation between children's mathematical modeling competency and curiosity was also significant ($r = 0.192$, $p_s < 0.01$). In addition, independent-sample t tests indicated significant gender differences in mathematical modeling competency, creativity, and curiosity; specifically, boys scored significantly higher than girls on those variables.

3.3. The mediating effect of curiosity

To examine the mediating effect of mathematical modeling competency on creativity *via* curiosity, we computed 95% confidence intervals (95% CIs) using the bootstrap method with 5,000 replications with the help of the PROCESS 4.0 plug-in (Model 4). After controlling for gender and SES, mathematical modeling competency significantly positively predicted curiosity ($\beta = 0.19$, $p < 0.001$). As shown in Figure 3, curiosity significantly positively predicted creativity ($\beta = 0.57$, $p < 0.001$), and mathematical modeling competency had a significant direct effect on creativity ($\beta = 0.03$, $p = 0.004$). The 95% confidence interval of the bootstrap mediation effect did not include zero (effect size = 0.108, 95% CI [0.092, 0.126], accounting for 78.57% of the total effect), thus indicating that mathematical modeling competency has a significant mediating effect on creativity.

3.4. The moderating effect of guided inquiry teaching

To test Hypothesis 3, which pertained to the moderating effect of guided inquiry teaching on the influence of curiosity on creativity, we used Model 14 in PROCESS 4.0 to conduct the relevant analysis. As shown in Table 2, after controlling for gender and SES, curiosity and guided inquiry teaching significantly positively predicted creativity ($\beta = 0.49$, 0.20, $p < 0.001$), and the interaction between guided inquiry teaching

TABLE 1 Means, SDs, gender difference, and intercorrelations among key variables.

	1	2	3	4	5	6
1. Gender	1					
2. SES	0.023	1				
3. Mathematical modeling competency	0.035*	0.040**	1			
4. Creativity	0.104**	0.084**	0.149**	1		
5. Curiosity	0.071**	0.048**	0.192**	0.586**	1	
6. Guided inquiry teaching	0.009	0.031*	0.167**	0.440**	0.507**	1
M ± SD (the whole sample)	–	–	483.57 ± 99.61	3.72 ± 0.71	3.91 ± 0.94	4.12 ± 0.89
M ± SD (Boys)	–	–	486.9 ± 103.1	3.79 ± 0.74	3.97 ± 0.97	4.12 ± 0.92
M ± SD (Girls)	–	–	479.9 ± 95.53	3.64 ± 0.66	3.83 ± 0.90	4.11 ± 0.87
t			2.39	7.06	4.85	0.59
p			0.017	<0.001	<0.001	0.554

$^{*}p < 0.05$; $^{**}p < 0.01$.

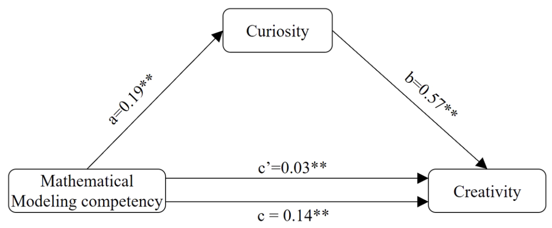


FIGURE 3
The mediating effect of curiosity on mathematical modeling competency and creativity.

TABLE 2 The moderating effect of guided inquiry teaching.

	Creativity		
	β	SE	*t*
Gender	0.121	0.024	5.139**
SES	0.002	0.000	4.570**
Mathematical modeling competency	0.022	0.012	1.869
Curiosity	0.478	0.014	34.785**
Guided inquiry teaching	0.203	0.014	14.455**
Guided inquiry teaching * curiosity	0.031	0.010	3.074**
R^2	0.380		
F	461.796**		

**$p < 0.01$.

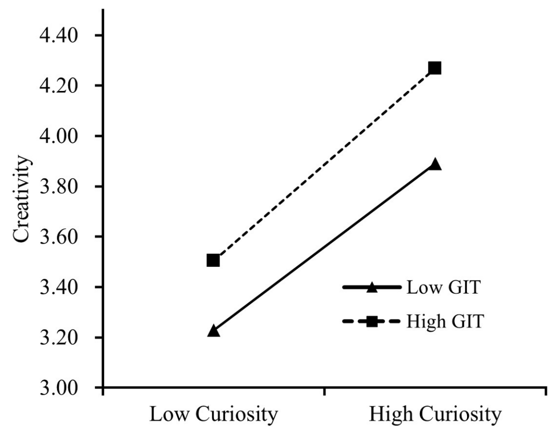


FIGURE 4
Simple slope test.

and curiosity was significantly related to creativity ($\beta = 0.03$, $p = 0.004$). Thus, the results confirm our hypothesis that guided inquiry teaching moderates the relationship between curiosity and creativity.

In addition, to improve our understanding of this interaction, we plotted the simple slope (Figure 4), and the results of the simple slope test show that in the case of high levels of guided inquiry teaching (M + 1SD), the slope is $B = 0.51$, $p < 0.001$, while in the case of low levels of guided inquiry teaching (M–1SD), the slope is $B = 0.44$, $p < 0.001$, thus indicating that curiosity has a stronger positive impact on creativity in the high guided inquiry teaching condition. Mathematical modeling competency also had a stronger mediating effect on creativity *via* curiosity in this condition (effect size = 0.10, 95% CI [0.08, 0.11]).

4. Discussion

4.1. Mathematical modeling positively affects the creativity of middle school students

This study found that the mathematical modeling of middle school students plays a positive role in their development of creativity; accordingly, Research Hypothesis 1 was verified. The stronger a student's mathematical modeling competency is, the higher that student's level of creativity. To a certain extent, these research results support the view that mathematical modeling helps improve the development of students' creativity (Wessels, 2014; Suh et al., 2017; Lu and Kaiser, 2022a). The modeling cycle, which is enriched by aspects of creativity proposed by Lu and Kaiser (2022a) (Figure 1), explains that students may generate a variety of new ideas during each step of mathematical modeling and may take different approaches to the task of solving problems, thus providing support for the development of the divergent thinking of students. Simultaneously, children exhibit a high level of natural enthusiasm for mathematics and are skilled at creativity, and so the open nature of mathematical modeling tasks can stimulate their creativity and choice (COMAP and SIAM, 2016). Therefore, teachers can use the uncertainty and openness of the mathematical modeling task in mathematics classrooms to guide and cultivate middle school students' divergent thinking, not merely by focusing on the knowledge and skills they need to solve practical problems but also by emphasizing and developing their creativity, which may be reflected in the modeling process.

4.2. Curiosity mediates the relationship between mathematical modeling competency and creativity

This study found that curiosity partially mediates the effect of the mathematical modeling competency of middle school students on creativity; accordingly, Research Hypothesis 2 was verified. Mathematical modeling plays a positive role in the development of individual curiosity. The stronger students' mathematical modeling competency is, the greater their curiosity is, which to some extent verifies the theory that curiosity originates from situation-based prediction errors and information-based prediction errors (Gruber and Ranganath, 2019). Based on Lu and Kaiser's (2022a) theoretical framework, the different perspectives and approaches used by students in solving mathematical modeling tasks may be the source of their curiosity. The mathematical modeling tasks used in this study are mainly based on personal and scientific situations, which are more realistic and feature more uncertainty than situations in which students solve problems in ordinary mathematics education. Errors trigger students' desire to explore different content in the situation at hand, thereby inspiring students' curiosity. Simultaneously, solving modeling tasks requires students to understand and collect relevant information, construct appropriate mathematical models independently, solve models, and explain and verify the solutions to realistic problems, a process which differs from the "standard process" used by students to solve conventional mathematical problems. Mathematical modeling tasks stimulate prior knowledge and information prediction errors between the process and reality, thereby enhancing students' creativity. For example, students collected different useful information to make different mathematical models in the example item (see Appendix A). Some of them chose "Use 1 liter of water every time" so that they make a more complete mathematical models and the others make the different mathematical models. The uncertainty of the information collection could promote the motivation of students to further explore different approaches to solve the problem.

On the other hand, curiosity has a positive impact on individual creativity. Middle school students with higher levels of curiosity tend to exhibit higher levels of creativity, a finding which supports the conclusions of previous studies (Celik et al., 2016; Hardy et al., 2017; Schutte and Malouff, 2020). According to Mumford and McIntosh (2017) creative thinking process model, problem definition and information collection are the initial steps in this process. On this basis, students are given the opportunity to understand the task situation and collect relevant data and information independently, which helps them participate in the process of creative thinking and enhance their creativity.

Based on the mediation path of "mathematical modeling competency-curiosity-creativity," teachers can focus on the following two strategies in their daily teaching using mathematical modeling. First, teachers should focus on guiding students to understand and explore realistic problems and situations and helping students consider the uncertainty involved in the task at hand. When middle school students face realistic problems, due to their limited knowledge of the real world, teachers must guide students to consider ways of understanding the problem situation and the possibilities that exist in this context. Teachers should help students understand the uncertainty involved in real situations as well as ways of using mathematics to solve real problems with the aim of effectively stimulating and cultivating students' curiosity and creativity. Second, when selecting and designing modeling tasks, teachers should control the difficulty and complexity of such tasks. The cognitive development of middle school students is not yet fully mature, and their mathematical knowledge and skills remain limited. Therefore, the realistic problems selected by teachers should enable students to understand the situation at hand and allow them to try to develop solutions; accordingly, these problems should not be too simple and routine to stimulate students' high-level cognitive processes, nor should they be too difficult and complex, thus causing students to lose their motivation and interest in inquiry.

4.3. Guided inquiry teaching moderates the relationship between curiosity and creativity

This study found that the influence of middle school students' curiosity on creativity is moderated by guided inquiry teaching; accordingly, Research Hypothesis 3 was validated. The impact of curiosity on creativity is higher in individuals who perceive a higher level of guided inquiry teaching than in individuals who perceive a lower level of guided inquiry teaching. In other words, guided inquiry-based teaching can promote the positive impact of middle school students' curiosity on their creativity. One possible explanation for this influence is related to the free exploration space provided by guided inquiry teaching; one characteristic of guided inquiry teaching is that after teachers provide students with tasks and the necessary explanations, students must independently design their own process of inquiry and approach to problem-solving (Blanchard et al., 2010; Chen et al., 2021), thus offering students the freedom to play and operate in this context. In a classroom in which teachers use more guided inquiry teaching, students may analyze and solve problems from additional perspectives, and they may have more opportunities to try multiple strategies and solutions; accordingly, their curiosity is more likely to promote the development of creativity (Zhao, 2018).

Based on the positive effect of guided inquiry teaching on the relationship between curiosity and creativity, teachers should consider the positive impact of teaching methods and classroom climate (Lucas and Spencer, 2017). In China, demonstration or lecturing are mostly applied by primary and middle school teachers in class, leading to students having few opportunities to think independently (Crehan, 2016). The results provide a suggestion that Chinese teachers should play guiding roles in students' learning processes and give students more freedom in inquiry. Teachers should avoid excessive participation that might reduce the effectiveness of students' creativity, thereby effectively promoting students' curiosity and creativity.

5. Limitations

This study faced certain limitations that should be addressed in future studies. First, this study adopted a cross-sectional research design, and in the future, it is necessary to investigate the causal relationship between mathematical modeling and creativity in further detail using experimental studies or longitudinal tracking studies. Second, mathematical modeling competency includes not only the sub-competencies of the modeling process but also metacognitive modeling competencies and other elements. In the future, researchers can use log data to measure metacognitive competency and incorporate it into mathematical modeling competency. Third, the influence of mathematical modeling competency on creativity *via* curiosity is only one of the possible paths associated with this relationship, and there may be other factors that mediate this relationship. In the future, additional empirical studies are necessary to verify the influence paths highlighted by this study, and more variables should be used to explore the possible factors mediating this relationship to provide more theoretical support for the cultivation of creativity. Moreover, students' self-reports may exhibit certain biases, and other, more objective methods should be used to measure relevant variables in the future.

6. Conclusion

In conclusion, this study expands our understanding of the relationship between mathematical modeling competency and creativity and explores the role of curiosity as a mediator and that of guided inquiry teaching as a moderator in this relationship. This model enriches the existing theories on the relationship between mathematical modeling competency and creativity and improves the theoretical basis for teachers to use mathematical modeling tasks and guided inquiry teaching to cultivate students' creativity. The results of this study were as follows. (1) Creativity can be influenced by middle school students' mathematical modeling competency. (2) Mathematical modeling can promote the creativity of middle school students by stimulating their curiosity. (3) Guided inquiry teaching can improve the impact of middle school students' curiosity on their creativity. Compared with low levels of guided inquiry teaching, high levels of guided inquiry teaching can improve the positive effect of curiosity on creativity.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) at the Collaborative Innovation Center of Assessment toward Basic Education Quality at Beijing Normal University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

TW: writing—original draft and review, writing—editing, and data—analysis. LZ: writing—original draft and review and data—analysis. ZX: writing—review and writing—editing. JL: writing—editing, supervision, and project administration. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Collaborative Innovation Center of Assessment Toward Basic Education Quality, Beijing Normal University [Grant number BJZK-2021A1-20016].

Acknowledgments

The data referenced by this research are drawn from the Regional Education Assessment Project of Collaborative Innovation Center of Assessment towards Basic Education Quality of Beijing Normal University.

Conflict of interest

The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1044580/full#supplementary-material

References

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., and Granger, E. M. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Sci. Educ.* 94, 577–616. doi: 10.1002/sce.20390

Blum, W., and Ferri, R. B. (2016). "Advancing the teaching of mathematical modeling: research-based concepts and examples" in *Mathematical Modeling and Modeling Mathematics*. eds. C. R. Hirsch and A. R. McDuffie (Reston, VA: NCTM)

Blum, W., Galbraith, P. L., and Niss, M. (2007). "Introduction" in *Modelling and Applications in Mathematics education*. eds. P. L. Galbraith, H. Henn and M. Niss (New York: Springer)

Brown, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit" in *Testing Structural Equation Models*. eds. K. A. Bollen and J. S. Long (Newbury Park, CA: Sage)

Cai, J. (2017). *Compendium for Research in Mathematics Education*. Reston, VA: NCTM.

Cai, J., and Xu, B. (2016). 也论数学核心素养及其构建. [on the core components of mathematical literacy.] global. *Education* 45, 3–12.

Camp, G. C. (1994). A longitudinal study of correlates of creativity. *Creat. Res. J.* 7, 125–144. doi: 10.1080/10400419409534519

Celik, P., Storme, M., Davila, A., and Myszkowski, N. (2016). Work-related curiosity positively predicts worker innovation. *J. Manag. Dev.* 35, 1184–1194. doi: 10.1108/JMD-01-2016-0013

Chen, C., Zheng, Y., and Shi, S. (2021). 不同探究式教学法对幼儿编程学习的影响. [the influence of different inquiry-based teaching methods on children's programming learning.] studies in early. *Child. Educ.* 3, 52–63. doi: 10.3969/j.issn.1007-8169.2021.03.006

Cheung, P. C., Lau, S., Chan, D. W., and Wu, W. Y. H. (2004). Creative potential of school children in Hong Kong: norms of the Wallach-Kogan creativity tests and their implications. *Creat. Res. J.* 16, 69–78. doi: 10.1207/s15326934crj1601_7

Choi, A., Klein, V., and Hershberger, S. (2015). Success, difficulty, and instructional strategy to enact an argument-based approach: experiences of elementary teachers. *Int. J. Sci. Math. Educ.* 13, 991–1011. doi: 10.1007/s10763-014-9525-1

COMAP and SIAM. (2016). *Guidelines for Assessment and Instruction in Mathematical Modelling Education*. Bedford, MA; Philadelphia, PA: COMAP; SIAM.

Common core state standards initiative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Crehan, L. (2016). *Cleverlands: The Secrets behind the Success of the World's Education Superpowers*. London: Unbound.

English, L. D. (2021). "Mathematical and interdisciplinary modeling in optimizing young Children's learning" in *Exploring Mathematical Modeling with Young Learners*. eds. J. M. Suh, M. H. Wickstrom and L. D. English (Cham: Springer)

Erbas, A. K., and Yenmez, A. A. (2011). The effect of inquiry-based explorations in a dynamic geometry environment on sixth grade students' achievements in polygons. *Comput. Educ.* 57, 2462–2475. doi: 10.1016/j.compedu.2011.07.002

Facione, P. A., Facione, N. C., and Giancarlo, C. A. F. (1994). The disposition toward critical thinking as a measure of competent clinical judgment: the development of the California critical thinking disposition inventory. *J. Nurs. Educ.* 33, 345–350. doi: 10.3928/0148-4834-19941001-05

Facione, P. A., Facione, N. C., and Giancarlo, C. A. F. (1998). *The California Critical Thinking Disposition Inventory Test Manual (revised)*. Millbrae, CA: California Academic Press.

Fletcher, N., Velamur, A., and Waid, B. E. (2013). *Mathematical Modeling Handbook II: the Assessments*, Bedford MA: COMAP.

Fredrickson, B. L., and Joiner, T. (2018). Reflections on positive emotions and upward spirals. *Perspect. Psychol. Sci.* 13, 194–199. doi: 10.1177/1745691617692106

Galbraith, P., and Holton, D. (2018). *Mathematical Modelling: a Guidebook for Teachers and Teams*. Camberwell: Australian Council for Educational Research.

Gan, Q., Bai, X., Liu, J., Wei, R., Ma, L., Xu, G., et al. (2020). 创新素养: 21世纪核心素养5C模型之三. [creativity competence: part III of the 5Cs framework for twenty-first century key competences.]. *J. East China Norm. Univ.* 38, 57–70. doi: 10.16382/j.cnki.1000-5560.2020.02.006

Geiger, V., Galbraith, P., Niss, M., and Delzoppo, C. (2022). Developing a task design and implementation framework for fostering mathematical modelling competencies. *Educ. Stud. Math.* 109, 313–336. doi: 10.1007/s10649-021-10039-y

Gottlieb, J., Oudeyer, P. Y., Lopes, M., and Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* 17, 585–593. doi: 10.1016/j.tics.2013.09.001

Gruber, M. J., and Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: the prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends Cogn. Sci.* 23, 1014–1025. doi: 10.1016/j.tics.2019.10.003

Hagtvedt, L. P., Dossinger, K., Harrison, S. H., and Huang, L. (2019). Curiosity made the cat more creative: specific curiosity as a driver of creativity. *Organ. Behav. Hum. Decis. Process.* 150, 1–13. doi: 10.1016/j.obhdp.2018.10.007

Hardy, J. H., Ness, A. M., and Mecca, J. (2017). Outside the box: epistemic curiosity as a predictor of creative problem solving and creative performance. *Personal. Individ. Differ.* 104, 230–237. doi: 10.1016/j.paid.2016.08.004

Hennessey, B., and Amabile, T. (2010). Creativity. *Annu. Rev. Psychol.* 61, 569–598. doi: 10.1146/annurev.psych.093008.100416

Hox, J. J., and Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Sci. Rev.* 10, 623–639. doi: 10.1080/01688638808402800

Hu, W. P., Wan, X. T., and Yu, L. (2011). 儿童青少年技术创造力的发展. [the development of technological creativity of children and adolescents.]. *Psychol. Res.* 4, 24–28.

Huang, J., Lu, X., Wang, Y., and Xu, B. (2019). 20 世纪以来中国数学课程标准中数学建模内涵的发展. [mathematical modelling in mathematics curriculum standards in China since the 20th Century.] journal of. *Math. Educ.* 28, 18–23.

Jiang, F., and McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: evidence from propensity score analysis of PISA data. *Int. J. Sci. Educ.* 37, 554–576. doi: 10.1080/09500693.2014.1000426

Kaiser, G. (2007). "Modelling and modelling competencies in school" in *Mathematical Modelling: Education, Engineering and Economics*. eds. C. Haines, P. Galbraith, W. Blum and S. Khan (West Sussex: Honvood Publishing Limited)

Kline, R.B. (2005). *Principles and Practices of Structural Equation Modeling*. New York: Guilford.

Leikin, R., and Elgrably, H. (2020). Problem posing through investigations for the development and evaluation of proof-related skills and creativity skills of prospective high school mathematics teachers. *Int. J. Educ. Res.* 102:101424. doi: 10.1016/j.ijer.2019.04.002

Litman, J. A. (2005). Curiosity and the pleasures of learning: wanting and liking new information. *Cognit. Emot.* 19, 793–814. doi: 10.1080/02699930541000101

Loewenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychol. Bull.* 116, 75–98. doi: 10.1037/0033-2909.116.1.75

Lu, X., and Kaiser, G. (2022a). Creativity in students' modelling competencies: conceptualisation and measurement. *Educ. Stud. Math.* 109, 287–311. doi: 10.1007/s10649-021-10055-y

Lu, X., and Kaiser, G. (2022b). Can mathematical modelling work as a creativity-demanding activity? An empirical study in China. ZDM–mathematics. *Education* 54, 67–81. doi: 10.1007/s11858-021-01316-4

Lucas, B., and Spencer, E. (2017). *Teaching Creative Thinking: Developing Learners who Generate Ideas and can Think Critically*. Wales: Crown House Publishing Limited.

Maaß, K. (2006). What are modelling competencies? *ZDM* 38, 113–142. doi: 10.1007/BF02655885

Malhotra, N. K., Kim, S. S., and Patil, A. (2006). Common method variance in IS research: a comparison of alternative approaches and a reanalysis of past research. *Manag. Sci.* 52, 1865–1883. doi: 10.1287/mnsc.1060.0597

Mei, Z. (2018). 美国M CM/ICM 特等奖论文对大学数学创新教育的启示. [enlightenment of MCM/ICM outstanding papers on innovative education of college mathematics.]. *J. Math. Educ.* 27, 10–13.

Minner, D. D., Levy, A. J., and Century, J. (2010). Inquiry-based science instruction: what is it and does it matter? Results from a research synthesis years 1984 to 2002. *J. Res. Sci. Teach.* 47, 474–496. doi: 10.1002/tea.20347

Mumford, M. D., and McIntosh, T. (2017). Creative thinking processes: the past and the future. *J. Creat. Behav.* 51, 317–322. doi: 10.1002/jocb.197

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

Niss, M., and Blum, W. (2020). *The Learning and Teaching of Mathematical Modelling*. New York: Routledge.

Niss, M., Blum, W., and Galbraith, P. (2007). "Introduction" in *Modelling and Applications in Mathematics Education: The 14th ICMI Study*. eds. W. Blum, P. L. Galbraith, H. W. Henn and M. Niss (Cham: Springer)

 10.3389/fpsyg.2022.1044580

O'Keefe, J., and Nadel, L. (1979). The cognitive map as a hippocampus. *Behav. Brain Sci.* 2, 520–533. doi: 10.1017/S0140525x00064256

OECD (2010). *The OECD Innovation Strategy: Getting a Head Start on Tomorrow*. Paris: OECD Publishing.

OECD (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.

OECD (2019). *PISA 2021 Creative Thinking Framework*. Paris: OECD Publishing.

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussion through professional development. *J. Res. Sci. Teach.* 47, 422–453. doi: 10.1002/tea.20345

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879

Rodríguez, G., Pérez, N., Núñez, G., Baños, J., and Carrió, M. (2019). Developing creative and research skills through an open and interprofessional inquiry-based learning course. *BMC Med. Educ.* 19:134. doi: 10.1186/s12909-019-1563-5

Schijndel, T. J. P., Jansen, B. R. J., and Jansen, M. E. J. (2018). Do individual differences in children's curiosity relate to their inquiry-based learning? *Int. J. Sci. Educ.* 40, 996–1015. doi: 10.1080/09500693.2018.1460772

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T., and Lee, Y. (2007). A meta-analysis of national research: effects of teaching strategies on student achievement in science in the United States. *J. Res. Sci. Teach.* 44, 1436–1460. doi: 10.1002/tea.20212

Schutte, N. S., and Malouff, J. M. (2020). A meta-analysis of the relationship between curiosity and creativity. *J. Creat. Behav.* 54, 940–947. doi: 10.1002/jocb.421

Smith, G. J., and Carlsson, I. M. (1990). *The Creative Process: A Functional Model based on Empirical Studies from Early Childhood to Middle Age*. Madison, CT: International Universities Press.

Suh, J. M., Matson, K., and Seshaiyer, P. (2017). Engaging elementary students in the creative process of mathematizing their world through mathematical modeling. *Educ. Sci.* 7:62. doi: 10.3390/educsci7020062

Tang, D., and Wen, Z. (2020). Statistical approaches for testing common method bias: problems and suggestions. *J. Psychol. Sci.* 43, 215–223. doi: 10.16719/j.cnki.1671-6981.20200130

The Ministry of Education of the People's Republic of China (2022). *Compulsory Mathematics Curriculum Standard for Compulsory Education* (*2022 Edn*). Beijing: Beijing Normal University Publishing Group.

Wessels, H. (2014). Levels of mathematical creativity in model-eliciting activities. *J. Math. Modell. Appl.* 1, 22–40.

Yang, K., Lin, S., Hong, Z., and Lin, H. (2016). Exploring the assessment of and relationship between elementary students' scientific creativity and science inquiry. *Creat. Res. J.* 28, 16–23. doi: 10.1080/10400419.2016.1125270

Zhao, L. (2018). 不同社会经济地位中小学生的创造力表现. [creativity performance of primary and secondary school students with different socioeconomic status.]. *J. Chin. Soc. Educ.* 10, 49–54.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fpsyg.2023.1012701

Check for updates

OPEN ACCESS

EDITED BY
Nelly Lagos San Martín,
University of the Bío Bío,
Chile

REVIEWED BY
María Cristina Richaud,
National Scientific and Technical Research
Council (CONICET), Argentina
Liu Shengnan,
East China Normal University,
China

*CORRESPONDENCE
Haitao Wang
✉ yshdwht@163.com
Xinlin Zhou
✉ zhou_xinlin@bnu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 05 August 2022
ACCEPTED 30 January 2023
PUBLISHED 16 February 2023

CITATION
Yu X, Wang X, Zheng H, Zhen X, Shao M, Wang H and Zhou X (2023) Academic achievement is more closely associated with student-peer relationships than with student-parent relationships or student-teacher relationships.
Front. Psychol. 14:1012701.
doi: 10.3389/fpsyg.2023.1012701

Academic achievement is more closely associated with student-peer relationships than with student-parent relationships or student-teacher relationships

Xiaodan Yu[1†], Xufei Wang[2†], Haoyue Zheng[1], Xin Zhen[1], Min Shao[1], Haitao Wang[1]* and Xinlin Zhou[2]*

[1]Department of Education, Ocean University of China, Qingdao, Shandong, China, [2]State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Siegler Center for Innovative Learning, Beijing Normal University, Beijing, China

Introduction: Personal relationships have long been a concern in education. Most studies indicate that good personal relationships are generally positively correlated with academic performance. However, few studies have compared how different types of personal relationships correlate with academic performance, and the conclusions of existing studies are inconsistent. Based on a large sample, the current study compared how the three closest types of personal relationships among students (with parents, teachers, and their peers) compared with their academic performance.

Methods: Cluster sampling was used to issue questionnaires to students in Qingdao City, Shandong Province, China in 2018 (Study 1) and in 2019 (Study 2). The actual sample size included 28168 students in Study 1 and 29869 students in Study 2 (both studies, Grades 4 and 8), thus totaling 58037 students. All students completed a personal relationship questionnaire and several academic tests.

Results: The results showed that: (1) the quality of personal relationships significantly and positively correlated with academic performance; (2) Among the three types of relationships tested, the quality of student-peer relationships was the most closely associated with academic achievement.

Discussion: This study gives insights into future research directions in this field and also reminds educators to pay attention to the personal relationships among their students, especially peer relationships.

KEYWORDS

personal relationships, student-parent relationships, student-teacher relationships, student-peer relationships, academic achievement

1. Introduction

Numerous studies have shown a significant positive correlation between personal relationships and academic performance (Wentzel, 1998; Kiuru et al., 2009; Martin and Dowson, 2009; Tobbell and O'Donnell, 2013; Castro et al., 2015). In these studies, student relationships have typically been divided into those with parents (student-parent), those with teachers (student-teacher), and those with peers (student-peer) because these relationships encompass the main scope and characteristics of their daily activities (Rubin et al., 2010; Kiuru et al., 2014; Moore et al., 2018). Studies have shown that each of these relationship categories is an important factor in student development (Walsh et al., 2010; Wang and Eccles, 2012; King and Ganotice, 2014; Collie et al., 2016). Only a few studies have compared how well different types of personal relationships correlate with student academic performance (Chen, 2005; Lam et al., 2012; Altermatt, 2019; Wentzel et al., 2016; Leung et al., 2021; Vargas-Madriz and Konishi, 2021), and the results were not consistent. Which type of personal relationships is most closely related to academic performance remains unclear. We classified personal relationships the same way in the current study. With a large sample size (58,037 primary and middle school students) and two experiments (Experiment 1 and Experiment 2), the current study focused on the associations between student academic performance and relationships with family, teachers, and peers.

1.1. The correlation between personal relationships and academic achievement

Most studies agree that there is a significant correlation between the quality of personal relationships and academic performance; positive relationships can predict good academic performance, while those ridden with conflict can predict poor academic performance. These studies focused on the effects of different types of personal relationships on academic performance.

First, considering the student-parent relationships, studies have found that parental support is a predictor of student achievement (Wentzel, 1998; Chiu, 2010; Castro et al., 2015; Igbo et al., 2015). For example, Igbo et al. (2015) used an Ex-Post Facto design to examined 48 public schools in the Otukpo Education Zone of Benue State and formulated and tested four null hypotheses with the t-test statistics at 0.05 level of significance. They found that student-parent relationships had a significant impact on math and English test scores. Castro et al. (2015) found that the frequency with which students interact with their family members had a positive impact on academic performance. Based on this, the authors concluded that parents influence students by shaping the positive value of education in academic-oriented behavior.

Second, the student-teacher relationships have also been shown impact academic performance (Roorda et al., 2011; Hughes et al., 2012; Ly et al., 2012; Hajovsky et al., 2020). High quality student-teacher relationships could contribute to students' cognitive skill development. For example, when Ly et al. (2012) characterized student-teacher relationships as either intimacy, conflict, or warmth, they found that the comprehensive warmth score (rated by the teachers) was significantly and positively correlated with children's math and reading performance. Hajovsky et al. (2020) studied two types of student-teacher relationships, intimacy and conflict, and found that while intimacy (as rated by the teachers) had only an indirect effect on math achievement, conflict affected math achievement both directly and indirectly.

Third, student-peer relationships have been shown to be significantly associated with student academic performance (Espelage et al., 2013; Li et al., 2020; Vignery and Laurier, 2020). For example, using both horizontal and longitudinal analysis, Espelage et al. (2013) showed that bullying and peer victimization were related to low academic performance. In the study of Gremmen et al. (2018), a longitudinal social network analysis (RSiena) showed that student academic engagement and achievement improved when their friends scored better, and vice versa, regardless of their physical position in the classroom. The selection effect (i.e., students for groups of friends that share the same levels of academic performance) and socialization process (i.e., peer performance is significantly associated with future student achievement), which were reported by Vignery and Laurier (2020), could be used to understand the links between student academic performance and peer relationships.

Despite this evidence, other studies have come to different conclusions. These have suggested that there is no significant correlation between personal relationships and academic achievement (Chen, 2008; Nokali et al., 2010; Barile et al., 2012; Hajovskya et al., 2017). For example, Chen (2008) found perceived peer support had no significant direct or indirect relationship with student achievement at any grade level, while perceived academic engagement was significantly and positively related with student achievement for adolescents at all grade levels. Nokali et al. (2010) found that improvements in parental involvement predicted declines in problematic behaviors and improvements in social skills, but did not predict changes in achievement. Barile et al. (2012) found that even if students have good relationships with their teachers, the relationship may not be enough to improve their math achievement, particularly in high school. Hajovskya et al. (2017) found that student-teacher conflict had a statistically significant effect on subsequent math performance ($\beta = 0.04$), but the size of the effect was not actually significant. The connection between personal relationships and academic achievement remains unclear and more research is needed to reach a definitive conclusion.

1.2. Comparative study of the association between different types of personal relationships and academic achievement in students

Determining which type of personal relationships is most closely related to academic performance will contribute to future theoretical and practical education planning. We found only a few studies that have compared how well different types of personal relationships correlate with student academic performance (Chen, 2005; Lam et al., 2012; Altermatt, 2019; Wentzel et al., 2016; Leung et al., 2021; Vargas-Madriz and Konishi, 2021). Importantly, among the existing research, results were not consistent.

Some studies state more specifically which type of personal relationships is important. For instance, Altermatt (2019) found that only perceived academic support from peers, neither parents nor professors, independently predicted academic self-efficacy. Gao and Xue (2020) found that parental participation had a greater impact on student academic performance than did peer influence, and was better at explaining differences in student academic performance. Vargas-Madriz and Konishi (2021) found that parental support showed a direct relationship with academic involvement, while the relationships between peer/teacher support and academic involvement was mediated by the students' sense of school belongingness.

Other studies found some personal relationships have less correlation with academic performance than other personal relationships. Some studies found that student-parent and student-peer relationships were more related to academic achievement than were student-teacher relationships. For instance, Leung et al. (2021) reported that while the quality of student-parent and student-peer relationships at Time 1 were significantly associated with academic achievement at Time 2 after controlling for Time 1 academic achievement, that of student-teacher relationships was not. In contrast, other studies found that student-parent and student-teacher relationships more related to academic achievement than were student-peer relationships. For instance, Wentzel et al. (2016) examined perceived emotional support and expectations from parents, teachers, and classmates in relation social behavior and academic functioning in Mexican American adolescents ($n = 398$). Results of a regression analyses indicated that teacher and parent variables were significant predictors of academic outcomes, while peer variables were significant predictors of social behavior. Lam et al. (2012) examined multiple-group structural equation modeling to revealed that, not perceptions of peer support, but teacher support and parent support were related indirectly to academic performance through student engagement.

Thus, the literature shows varying results regarding the correlation between different types of student relationships and student academic performance. There are several reasons that could explain why these results have been inconsistent.

First, the samples in these studies had differences in age, size, and social and cultural backgrounds. For example, consider the following two studies. In Altermatt (2019), the participants were 107 undergraduate students (79 women, 28 men), who were of traditional college age (mean = 20.14 years) and the majority (92%) identified as Caucasian, enrolled in introductory psychology courses at a liberal arts college in the Midwest United States. In contrast, Leung et al. (2021) conducted a longitudinal study in Hong Kong, China in which data were collected in two waves from a sample of 786 primary school students who had similar socioeconomic status and academic performance.

In addition, the disciplines chosen in the literature to test academic performance varied considerably. For instance, Teng et al. (2018) used final exam results in Chinese, mathematics, and English from the semester just before the experiment. The scores of each participant were standardized within every class, which yielded Z-scores for the three subjects. The sum of the three Z scores was used as an index of each student's academic achievement. In contrast, García Bacete et al. (2021) used student marks in mathematics and Spanish at the end-of-year exams for 1st, 2nd, 4th, and 6th grades, applying a 5-point scale (fail, pass, good, very good, and excellent).

Based on these factors that might influence the research results, the current study makes adjustments in a large sample, which can ameliorate the limitations of previous studies. We focused on comparing three student relationships using different disciplines in two studies. This method allows us to determine which relationship is most closely related to academic achievement.

1.3. How the association between personal relationships and academic achievement develops

Previous research has focused on whether the correlation between personal relationships and academic achievement varies by grade. Most researchers have found that the degree to which personal relationships affect academic achievement varies by grade. For example, Furman and Buhrmester (1992) observed age differences in perceptions of relationships with parents, grandparents, teachers, and siblings. Parents were seen as the most frequent providers of support in the fourth grade, same-sex friends were perceived to be as supportive as parents in the seventh grade, and were the most frequent providers of support in the tenth grade. Another good example of this can be found in Gallardo' research (2016). Their regression analysis indicated that peer acceptance positively predicted subsequent academic achievement and this relationship was also moderated by age, with the effect of peer acceptance on subsequent academic achievement being weaker during mid-adolescence than in early adolescence. This consistent result was also present in Wentzel' study. Wentzel et al. (2020) found an association between peer social acceptance and academic performance as twice as strong for students of primary education with respect to students of secondary education.

These findings indicate that various relationships are perceived as playing different roles at different points in development, which is consistent with physical and mental development in children. Due to the development of self-awareness and independence, children and teenagers show completely different personal communication characteristics; the focus of their personal relationships gradually shift to close friendships. Therefore, we have reason to believe that with this change in focus, the correlation between personal relationships and academic performance also changes.

1.4. Current study

Based on the existing theoretical and empirical results, this study used a big sample to focus on three different personal relationships among students and their academic performance. The data came from monitoring education quality in the Chinese city of Qingdao (Shandong Province) for two consecutive years, which includes a survey that assesses personal relationships in fourth and eighth graders. Part of the personal relationship scale was updated in the second year (see Measurements in Study 2 for details). As indexes of academic achievement, we chose math scores in the first year and scientific scores in the second year.

Social Impact Theory (SIT) provides theoretical support for our hypothesis. The SIT proposed by Latan (1981) suggests that an individual's feelings, attitudes, and behavior can be influenced by the presence of others. When some numbers of social sources are acting on a target individual, the amount of impact experienced by the target should be a multiplicative function of the strength, S, the immediacy, I, and the number, N, of sources present. The three aspects of SIT is to say that how important the group of social sources is to oneself (strength of influence), closeness to the group (in proximity), and size of the group (numbers) all combine to influence individuals. Pedersen et al. (2008) proved the principles of SIT may contribute to differences between assessments performed individually and those completed when surrounded by members of one's salient reference group. They examined 284 members of campus organizations in two contexts (online and group) to determine if individuals endorse different responses to questions of perceived and actual drinking norms across contexts. Results showed that all participants endorsed higher responses on questions of actual and perceived group behavior and of perceived group attitudes toward drinking during the group assessment than online assessment.

From this perspective, the characteristics of peer relationships are consistent with the important factors that influence individuals as emphasized in SIT theory. Usually, a class size is around 30 to 50 students. School-aged children spend most of their time in school and have more contact with their classmates and friends. The number of peers and communication time with peers are far greater than that of parents and teachers. A Chinese classic is consistent with this saying, "If you stay close to vermilion, you will be red, and if you stay close to ink, you will be black." (from Fu Xuan, Jin Dynasty, China), which addressed the importance of immediacy.

An equal relationship with peers can provide a sense of psychological stability and identity, as well as opportunities and places to use their own initiative, which parents and teachers cannot provide. Students can meet their self-development needs through mutual help and respect by sharing common feelings, conflicts, worries and difficulties with friends.

In addition, the process through which children acquire knowledge is not done alone, but requires communication and experience. With improved understanding of classroom teaching, cooperative learning has become a focus in education and many studies have confirmed the positive effects it has on different outcomes (Kyndt et al., 2013). It is expected that with the strong advocacy of cooperative learning in recent years, students are more inclined to cooperate with their peers to complete their learning tasks. Such a reform further promotes the degree of intimacy between peers and the importance of peers in student learning.

To sum up, we hypothesized that among the three typical personal relationships (student-parent, student-teacher, and student-peer), the student-peer relationships would have the highest correlation with academic achievement.

2. Study 1

2.1. Methods

2.1.1. Participants

With the help of the center of Assessment for Basic Education Quality in Qingdao City, Shandong Province, China, we used the cluster-sampling method to send questionnaires to each school. The data were collected in May 2018, and included 28,726 students in Grades 4 and 8 who came from 489 primary schools and 238 middle schools. Excluding 558 students who incorrectly filled out the forms (erroneous gender, grade level, or age), the actual sample size was 28,168. Parental consent was obtained prior to classroom-based testing.

After excluding invalid questionnaires, 17,112 participants (mean age = 9.80 years) remained in Grade 4, which included 8,847 (51.7%) boys and 8,265 (48.3%) girls, and 11,056 participants (mean age = 13.65 years) remained in Grade 8, which included 5,448 (49.28%) boys and 5,608 (50.72%) girls. Detailed demographic information is shown in Table 1.

2.1.2. Measurements

Personal relationships were measured using a revised version of the Personal Relationship Assessment Scale (PRAS) developed by the center of Qingdao Education Evaluation and Quality Monitoring Center, Ocean University of China. PRAS comprises 10 items that focus on perceived personal interaction of students with parents, teachers, and their peers (e.g., “When I’m in trouble, teacher help me in time”).Students responded to each item about their particular situations on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Scores were summed across the 10 questions to generate the final score with higher scores indicating higher quality of personal relationships students perceived. The scale has good internal consistency, with Cronbach’s alpha of 0.648 for Grade 4 and 0.763 for Grade 8.

The math achievement test was designed based on the mathematic curriculum standards of China, and included single choice, fill-in-the-blank, and essay problems that test math knowledge for Grades 4 and 8. For the sake of scientific research, we divided the test into six sets of test papers with different content and the same question type combination, which were randomly distributed to students in each grade. Test papers for each grade (6 sets for each grade) had good internal consistency. Cronbach’s alpha values for the 6 sets of Grade 4 and Grade 8 papers ranged from 0.720 to 0.808.

2.1.3. Procedure

After obtaining the informed consent of teachers and students, we began data collection in 2018, requiring all participants to complete the PRAS and a random math test corresponding to their grade. The math scores were graded by a quality inspection team of professionals.

TABLE 1 Sample size and mean age (years) by gender, grade level, and math test version.

	Grade 4			Grade 8	
	Number (Male, Female)	Mean age (Male, Female)		Number (Male, Female)	Mean age (Male, Female)
Test 1	2,859	9.79	Test 1	1840	13.66
	(1,458, 1,401)	(9.81, 9.77)		(947, 893)	(13.69, 13.64)
Test 2	2,852	9.79	Test 2	1847	13.64
	(1,429, 1,423)	(9.79, 9.80)		(901, 946)	(13.66, 13.61)
Test 3	2,856	9.79	Test 3	1832	13.66
	(1,463, 1,393)	(9.79, 9.80)		(906, 926)	(13.68, 13.63)
Test 4	2,854	9.79	Test 4	1843	13.66
	(1,488, 1,366)	(9.80, 9.78)		(884, 959)	(13.70, 13.62)
Test 5	2,852	9.80	Test 5	1850	13.66
	(1,488, 1,364)	(9.81, 9.80)		(891, 959)	(13.69, 13.63)
Test 6	2,839	9.81	Test 6	1844	13.64
	(1,521, 1,318)	(9.82, 9.79)		(919, 925)	(13.70, 13.59)
Total	17,112	9.80	Total	11,056	13.65
	(8,847, 8,265)	(9.80, 9.79)		(5,448, 5,608)	(13.69, 13.62)

We divided the math questions for each grade into six sets of tests in which the specific content differed, but the combination of questions was the same.

Moreover, the quality inspection team was also responsible for processing the questionnaire data. Questionnaire answers and math test results were returned to the researchers in excel format.

All tasks were conducted in a school environment and started at the same time. Relationships were measured online with no time limit, and all students received the same instructions. The math tests were completed in written form with a 60-min time limit.

2.1.4. Data analysis

Before analysis, we cleaned up the data using Microsoft Visual FoxPro 9.0. This was mainly to replace variable names, delete outliers, null values, and add dimensions. The purpose was to ensure that the data being analyzed tended to be true and accurate.

The data analyses were conducted with IBM SPSS Statistics for Windows, Version 26.0. Reliability and validity analysis, descriptive analysis, and inferential analyses were performed to examine the distribution of all variables and the associations between the main variables and other variables. The inferential analysis used Pearson correlation analysis and repeated-measures one-way analysis of variance (ANOVA) to explore the association between personal relationships and math achievement, as well as gender and age differences among them.

2.2. Results

2.2.1. Mean values and gender differences in personal relationships

Table 2 shows mean values and gender differences in overall personal relationships and the three subdivisions. There is a significant difference in gender for students' personal relationships that PRAS scores were higher for girls than for boys in both Grade 4 [girls vs. boys: 4.34 vs. 4.14, $t(df) = -19.341$ (17098.681), $p < 0.001$] and Grade 8 [4.41

TABLE 2 Mean (SD) values and independent sample T test results for gender in personal relationships.

Relationship		Grade 4				Grade 8			
		Boys	Girls	*t*	*p*	Boys	Girls	*t*	*p*
		M(SD)	*M(SD)*			*M(SD)*	*M(SD)*		
Student-parent	Test 1	4.05 (0.74)	4.16 (0.71)	−4.010	<0.001	4.13 (0.79)	4.23 (0.75)	−2.837	0.005
	Test 2	4.03 (0.70)	4.14 (0.70)	−4.171	<0.001	4.19 (0.76)	4.25 (0.78)	−1.509	0.131
	Test 3	4.06 (0.73)	4.15 (0.68)	−3.262	0.001	4.10 (0.79)	4.21 (0.74)	−3.141	0.002
	Test 4	4.01 (0.72)	4.17 (0.67)	−6.322	<0.001	4.12 (0.82)	4.25 (0.77)	−3.348	0.001
	Test 5	4.04 (0.70)	4.17 (0.69)	−5.177	<0.001	4.19 (0.75)	4.20 (0.74)	−0.373	0.709
	Test 6	4.04 (0.74)	4.16 (0.70)	−4.325	<0.001	4.15 (0.81)	4.24 (0.75)	−2.658	0.008
	Total	4.04 (0.72)	4.16 (0.69)	−11.111	<0.001	4.15 (0.79)	4.23 (0.75)	−5.711	<0.001
Student-teacher	Test 1	4.58 (0.81)	4.69 (0.69)	−3.686	<0.001	4.51 (0.89)	4.55 (0.80)	−1.201	0.230
	Test 2	4.57 (0.82)	4.69 (0.70)	−4.263	<0.001	4.55 (0.85)	4.56 (0.82)	−0.337	0.736
	Test 3	4.58 (0.79)	4.68 (0.70)	−3.358	0.001	4.47 (0.86)	4.55 (0.75)	−2.184	0.029
	Test 4	4.57 (0.82)	4.67 (0.70)	−3.794	<0.001	4.47 (0.91)	4.59 (0.77)	−3.013	0.003
	Test 5	4.55 (0.83)	4.68 (0.71)	−4.364	<0.001	4.53 (0.83)	4.52 (0.78)	0.227	0.821
	Test 6	4.58 (0.83)	4.67 (0.70)	−3.324	0.001	4.52 (0.88)	4.54 (0.78)	−0.499	0.618
	Total	4.57 (0.82)	4.68 (0.70)	−9.322	<0.001	4.51 (0.87)	4.55 (0.79)	−2.878	0.004
Student-peer	Test 1	3.71 (1.09)	4.06 (1.01)	−8.811	<0.001	4.01 (0.99)	4.35 (0.83)	−7.851	<0.001
	Test 2	3.62 (1.12)	4.04 (1.02)	−10.379	<0.001	4.05 (0.99)	4.40 (0.81)	−8.308	<0.001
	Test 3	3.69 (1.11)	4.08 (0.99)	−9.888	<0.001	3.97 (0.98)	4.32 (0.85)	−8.285	<0.001
	Test 4	3.64 (1.10)	4.01 (1.03)	−9.298	<0.001	4.01 (0.99)	4.35 (0.83)	−7.881	<0.001
	Test 5	3.66 (1.11)	4.03 (1.03)	−9.122	<0.001	4.07 (0.95)	4.31 (0.85)	−5.673	<0.001
	Test 6	3.73 (1.10)	4.08 (1.02)	−8.737	<0.001	4.00 (0.99)	4.37 (0.80)	−8.954	<0.001
	Total	3.68 (1.11)	4.05 (1.02)	−22.937	<0.001	4.02 (0.98)	4.35 (0.83)	−19.179	<0.001
Total	Test 1	4.16 (0.72)	4.33 (0.67)	−6.588	<0.001	4.25 (0.77)	4.41 (0.69)	−4.724	<0.001
	Test 2	4.12 (0.71)	4.33 (0.67)	−8.300	<0.001	4.31 (0.73)	4.45 (0.71)	−4.082	<0.001
	Test 3	4.15 (0.70)	4.35 (0.68)	−7.521	<0.001	4.20 (0.77)	4.39 (0.69)	−5.737	<0.001
	Test 4	4.12 (0.70)	4.33 (0.68)	−8.167	<0.001	4.24 (0.76)	4.45 (0.68)	−6.273	<0.001
	Test 5	4.12 (0.71)	4.34 (0.67)	−8.586	<0.001	4.31 (0.71)	4.39 (0.70)	−2.412	0.016
	Test 6	4.15 (0.71)	4.36 (0.68)	−8.240	<0.001	4.26 (0.75)	4.40 (0.69)	−4.360	<0.001
	Total	4.14 (0.71)	4.34 (0.68)	−19.341	<0.001	4.26 (0.75)	4.41 (0.70)	−11.298	<0.001

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

TABLE 3 Correlations between personal relationships and math scores.

Relationship		Grade 4			Grade 8		
		Boys	Girls	Total	Boys	Girls	Total
Student-parent	Test 1	0.134**	0.254**	0.187**	0.191**	0.156**	0.176**
	Test 2	0.218**	0.221**	0.220**	0.147**	0.112**	0.129**
	Test 3	0.189**	0.209**	0.199**	0.197**	0.104**	0.153**
	Test 4	0.178**	0.220**	0.195**	0.097**	0.078*	0.090**
	Test 5	0.221**	0.153**	0.188**	0.149**	0.128**	0.138**
	Test 6	0.192**	0.196**	0.190**	0.171**	0.150**	0.161**
	Total	0.182**	0.189**	0.185**	0.154**	0.114**	0.136**
Student-teacher	Test 1	0.145**	0.160**	0.149**	0.170**	0.127**	0.152**
	Test 2	0.171**	0.215**	0.191**	0.130**	0.136**	0.133**
	Test 3	0.135**	0.172**	0.153**	0.142**	0.143**	0.142**
	Test 4	0.112**	0.133**	0.120**	0.102**	0.073*	0.091**
	Test 5	0.188**	0.124**	0.160**	0.158**	0.165**	0.161**
	Test 6	0.145**	0.158**	0.148**	0.135**	0.126**	0.130**
	Total	0.141**	0.151**	0.145**	0.135*	0.122**	0.129**
Student-peer	Test 1	0.206**	0.309**	0.245**	0.197**	0.187**	0.194**
	Test 2	0.237**	0.262**	0.247**	0.223**	0.199**	0.211**
	Test 3	0.278**	0.244**	0.265**	0.253**	0.199**	0.225**
	Test 4	0.247**	0.279**	0.257**	0.209**	0.164**	0.191**
	Test 5	0.301**	0.292**	0.291**	0.224**	0.240**	0.229**
	Test 6	0.261**	0.221**	0.234**	0.198**	0.206**	0.197**
	Total	0.239**	0.250**	0.241**	0.212**	0.188**	0.200**
Totals	Test 1	0.215**	0.329**	0.262**	0.230**	0.180**	0.209**
	Test 2	0.271**	0.296**	0.282**	0.200**	0.196**	0.199**
	Test 3	0.252**	0.258**	0.256**	0.234**	0.170**	0.203**
	Test 4	0.233**	0.263**	0.243**	0.165**	0.113**	0.144**
	Test 5	0.300**	0.254**	0.275**	0.216**	0.208**	0.212**
	Test 6	0.267**	0.252**	0.252**	0.223**	0.196**	0.209**
	Total	0.242**	0.256**	0.246**	0.208**	0.170**	0.191**

*$p < 0.05$, **$p < 0.01$.

vs. 4.26, $t(df) = -11.298$ (10936.739), $p < 0.001$]. Looking at individual relationship types, the same pattern held for the student-parent [Grade 4: 4.16 vs. 4.04, $t(df) = -11.111$ (17095.336), $p < 0.001$, Grade 8: 4.23 vs. 4.15, $t(df) = -5.711$ (11054), $p < 0.001$]; student-teacher, Grade 4: 4.68 vs. 4.57, $t(df) = -9.322$ (16976.821), $p < 0.001$, Grade 8: 4.55 vs. 4.51, $t(df) = -2.878$ (10860.621), $p < 0.01$ and student-peer, Grade 4: 4.05 vs. 3.68, $t(df) = -22.937$ (17106.491), $p < 0.001$, Grade 8: 4.35 vs. 4.02, $t(df) = -19.179$ (10637.617), $p < 0.001$.

2.2.2. The correlation analysis between personal relationships and math achievement

PRAS scores correlated significantly and positively with math achievement in Grade 4 ($r = 0.246$, $p < 0.01$) and Grade 8 ($r = 0.191$, $p < 0.01$). To further explore this relationship, we analyzed the data according to the test version, gender, and grade. We found that the correlation between personal relationships and math achievement was roughly consistent and stable. See Table 3 for details.

Student-parent relationships. For boys in Grade 4, the correlation coefficient between the student-parent PRAS score and math achievement overall was 0.182 ($p < 0.01$). When looking each math test version separately, the correlation coefficients ranged from 0.134 to 0.221 (all $p < 0.01$). For girls in Grade 4, the correlation coefficient for the same relationship was 0.189 overall ($p < 0.01$), and ranged from 0.153 to 0.254 (all $p < 0.01$) for each math test individually. Looking at the same relationship in Grade 8, the correlation coefficient was 0.154 for boys ($p < 0.01$; ranging from 0.097 to 0.197, all $p < 0.01$) and 0.114 for girls ($p < 0.01$; ranging from 0.078 to 0.156, $p < 0.05$).

In Grade 4, the correlation coefficient between student-teacher PRAS score and math achievement was 0.141 for boys ($p < 0.01$; ranging from 0.112 to 0.188 for the six math tests, all $p < 0.01$) and 0.151 for girls ($p < 0.01$; ranging from 0.124 to 0.215, $p < 0.01$). In Grade 8, the correlation coefficient was 0.135 for boys ($p < 0.05$; ranging from 0.102 to 0.170, all $p < 0.01$) and 0.122 for girls ($p < 0.01$; ranging from 0.073 to 0.165, all $p < 0.05$).

In Grade 4, the correlation coefficient between student-peer PRAS score and math achievement was 0.239 for boys ($p < 0.01$; ranging from 0.206 to 0.301, all $p < 0.01$) and 0.250 for girls ($p < 0.01$; ranging from 0.221 to 0.309, all $p < 0.01$). In Grade 8, the correlation coefficient was 0.212 for boys ($p < 0.01$; ranging from 0.197 to 0.253, all $p < 0.01$) and 0.188 for girls ($p < 0.01$; ranging from 0.164 to 0.240, all $p < 0.01$).

In Grade 4, the correlation coefficient between the overall PRAS score and math achievement was 0.242 for boys ($p < 0.01$; ranging from 0.215 to 0.300, all $p < 0.01$) and 0.256 for girls ($p < 0.01$; ranging from 0.252 to 0.329, all $p < 0.01$). In Grade 8, it was 0.208 for boys ($p < 0.01$; ranging from 0.165 to 0.234, all $p < 0.01$) and 0.170 for girls ($p < 0.01$; ranging from 0.113 to 0.208, all $p < 0.01$).

Based on these results, we analyzed the data using a repeated-measures General Linear Model. The in-between factors were the correlation coefficients between the three types of student relationships and math achievement. The factors were Grade (4 or 8) and gender, and the Sphericity Assumption method was adopted to obtain the results of the intra-group Factor test (see Table 4). We found significant differences between the correlations ($p < 0.001$). Based on Table 3, we can see that among the three relationship types, the student-peer relationships had the closest association with math achievement.

2.2.3. Multivariate ANOVA

Multi-factor ANOVAs were used to investigate gender and grade differences in the correlation between student personal relationships and math achievement. We found no gender difference for any of the three relationship types (all $p > 0.05$, Bonferroni-corrected). In contrast, we did find a difference in grade for the student-parent [$F(1, 20) = 19.134$, $p < 0.001$, Bonferroni-corrected, η^2 $p = 0.489$] and student-peer [$F(1, 20) = 21.236$, $p < 0.001$, Bonferroni-corrected, η^2 $p = 0.515$] relationships. Combined with Tables 3, 5, we conclude that the correlation between the quality of the student-peer relationships and math achievement was stronger in Grade 4 than in Grade 8, and it was same between student-parent relationships and math achievement. See Table 5 for details.

2.3. Discussion

The results of Study 1 supported our hypothesis that student-peer relationships have the closest association with academic achievement (as indexed by math achievement) among the three types of student relationships.

The correlation analysis showed that personal relationships significantly and positively correlate with math performance, which is consistent with numerous other studies (Dhingra and Manhas, 2009; Lee, 2012; Oberle and Schonert-Reichl, 2013). Further analysis showed that, among the three types of student relationships, the student-peer relationships had the closest association with math scores in both Grade 4 and Grade 8, among boys and girls. To the best of our knowledge, this study is the first evidence showing that the student-peer relationships have a closer association with academic achievement than do student-parent or student-teacher relationships.

The ANOVA results showed the correlation between student-peer relationships quality and math performance depended on grade: the correlation became less as students got older. In terms of changes in peer relationships themselves, it has been confirmed that increased cognitive ability may encourage older students to seek independence more than younger students do (Furman and Buhrmester, 1992; Fuligni Andrew and Eccles Jacquelynne, 1993). Therefore, we speculate that eighth grade students may be more mature in thinking and more independent in their learning compared with fourth grade students, which could explain the grade difference.

Further inspiration comes from Chen (2005). Chen tested a hypothesized model that students' self-perceived academic support (from parents, teachers, and peers) is directly or indirectly related to their achievement through their own perceived academic engagement. In this process, the author found that personal relationships have different effects on different academic subjects: student-teacher relationships had the greatest impact on English and math, while student-peer relationships had the least impact. The student-teacher relationships had the greatest influence on Chinese, and the student-parent relationships had the least influence. These findings are not consistent with our results from Study 1, and we are curious whether differences in discipline is the key factor leading to different results. Therefore, in the second year, we focused on the connection between personal relationships and science achievement to test our hypothesis that among the three typical personal relationships (student-parent, student-teacher, and student-peer), the student-peer relationships would have the highest correlation with academic achievement.

TABLE 4 Results of intra-group factor test.

Source	SS	df	MS	F	P
Personal relationships	0.105	2	0.052	95.386	<0.001
Gender	0.000	2	0.000	0.208	0.813
Grade	0.005	2	0.003	4.558	0.016
Gender × Grade	0.001	2	0.001	0.924	0.405
Error (personal relationships)	0.022	40	0.001		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Bonferroni-corrected.

3. Study 2

3.1. Method

3.1.1. Participants

We used the same methods to collect student data as in Study 1. The data were collected in May 2019, and included 30,596 students in Grade 4 and Grade 8 who came from 545 primary schools and 241 middle schools. Excluding 727 students who incorrectly filled out the forms (erroneous gender, grade level, or age), the actual sample size was 29,869. After excluding invalid questionnaires, there were 17,752 participants (mean age = 9.80 years) in Grade 4, which included 9,210 (51.9%) boys and 8,542 (48.1%) girls, and 12,117 participants (mean age = 13.75 years) in Grade 8, which included 6,283 (51.9%) boys and 5,834 (48.1%) girls. Parental consent was obtained prior to classroom-based testing. Detailed demographic information is shown in Table 6.

3.1.2. Measurements

Personal relationships were measured using PRAS (2019), an updated version of PRAS (2018) that still focus on perceived personal interaction of students with parents, teachers, and their peers (e.g., "I have good times with my classmates"). Since large-scale educational

TABLE 5 ANOVA for gender and grade in relation to personal relationships and achievement.

	Student-parent relationship				Student-teacher relationship				Student-peer relationship			
	F	*df*	*p*	η^2 *p*	*F*	*df*	*p*	η^2 *p*	*F*	*df*	*p*	η^2 *p*
Gender	0.408	1	0.530	0.020	0.000	1	0.994	0.000	0.053	1	0.820	0.003
Grade	19.134	1	<0.001	0.489	3.225	1	0.088	0.139	21.236	1	<0.001	0.515
Gender × Grade	4.582	1	0.045	0.186	0.905	1	0.353	0.043	1.805	1	0.194	0.083
Error		20				20				20		

**p < 0.05, **p < 0.01, ***p < 0.001, Bonferroni-corrected.*

TABLE 6 Sample size and mean age (years) by gender, grade level, and science test version.

	Grade 4			Grade 8	
	Number (Male, Female)	Mean age (Male, Female)		Number (Male, Female)	Mean age (Male, Female)
Test 1	4,482	9.81	Test 1	1,836	13.74
	(2,320, 2,162)	(9.82, 9.79)		(945, 891)	(13.74, 13.74)
Test 2	4,460	9.79	Test 2	2,065	13.75
	(2,328, 2,132)	(9.79, 9.80)		(1,071, 994)	(13.75, 13.75)
Test 3	4,418	9.80	Test 3	2,059	13.76
	(2,330, 2088)	(9.80, 9.80)		(1,033, 1,026)	(13.79, 13.72)
Test 4	4,392	9.80	Test 4	2,056	13.76
	(2,232, 2,160)	(9.81, 9.78)		(1,069, 987)	(13.80, 13.72)
			Test 5	2,052	13.75
				(1,076, 976)	(13.77, 13.73)
			Test 6	2,049	13.74
				(1,089, 960)	(13.76, 13.73)
Total	17,752	9.80	Total	12,117	13.75
	(9,210, 8,542)	(9.80, 9.79)		(6,283, 5,834)	(13.77, 13.73)

We divided the science questions for Grade 4 into four sets, and the science knowledge for Grade 8 into six sets, which the specific content differed, but the combination of questions was the same.

quality monitoring generally requires a certain amount of annual updates, PRAS (2019) has been revised based on last year's version: the student-parent section was revised from four items, adding reverse scoring questions to eight items; a reverse scoring question was added to the student-teacher section and a forward scoring question was added to the student-peer section. The scoring rules remain the same. Students responded to each of 17 items about their particular situations. The scale has good internal consistency, with Cronbach's alphas of 0.869 for Grade 4 and 0.916 for Grade 8.

The Science achievement test was designed based on the science curriculum standards of China, which include single choice questions and non-choice questions and content about science knowledge in Grade 4 and Grade 8. For the experiment, we divided the Grade 4 test into four sets and Grade 8 test into six sets, each with different content but with the same combination of questions, which were randomly distributed to students in each grade. Test papers for each grade had good internal consistency, with Cronbach's alpha values for the 4 sets of test papers in Grade 4 and Grade 8 ranging from 0.723 to 0.922.

3.1.3. Procedure

After obtaining the informed consent of teachers and students, we began data collection in 2019, requiring all participants to complete the PRAS and a random science test corresponding to their grade. The science scores were graded by a quality inspection team of professionals. All other aspects of the procedure were the same as in Study 1.

3.1.4. Data analysis

Data analysis was conducted in the same manner as in Study 1.

3.2. Results

3.2.1. Mean values and gender differences in personal relationships

Table 7 shows mean values and gender differences in overall personal relationships and the three subdivisions. There is a significant difference in gender for students' personal relationships that PRAS scores were higher for girls than for boys in both Grade 4 [girls vs. boys: 4.40 vs. 4.25, $t(df) = 12.940$ (17747.529), $p < 0.001$] and Grade 8 [4.16 vs. 4.07, $t(df) = 5.854$ (12115), $p < 0.001$]. Looking at individual relationship types, the same pattern held for the student-parent, Grade 4: 4.31 vs. 4.20, $t(df) = 7.817$ (17629.440), $p < 0.001$, Grade 8: 3.95 vs. 3.92, $t(df) = 1.548$ (12115), $p = 0.122$; student-teacher, Grade 4: 4.65 vs. 4.50, $t(df) = 12.694$ (17545.281), $p < 0.001$, Grade 8: 4.38 vs. 4.28, $t(df) = 5.827$ (12107.187), $p < 0.001$ and student-peer, Grade 4: 4.39 vs. 4.21,

TABLE 7 Mean (SD) values and independent sample T test results for gender in personal relationships.

Relationship		Grade 4					Grade 8			
		Boys	Girls	*t*	*p*		Boys	Girls	*t*	*p*
		M(SD)	*M(SD)*				*M(SD)*	*M(SD)*		
Student-parent	Test 1	4.20 (0.86)	4.32 (0.86)	4.663	<0.001	Test 1	3.93 (1.02)	3.99 (0.97)	1.282	0.200
	Test 2	4.25 (0.84)	4.30 (0.86)	2.097	0.036	Test 2	3.90 (1.00)	4.00 (1.00)	2.235	0.026
	Test 3	4.18 (0.86)	4.31 (0.84)	5.237	<0.001	Test 3	3.88 (1.01)	3.93 (1.01)	1.074	0.283
	Test 4	4.19 (0.86)	4.29 (0.88)	3.649	<0.001	Test 4	3.91 (1.02)	3.90 (1.02)	−0.225	0.822
						Test 5	3.92 (1.01)	3.92 (1.03)	0.048	0.962
						Test 6	3.96 (0.97)	3.94 (1.04)	−0.516	0.606
Student-teacher	Total	4.20 (0.86)	4.31 (0.86)	7.817	<0.001	Total	3.92 (1.00)	3.95 (1.01)	1.548	0.122
	Test 1	4.50 (0.84)	4.65 (0.69)	6.67	<0.001	Test 1	4.25 (0.96)	4.36 (0.90)	2.582	0.010
	Test 2	4.52 (0.82)	4.65 (0.70)	5.594	<0.001	Test 2	4.28 (0.96)	4.40 (0.87)	3.206	0.001
	Test 3	4.48 (0.85)	4.65 (0.68)	7.472	<0.001	Test 3	4.29 (0.95)	4.33 (0.90)	1.048	0.295
	Test 4	4.51 (0.82)	4.64 (0.70)	5.638	<0.001	Test 4	4.30 (0.94)	4.36 (0.87)	1.580	0.114
						Test 5	4.28 (0.96)	4.39 (0.84)	2.959	0.003
						Test 6	4.31 (0.96)	4.43 (0.82)	3.060	0.002
Student-peer	Total	4.50 (0.83)	4.65 (0.69)	12.694	<0.001	Total	4.28 (0.96)	4.38 (0.87)	5.827	<0.001
	Test 1	4.21 (0.86)	4.41 (0.80)	7.989	<0.001	Test 1	4.25 (0.85)	4.44 (0.74)	5.070	<0.001
	Test 2	4.26 (0.84)	4.37 (0.82)	4.707	<0.001	Test 2	4.28 (0.82)	4.46 (0.75)	5.137	<0.001
	Test 3	4.19 (0.86)	4.38 (0.80)	7.571	<0.001	Test 3	4.27 (0.84)	4.40 (0.78)	3.448	0.001
	Test 4	4.20 (0.88)	4.38 (0.79)	7.367	<0.001	Test 4	4.31 (0.83)	4.41 (0.76)	2.700	0.007
						Test 5	4.28 (0.82)	4.42 (0.78)	4.093	<0.001
						Test 6	4.31 (0.82)	4.44 (0.75)	3.797	<0.001
Total	Total	4.21 (0.86)	4.39 (0.80)	13.822	<0.001	Total	4.28 (0.83)	4.43 (0.76)	9.844	<0.001
	Test 1	4.25 (0.78)	4.42 (0.73)	7.294	<0.001	Test 1	4.05 (0.88)	4.19 (0.81)	3.395	0.001
	Test 2	4.29 (0.77)	4.40 (0.74)	4.848	<0.001	Test 2	4.06 (0.84)	4.19 (0.83)	3.633	<0.001
	Test 3	4.22 (0.79)	4.40 (0.73)	7.783	<0.001	Test 3	4.05 (0.85)	4.12 (0.84)	2.080	0.038
	Test 4	4.24 (0.79)	4.38 (0.75)	5.947	<0.001	Test 4	4.08 (0.87)	4.13 (0.83)	1.235	0.217
						Test 5	4.06 (0.86)	4.15 (0.84)	2.332	0.020
						Test 6	4.10 (0.83)	4.16 (0.81)	1.787	0.074
	Total	4.25 (0.78)	4.40 (0.74)	12.940	<0.001	Total	4.07 (0.85)	4.16 (0.83)	5.854	<0.001

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

$t(df) = 13.822$ (17749.549), $p < 0.001$, Grade 8: 4.43 vs. 4.28, $t(df) = 9.844$ (12113.495), $p < 0.001$.

3.2.2. The correlation analysis between personal relationships and science achievement

PRAS scores were significantly and positively correlated with science achievement in Grade 4 ($r = 0.180$, $p < 0.01$) and Grade 8 ($r = 0.159$, $p < 0.01$). To further explore the correlation between student personal relationships and science achievement, we analyzed the data according to the test version, gender, and grade. We found that the correlation was roughly consistent and stable across these dimensions. See Table 8 for details.

For boys in Grade 4, the correlation coefficient between the student-parent PRAS score and science achievement was 0.149 ($p < 0.01$) and those for each version of the science test ranged from 0.118 to 0.216 (all $p < 0.01$). For girls in Grade 4, the correlation coefficient was 0.110 ($p < 0.01$; ranging from 0.103 to 0.124 for each test version, all $p < 0.01$). In Grade 8, the correlation coefficient was 0.095 for boys ($p < 0.01$; ranging from 0.071 to 0.164, all $p < 0.01$) and 0.111 for girls ($p < 0.01$; ranging from 0.096 to 0.176, all $p < 0.01$).

In Grade 4, the correlation coefficient between student-teacher PRAS score and science achievement was 0.172 for boys ($p < 0.01$; ranging from 0.154 to 0.197, all $p < 0.01$) and 0.149 for girls ($p < 0.01$; ranging from 0.126 to 0.189, all $p < 0.01$). In Grade 8, it was 0.180 for boys ($p < 0.01$; ranging from 0.173 to 0.207, all $p < 0.01$) and 0.155 for girls ($p < 0.01$; ranging from 0.113 to 0.212, all $p < 0.01$).

In Grade 4, the correlation coefficient between student-peer PRAS scores and science achievement was 0.221 for boys ($p < 0.01$; ranging from 0.198 to 0.262, all $p < 0.01$) and 0.199 for girls ($p < 0.01$; ranging from 0.184 to 0.233, all $p < 0.01$). In Grade 8, it was 0.216 for boys ($p < 0.01$; ranging from 0.195 to 0.255, all $p < 0.01$) and 0.168 for girls ($p < 0.01$; ranging from 0.164 to 0.220, all $p < 0.01$).

TABLE 8 Correlations between personal relationships and science scores.

Relationship	Grade 4				Grade 8			
		Boys	Girls	Total		Boys	Girls	Total
Student-parent	Test 1	0.216**	0.103**	0.163**	Test 1	0.164**	0.164**	0.164**
	Test 2	0.161**	0.108**	0.135**	Test 2	0.095**	0.176**	0.133**
	Test 3	0.120**	0.119**	0.116**	Test 3	0.100**	0.105**	0.103**
	Test 4	0.118**	0.124**	0.117**	Test 4	0.084**	0.096**	0.089**
					Test 5	0.071*	0.107**	0.087**
					Test 6	0.090**	0.127**	0.106**
	Total	0.149**	0.110**	0.128**	Total	0.095**	0.111**	0.102**
Student-teacher	Test 1	0.197**	0.156**	0.177**	Test 1	0.203**	0.230**	0.215**
	Test 2	0.197**	0.189**	0.190**	Test 2	0.177**	0.163**	0.173**
	Test 3	0.154**	0.126**	0.136**	Test 3	0.173**	0.212**	0.191**
	Test 4	0.178**	0.144**	0.157**	Test 4	0.184**	0.139**	0.165**
					Test 5	0.188**	0.193**	0.189**
					Test 6	0.207**	0.113**	0.170**
	Total	0.172**	0.149**	0.158**	Total	0.180**	0.155**	0.170**
Student-peer	Test 1	0.262**	0.197**	0.230**	Test 1	0.217**	0.211**	0.213**
	Test 2	0.240**	0.226**	0.231**	Test 2	0.246**	0.220**	0.238**
	Test 3	0.198**	0.233**	0.207**	Test 3	0.195**	0.191**	0.196**
	Test 4	0.216**	0.184**	0.194**	Test 4	0.234**	0.166**	0.206**
					Test 5	0.208**	0.191**	0.198**
					Test 6	0.255**	0.164**	0.218**
	Total	0.221**	0.199**	0.206**	Total	0.216**	0.168**	0.196**
Total	Test 1	0.257**	0.176**	0.218**	Test 1	0.205**	0.209**	0.206**
	Test 2	0.223**	0.189**	0.204**	Test 2	0.182**	0.187**	0.187**
	Test 3	0.179**	0.167**	0.167**	Test 3	0.185**	0.162**	0.176**
	Test 4	0.176**	0.164**	0.164**	Test 4	0.165**	0.148**	0.158**
					Test 5	0.138**	0.160**	0.147**
					Test 6	0.176**	0.149**	0.164**
	Total	0.200**	0.167**	0.180**	Total	0.167**	0.148**	0.159**

*$p < 0.05$, **$p < 0.01$.

TABLE 9 Results of the intra-group factor test.

Source	SS	df	MS	F	P
Personal relationships	0.077	2	0.039	72.767	<0.001
Gender	0.001	2	0.001	1.075	0.353
Grade	0.003	2	0.001	2.692	0.083
Gender × Grade	0.004	2	0.002	4.193	0.024
Error(personal relationships)	0.017	32	0.001		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Bonferroni-corrected.

In Grade 4, the correlation coefficient between the overall PRAS score and science achievement was 0.200 for boys ($p < 0.01$; ranging from 0.176 to 0.257, all $p < 0.01$) and 0.167 for girls ($p < 0.01$; ranging from 0.164 to 0.189, $p < 0.01$). In Grade 8, it was 0.167 for boys ($p < 0.01$; ranging from 0.138 to 0.205, all $p < 0.01$) and 0.148 for girls ($p < 0.01$; ranging from 0.148 to 0.209, all $p < 0.01$).

Correlation data were analyzed using the same repeated-measures General Linear Model as in Study 1 (see Table 9). We again found significant differences between personal relationships and achievement, which depended on the type of relationships ($p < 0.001$). Based on Table 8, we can see that among the three relationship types, the student-peer relationships were most closely related to science achievement.

3.2.3. Multivariate ANOVA

Multi-factor ANOVAs were used to investigate gender and grade differences in the relationship between student personal relationships and science achievement. We found no significant grade or gender differences

TABLE 10 ANOVA for gender and grade in relation to personal relationships and achievement.

	Student-parent relationship					Student-teacher relationship				Student-peer relationship			
	F	*df*	*P*	$\eta^2 p$		*F*	*df*	*p*	$\eta^2 p$	*F*	*df*	*p*	$\eta^2 p$
Gender	0.151	1	0.702	0.009		2.302	1	0.149	0.126	6.138	1	0.025	0.277
Grade	1.534	1	0.233	0.087		1.084	1	0.313	0.063	1.068	1	0.317	0.063
Gender × Grade	5.177	1	0.037	0.244		0.266	1	0.613	0.016	0.555	1	0.467	0.034
Error		16					16				16		

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Bonferroni-corrected.

in the correlations between the three relationship types and science achievement ($p > 0.05$, Bonferroni-corrected). See Table 10 for details.

3.3. Discussion

The results of Study 2 support our hypothesis the conclusion of Study 1; student-peer relationships had the closest association with academic achievement, compared with student-parent and student-teacher relationships. In Study 2, we used science performance as the index of academic performance to verify the relationships we found in Study 1.

Results of the correlation analysis confirmed a significant positive correlation between personal relationships and academic performance, with the student-peer relationships being the most strongly associated academic performance, in both Grade 4 and Grade 8 and among both boys and girls, even after changing the discipline categories.

However, we found that the correlation between student-peer relationships and science achievement did not depend on grade, which was inconsistent with the results from Study 1 in which the same relationship did depend on grade when the subject matter was mathematics. Thus, performance in different disciplines (math vs. science) might not be affected by social relationships in the same way. We speculate that the differences between math and science could be related to the extent that the students care, the time spent studying, and the point-weight of the exam. In China, because math is a subject used for school entrance examinations, it is considered important in every grade. Due to the constant external factors in math learning, we speculate that the grade difference in the correlation between the student-peer relationships and math achievement (Study 1) is simply related to age-related developmental differences, such as cognition. In contrast, it is true that neither students, teachers, or parents pay much attention or invest much in science education in primary schools, as science is not part of any entrance exam. However, science does receive more attention in junior high school as it begins to become incorporated in entrance exams. According to the report from the Program for International Student Assessment (Schleicher, 2019), the average time that Chinese students spent on math and science courses were about 5h per day, each. We speculate that the change in external factors that affect science learning eclipsed age-related developmental differences, resulting in no significant grade difference in the correlation between the student-peer relationships and science achievement.

4. General discussion

In 2018 and 2019, a large sample of primary and middle school students were selected as research participants, and different disciplines were used to explore the relationship between three types of personal relationships and academic performance. The results from Study 1 and Study 2 strongly support our hypothesis that among the three types of important personal relationships (student-parent, student-teacher, and student-peer), the student-peer relationships were the most closely related to academic performance.

4.1. The significant positive correlation between personal relationships and academic achievement

The positive correlation between personal relationships and academic performance in this study is consistent with most previous studies. According to the self-determination theory (SDT), supportive relationships may fulfill students basic psychological need for social relatedness (Deci and Ryan, 2000). When this need is met, adolescents feel connected to their teachers and peers, which fosters their motivation to behave in socially appropriate ways and concentrate on learning. Kiuru et al. (2014) also found that student academic performance can be promoted by increasing the support students receive from peers, parents, and teachers because such increased support leads to better task focus when learning. According to Soe (2020), the relationship between social support that students received from their parents, teachers, and peers and academic achievement were significantly correlated with each other; the more students get social support, the better their academic performance.

4.2. The student-peer relationships were highly related to academic achievement

This study confirmed our hypothesis that student-peer relationships would have a closer association with academic achievement than student-parent or student-teacher relationships. This result is consistent with previous studies. Kindermann (2016) found that students interactions with agemates enhanced their learning over and above the provisions of adult educators; many children appear to go to school or to like school (better) because of their peers and friends. Peer acceptance can have motivational benefits that enhance the learning process; children who enjoy positive relationships with peers also tend to be more engaged in academic tasks and even excel at academic tasks more than those who have problems in their peer relationships (Wentzel, 2017; Wentzel et al., 2020). Adolescents who are victimized by their peers are less motivated to attend school and may miss learning opportunities (Eisenberg et al., 2003).

The current study strongly supports this view. The item on the PRAS that correlated most with performance (highest overall correlation, −0.185 in Study 1; fourth highest overall correlation, −0.156 in Study 2) was, "It is difficult for me to participate in the discussions and activities of my classmates." The PRAS question that correlated second-most overall with achievement scores (−0.181) was an item from Study 2, "I want to participate in discussions and activities with my class." See Supplementary Materials for details. Thus, successful participation in discussions and activities with their peers is closely related to students' academic performance. We propose that in a safe and effective learning environment created by good peer relationships, students are more willing to participate in activities, and the higher the frequency of their collaborative interactions, the better their academic performance.

This conclusion verifies the Social Impact Theory (SIT). The strength (S), the immediacy (I), and the number (N), as proposed by SIT, is further discussed below based on the results of this study, existing studies, and the current situation of education.

First, compared with other groups, peers have more influence on students. Strength (S) is strongly supported by the high correlation between academic performance and peer relationships. As children move from primary school to middle school, the social support function from parents, teachers, and peers seems to change, and by middle and late childhood, close friendships become an important social support system for children. Furman and Buhrmester (1992) compared with parents, school-aged students are more willing to rely on friends to encourage and support them in coping with academic pressure.

Second, in the context of learning, students naturally have more direct contact and closer relationships with their peers than with others. Immediacy (I) is strongly supported by the high correlation between academic performance and peer relationships. In the famous Chinese story "The three times moving of Meng Ke's mother ", Meng Ke's mother chose good environments and companions for the child and moves many times. Known since ancient times, people have paid attention to the effect that companions have on their children.

Finally, peers are the people with whom students have the most contact. The Number (N) again is strongly supported by the high correlation between academic performance and peer relationships. Most researchers agree that the number of partners affects an individual's performance. For example, Carter and Hughes (2005) found that students with disabilities had higher levels of social interaction and contact with the general curriculum when they worked with two peers than when they worked with one.

In addition, we believe that cooperative inquiry teaching, which has been advocated by educators in recent years, has a positive effect on peer closeness. As early as 2001, China mentioned and advocated cooperative learning in the Decision of the State Council on the Reform and Development of Basic Education, pointing out that it encourages cooperative learning, promotes mutual communication and development among students, and promotes teachers and students to learn from each other. The Outline of Basic Education Curriculum Reform (Trial), issued in the same year, once again proposed that comprehensive practical activities should be taken as compulsory courses from primary school to senior high school. Research-based learning is a key focus of this new plan, implemented in order to develop communication and cooperation among students.

In the following years, cooperative learning was discussed as China's education and teaching reform deepened. Most teachers are actively practicing and guiding students to cooperate in learning. A series of studies confirmed the effect of peer cooperative learning. For instance, Veldman et al. (2020) found that cooperative learning may lead to improved group work behavior in young pupils (6–7 years old). Molla and Muche (2018) showed that a significant gain in learning occurred *via* a cooperative learning-achievement division followed by a cooperative discussion group. We speculate that in efficient cooperative learning processes, students and peers become closer and the importance of peers in a student's learning increases.

4.3. Educational implications

4.3.1. The importance of participation and cooperation

The results of this study remind educators that in addition to cognitive factors, personal relationships, especially peer relationships, need to be considered with regards to student academic performance. Student interactions with their peers in discussions or activities is closely related to academic performance. This enlightens our daily teaching work: (1) Teachers should guide students to integrate into the collective learning process, create learning opportunities through communication, encourage students to participate in discussions and activities, help them learn how to cooperate, and should not deprive students of time and space for independent communication; (2) teachers should also create a class and campus atmosphere of unity, mutual assistance, and friendship, guide students with a positive learning attitude to inspire others around them, and should not cultivate antagonistic emotions or over-emphasize competition; (3) teachers should also strive to improve the environmental conditions and adjust the space between students (such as seat adjustment) to promote peer relationships and increase the possibility of discussion and exchange. A series of studies have shown that learning spaces are becoming an important fulcrum of school reform. Flexible learning spaces have a significant impact on learning outcomes, including improving academic performance, promoting teacher-student interactions in class, and improving the learning experience (Brooks and Baepler, 2012; Baepler and Walker, 2014; Gremmen et al., 2018).

4.3.2. The importance of school and class construction

The results provided some evidences for the valuable of the existence of schools and classrooms. In recent years, many schools are exploring the abolition of classroom-teaching systems. Other are considering improving the classroom-teaching system by using an "Optional Class System" or other teaching methods. "With the rise of intelligent technology, the widespread availability of free learning opportunities signals the decline of existing curriculum structures and the collapse of school systems, the disappearance of traditional teaching staff, and the emergence of individuals as producers and practitioners of knowledge" (Organization for Economic Cooperation and Development, 2020). The natural limitations of traditional school education and classroom-teaching systems (such as unified teaching requirements and methods) make it difficult to meet the diverse learning needs of individuals.

However, the results of this study also remind us to think again. Collective teaching is not without advantages, and classrooms are also necessary. We believe that although ubiquitous learning resources are everywhere, the physical learning space is still irreplaceable. The process of knowledge acquisition is bets when it is not a solitary endeavor. It can be affected by many social factors; peer interactions and the quality of peer relationships within physical learning spaces are closely related to academic performance. Our view is supported by the report by the United Nations Educational, Scientific, and Cultural Organization, which states in 2015 that "learning should not be an individual matter, but a social experience that requires learning with others and through discussion and debate with peers and teachers." The classroom structure provides an environment for students to promote and interact with their peers as a collective, where they study together and gain a sense of belonging and security. It provides stable personal support and assists with learning, which is difficult to establish in an "Optional Class System." We know that in the "Optional Class System," choosing courses according to interest and level can give more space to a group of students with strong autonomy and good learning foundation, but for other students, it might bring about confusion. This could lead to the strong becoming stronger and the weak becoming weaker (Schofield, 2010; Hamilton and O'Hara, 2011; Wilkinson et al., 2015; Smyth, 2016).

Based on this, we believe that schools and classrooms remain valuable and should continue to exist. In the future, the construction of schools and classrooms needs to be constantly improved, which will be through a process that connects the past to the future.

4.4. Limitations of the study and implications for future research

To the best of our knowledge, this is the first analysis that has compared how well three important types of personal relationships correlate with academic performance. Therefore, more research is needed to confirm our conclusion that student-peer relationships are the most important. Additionally, some limitations need to be noted.

First, this study only divided personal relationships into three types, but did not continue to refine the indicators. According to existing studies, student-parent, student-teacher, and student-peer relationships are defined in various ways and classified by indicators (for example, student-peer relationships are refined into peer acceptance and peer rejection (Zhang et al., 2013)). Future studies can consider further refining the three dimensions we used here.

Second, although we designed two studies to confirm our results, the analyses were correlational; therefore, causal inferences cannot be drawn. Little is known about how the student-peer relationships as a variable affects academic performance and the underlying processes that can explain these relationships. The same goes for our speculation that cooperative learning is closely related to our conclusion. It is necessary to carefully consider how learning style affects student relationships and academic performance, and what variables should be controlled for further research on this scientific issue.

Third, this study selected Grade 4 and Grade 8 students in Qingdao, Shandong Province, China as representatives for the investigation. In order to make the research results more representative, future research should focus on students from more grade levels with more varied demographic profiles. This might reveal a pattern of findings and developmental characteristics not captured by the current study. Here, we have discussed the correlation between personal relationships and academic performance in math and science. In the next step, other disciplines should be considered to enrich the existing research results. Furthermore, the samples of this study were only from China, where interpersonal relationships were of particular importance, and the representativeness has certain limitations. Future studies could be extended to more diverse samples to investigate the correlation between interpersonal relationships and academic performance.

5. Conclusion

Through the investigation and analysis of a large data sample, we have compared how well three important personal relationships (student-parent, student-teacher, and student-peer) correlate with academic performance. We found that the student-peer relationships were most closely related to academic performance. The conclusion of this study helps us to further understand the relationship between personal relationships and academic performance. At the same time, it also reminds educators to pay attention to the personal relationships among their students, especially the peer relationships. It also reminds educators to create opportunities and environmental conditions for exchange and learning, encourage discussion and cooperation, and create a united classroom and campus atmosphere.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Experimental Ethics Committee of basic teaching center of Ocean University of China. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

XY, HW and XZ designed experiments and revised the manuscript. XY and HW collected data and analyzed data. XW, HZ, XZ and MS analyzed data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the MOE (Ministry of Education of the People's Republic of China) Project of Humanities and Social Sciences [No. 18YJC190030] and Qingdao Municipal Planning Project of Social Sciences [No. QDSKL2001017].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1012701/full#supplementary-material

References

Altermatt, E. R. (2019). Academic support from peers as a predictor of academic self-efficacy among college students. *J. College Student Retention Res. Theory Prac.* 21, 21–37. doi: 10.1177/1521025116686588

Baepler, P., and Walker, J. D. (2014). Active learning classrooms and educational alliances: changing relationships to improve learning [article]. *New Dir. Teach. Learn.* 2014, 27–40. doi: 10.1002/tl.20083

Barile, J. P., Donohue, D. K., Anthony, E. R., Baker, A. M., Weaver, S. R., and Henrich, C. C. (2012). Teacher–student relationship climate and school outcomes: implications for educational policy initiatives. *J. Youth Adoles.* 41, 256–267. doi: 10.1007/s10964-011-9652-8

Brooks, D. C., and Baepler, P. (2012). What's Going On? How What Happens in Active Learning Classrooms (ALCs) Matters. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/161414.

Carter, E. W., and Hughes, C. (2005). Increasing social interaction among adolescents with intellectual disabilities and their general education peers: effective interventions. *Res. Practice for Persons with Severe Disab.* 30, 179–193. doi: 10.2511/rpsd.30.4.179

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., and Gaviria, J. L. (2015). Parental involvement on student academic achievement: a meta-analysis. *Educ. Res. Rev.* 14, 33–46. doi: 10.1016/j.edurev.2015.01.002

Chen, J. J. (2005). Relation of academic support from parents, teachers, and peers to Hong Kong adolescents' academic achievement: the mediating role of academic engagement. *Genet. Soc. Gen. Psychol. Monogr.* 131, 77–127. doi: 10.3200/MONO.131.2.77-127

Chen, J. J. (2008). Grade-Level Differences. *Sch. Psychol. Int.* 29, 183–198. doi: 10.1177/0143034308090059

Chiu, M. M. (2010). Effects of inequality, family and school on mathematics achievement: country and student differences. *Soc. Forces* 88, 1645–1676. doi: 10.1353/sof.2010.0019

Collie, R. J., Martin, A. J., Papworth, B., and Ginns, P. (2016). Students' interpersonal relationships, personal best (pb) goals, and academic engagement. *Learn. Individ. Differ.* 45, 65–76. doi: 10.1016/j.lindif.2015.12.002

Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inq.* 11, 227–268. doi: 10.1207/S15327965PLI1104_01

Dhingra, R., and Manhas, S. (2009). Academic performance of children as a function of interaction with parents and teachers. *J. Social Sci.* 18, 59–64. doi: 10.1080/09718923.2009.11892666

Eisenberg, M. E., Neumark-Sztainer, D., and Perry, C. L. (2003). Peer harassment, school connectedness, and academic achievement. *J. Sch. Health* 73, 311–316. doi: 10.1111/j.1746-1561.2003.tb06588.x

Espelage, D. L., Hong, J. S., Rao, M. A., and Low, S. (2013). Associations between peer victimization and academic performance. *Theory Pract.* 52, 233–240. doi: 10.1080/00405841.2013.829724

Fuligni Andrew, J., and Eccles Jacquelynne, S. (1993). Perceived parent-child relationships and early adolescents' orientation toward peers. *Dev. Psychol.* 29, 622–632. doi: 10.1037/0012-1649.29.4.622

Furman, W., and Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child Dev.* 63, 103–115. doi: 10.2307/1130905

Gao, X., and Xue, H. P. (2020). Parental involvement, peer influence and academic achievement. *Educ. Sci. Res.* 6, 55–63.

García Bacete, F. J., Muñoz Tinoco, V., Marande Perrin, G., and Rosel Remírez, J. F. (2021). Stability of peer acceptance and rejection and their effect on academic performance in primary education: a longitudinal research. *Sustain.* 13:2650. doi: 10.3390/su13052650

Gremmen, M. C., Berg, Y. H. M. V. D., Steglich, C., Veenstra, R., and Dijkstra, J. K. (2018). The importance of near-seated peers for elementary students' academic engagement and achievement. *J. Appl. Dev. Psychol.* 57, 42–52. doi: 10.1016/j.appdev.2018.04.004

Hajovsky, D. B., Oyen, K. A., Chesnut, S. R., and Curtin, S. J. (2020). Teacher-student relationship quality and math achievement: the mediating role of teacher self-efficacy. *Psychol. Sch.* 57, 111–134. doi: 10.1002/pits.22322

Hajovskya, D. B., Masonb, B. A., and McCune, L. A. (2017). Teacher-student relationship quality and academic achievement in elementary school: a longitudinal examination of gender differences. *J. Sch. Psychol.* 63, 119–133. doi: 10.1016/j.jsp.2017.04.001

Hamilton, L., and O'Hara, P. (2011). The tyranny of setting (ability grouping): challenges to inclusion in Scottish primary schools. *Teach. Teach. Educ.* 27, 712–721. doi: 10.1016/j.tate.2010.11.009

Hughes, J. N., Wu, J.-Y., Kwok, O. M., Villarreal, V., and Johnson, A. Y. (2012). Indirect effects of child reports of teacher-student relationship on achievement. *J. Educ. Psychol.* 104, 350–365. doi: 10.1037/a0026339

Igbo, J. N., Sam, O. A., Onu, V. C., and Dan, M. (2015). Parent-child relationship motivation to learn and students academic achievement in mathematics. *Impact J.* 3, 87–108.

Kindermann, T. A. (2016). "Peer group influences on students' academic motivation," in *Handbook of social influences in school contexts* eds. K. Wentzel and G. Ramani (Routledge), 31–47.

King, R. B., and Ganotice, F. A. (2014). The social underpinnings of motivation and achievement: investigating the role of parents, teachers, and peers on academic outcomes. *Asia Pac. Educ. Res.* 23, 745–756. doi: 10.1007/s40299-013-0148-z

Kiuru, N., Nurmi, J.-E., Aunola, K., and Salmela-Aro, K. (2009). The role of peer groups in adolescents' educational trajectories. *Eur. J. Dev. Psychol.* 6, 521–547. doi: 10.1080/17405620701330599

Kiuru, N., Pakarinen, E., Vasalampi, K., Silinskas, G., Aunola, K., Poikkeus, A. M., et al. (2014). Task-focused behavior mediates the associations between supportive interpersonal environments and students' academic performance. *Psychol. Sci.* 25, 1018–1024. doi: 10.1177/0956797613519111

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., and Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educ. Res. Rev.* 10, 133–149. doi: 10.1016/j.edurev.2013.02.002

Lam, S. F., Jimerson, S., Kikas, E., Cefai, C., Veiga, F. H., Nelson, B., et al. (2012). Do girls and boys perceive themselves as equally engaged in school? The results of an international study from 12 countries. *J. Sch. Psychol.* 50, 77–94. doi: 10.1016/j.jsp.2011.07.004

Latan, B. (1981). The psychology of social impact. *Am. Psychol.* 36, 343–356. doi: 10.1037/0003-066X.36.4.343

Lee, J. S. (2012). The effects of the teacher–student relationship and academic press on student engagement and academic performance. *Int. J. Educ. Res.* 53, 330–340. doi: 10.1016/j.ijer.2012.04.006

Leung, C., Leung, J., Kwok, S., Hui, A., Lo, H., Tam, H. L., et al. (2021). Predictors to happiness in primary students: positive relationships or academic achievement. *Appl. Res. Qual. Life* 16, 2335–2349. doi: 10.1007/s11482-021-09928-4

Li, L., Peng, Z., Lu, L., Liao, H., and Li, H. (2020). Peer relationships, self-efficacy, academic motivation, and mathematics achievement in Zhuang adolescents: a moderated mediation model. *Child Youth Serv. Rev.* 118:105358. doi: 10.1016/j.childyouth.2020.105358

Ly, J., Zhou, Q., Chu, K., and Chen, S. H. (2012). Teacher-child relationship quality and academic achievement of Chinese American children in immigrant families. *J. Sch. Psychol.* 50, 535–553. doi: 10.1016/j.jsp.2012.03.003

Martin, A. J., and Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: yields for theory, current issues, and educational practice. *Rev. Educ. Res.* 79, 327–365. doi: 10.3102/0034654308325583

Molla, E., and Muche, M. (2018). Impact of cooperative learning approaches on students' academic achievement and laboratory proficiency in biology subject in selected rural schools, Ethiopia. *Educ. Res. Int.* 2018, 1–9. doi: 10.1155/2018/6202484

Moore, G. F., Cox, R., Evans, R. E., Hallingberg, B., Hawkins, J., Littlecott, H. J., et al. (2018). School, peer and family relationships and adolescent substance use, subjective wellbeing and mental health symptoms in Wales: a cross sectional study. *Child Indic. Res.* 11, 1951–1965. doi: 10.1007/s12187-017-9524-1

Nokali, N. E. E., Bachman, H. J., and Votruba-Drzal, E. (2010). Parent involvement and Children's academic and social development in elementary school. *Child Dev.* 81, 988–1005. doi: 10.1111/j.1467-8624.2010.01447.x

Oberle, E., and Schonert-Reichl, K. A. (2013). Relations among peer acceptance, inhibitory control, and math achievement in early adolescence. *J. Appl. Dev. Psychol.* 34, 45–51. doi: 10.1016/j.appdev.2012.09.003

Organization for Economic Cooperation and Development (2020). Back to the future of education: four OECD scenarios for schooling. *Educ. Res. Innov.* 4, 52–55. doi: 10.1787/178ef527-en

Pedersen, E. R., Labrie, J. W., and Lac, A. (2008). Assessment of perceived and actual alcohol norms in varying contexts: exploring social impact theory among college students. *Addict. Behav.* 33, 552–564. doi: 10.1016/j.addbeh.2007.11.003

Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., and Oort, F. J. (2011). The influence of affective teacher-student relationships on Students' School engagement and achievement: a meta-analytic approach. *Rev. Educ. Res.* 81, 493–529. doi: 10.3102/0034654311421793

Rubin, K. H., Bukowski, W., and Laursen, B. (2010). Handbook of peer interactions, relationships, and groups. Handbook of peer interactions, relationships, and groups. *J. Autism Dev. Disord.* 40, 264–265.

Schleicher, A. (2019). PISA 2018: Insights and interpretations. Available at: https://www.oecd.org/pisa/publications/pisa-2018-results.htm

Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teach. Coll. Rec.* 112, 1492–1528. doi: 10.1177/016146811011200506

Smyth, E. (2016). *Students' experiences and perspectives on secondary education*. London: Palgrave Macmillan Publications.

Soe, T. T. (2020). The relationship of social support with academic achievement and career aspiration of Grade-8 students in Myanmar. *Technium Social Sci. J.* 9:144. doi: 10.47577/tssj.v9i1.1102

Teng, X., Liu, G., and Song, T. (2018). Influence of teacher-student relationship on junior high Students' academic achievements: mediating role of self-esteem. *2018 2nd International Conference on Management, Education and Social Science (ICMESS 2018).*

Tobbell, J., and O'Donnell, V. L. (2013). The formation of interpersonal and learning relationships in the transition from primary to secondary school: students, teachers and school context. *Int. J. Educ. Res.* 59, 11–23. doi: 10.1016/j.ijer.2013.02.003

Vargas-Madriz, L. F., and Konishi, C. (2021). The relationship between social support and student academic involvement: the mediating role of school belonging. *Can. J. Sch. Psychol.* 36, 290–303. doi: 10.1177/08295735211034713

Veldman, M. A., Doolaard, S., Bosker, R. J., and Snijders, T. (2020). Young children working together. Cooperative learning effects on group work of children in grade 1 of primary education. *Learn. Instr.* 67:101308. doi: 10.1016/j.learninstruc.2020.101308

Vignery, K., and Laurier, W. (2020). Achievement in student peer networks: a study of the selection process, peer effects and student centrality. *Int. J. Educ. Res.* 99:101499. doi: 10.1016/j.ijer.2019.101499

Walsh, S. D., Harel-Fisch, Y., and Fogel-Grinvald, H. (2010). Parents, teachers and peer relations as predictors of risk behaviors and mental well-being among immigrant and israeli born adolescents. *Soc. Sci. Med.* 70, 976–984. doi: 10.1016/j.socscimed.2009.12.010

Wang, M. T., and Eccles, J. S. (2012). Social support matters: longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Dev.* 83, 877–895. doi: 10.1111/j.1467-8624.2012.01745.x

Wentzel, K. R. (1998). Social relationships and motivation in middle school: the role of parents, teachers, and peers. *J. Educ. Psychol.* 90, 202–209. doi: 10.1037/0022-0663.90.2.202

Wentzel, K. R. (2017). "Peer relationships, motivation, and academic performance at school" in *Handbook of competence and motivation: Theory and application* eds. A. J. Elliot, C. S. Dweck and D. S. Yeager (The Guilford Press), 586–603.

Wentzel, K. R., Jablansky, S., and Scalise, N. R. (2020). Peer social acceptance and academic achievement: a meta-analytic study. *J. Educ. Psychol.* 113, 157–180. doi: 10.1037/edu0000468

Wentzel, K. R., Russell, S., and Baker, S. (2016). Emotional support and expectations from parents, teachers, and peers predict adolescent competence at school. *J. Educ. Psychol.* 108, 242–255. doi: 10.1037/edu0000049

Wilkinson, S., Penney, D., and Allin, L. (2015). Setting and within-class ability grouping: a survey of practices in physical education. *Eur. Phys. Educ. Rev.* 22, 336–354. doi: 10.1177/1356336X15610784

Zhang, J., Tian, L. M., and Zhang, W. X. (2013). Relations between peer rejection and early Adolescents' academic achievement: the moderating effects of peer acceptance and friendship support. *Psycholog. Develop. Educ. 2013* 29, 353–360.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fpsyg.2023.1108740

Check for updates

OPEN ACCESS

EDITED BY
Nelly Lagos San Martín,
University of the Bío Bío,
Chile

REVIEWED BY
Murat Ekici,
Uşak University,
Türkiye
Silvana Watson,
Old Dominion University,
United States

*CORRESPONDENCE
Bin Xiong
✉ bxiong@math.ecnu.edu.cn
Yanyun Jiang
✉ jarenyry@mail.bnu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 26 November 2022
ACCEPTED 03 February 2023
PUBLISHED 22 February 2023

CITATION
Jiang P, Zhang X, Ruan X, Feng Z, Xiong B and Jiang Y (2023) A comparative study of high school mathematics teachers' audible teaching language: A student satisfaction perspective. *Front. Psychol.* 14:1108740.
doi: 10.3389/fpsyg.2023.1108740

A comparative study of high school mathematics teachers' audible teaching language: A student satisfaction perspective

Peijie Jiang[1], Xiangjun Zhang[2], Xiaomeng Ruan[1], Zirong Feng[1], Bin Xiong[3,4]* and Yanyun Jiang[2,5]*

[1]School of Mathematics and Statistics, Hunan Normal University, Changsha, Hunan, China, [2]The High School Attached to Hunan Normal University, Changsha, Hunan, China, [3]School of Mathematical Sciences, East China Normal University, Shanghai, China, [4]Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice, Shanghai, China, [5]School of Future Education, Beijing Normal University, Zhuhai, Guangdong, China

Teachers' audible teaching language is essential for organizing classroom instruction. This study used a questionnaire to compare expert, skilled, and novice high school mathematics teachers' audible teaching language from the perspective of student satisfaction. The sample was selected using a purposive sampling technique, and the participants were students from a key high school in Changsha, China. A research framework and research instrument with good reliability and validity were constructed for this study. The data were analyzed using SPSS 22.0 and AMOS 22.0. The results showed 263 valid questionnaires, good measurement model fit, and high reliability and validity of the questionnaire. It was found that: (1) students were highly satisfied with the audible teaching language of high school mathematics teachers; (2) student satisfaction with the audible teaching language of skilled, expert, and novice mathematics teachers declined in order, but there was no significant difference overall; (3) students were more satisfied with expert mathematics teachers than with novice teachers in terms of the tone and adaptability of the audible teaching language. The researchers discussed the study's results, suggested how pre-service and post-service mathematics teachers can improve the quality of their audible teaching language, and pointed out the value and limitations of the study.

KEYWORDS

audible teaching language, mathematics teacher, language of instruction, student satisfaction, comparative study, teaching tone, language speed

1. Introduction

As research in mathematics education continues to evolve, the mathematics classroom has gradually become the focus of research. Focusing on the role and functions of mathematics teachers in the classroom (Jiang et al., 2022), on what happens in the classroom (Lester, 2013), and on the influence of the classroom environment on students' interest in learning (Liu et al., 2022) are all important for academic research. As an essential part of the mathematics classroom, teachers play a significant role in imparting knowledge. The primary way teachers transmit knowledge is through verbal expressions, including lectures, questions, and conversations, all of which are forms of audible teaching language. An audible teaching language is a form of

 10.3389/fpsyg.2023.1108740

language that transmits knowledge through sound, with pauses, soothing, and other intonations that provide information not easily conveyed through text (Carr and Kemmis, 1986). It can help learners to relate the information they hear to existing knowledge more quickly (Gaver and Gaver, 1993). Students facilitate their knowledge acquisition by listening to the teacher. Teachers can maximize the purpose of questioning and facilitate students' learning by using verbal language to ask questions (Aziza, 2021). Productive classroom discourse is also beneficial to improve students' performance in the classroom (Mok et al., 2022). In addition, one study found that teachers' audible teaching language could influence classroom climate, and it indicated that a positive classroom climate could compensate for the negative effects of gender and low socioeconomic status on students' performance in mathematics (López et al., 2022). Therefore, teachers' classroom language is closely related to students' learning outcomes, and teachers should use engaging and expressive language in the classroom (Ormond, 2021).

There have been several studies related to the audible teaching language for mathematics teachers. Examples include research on classroom dialog and classroom feedback. Effective classroom dialog is essential for high-quality mathematics instruction (Høynes et al., 2019), and dialog clearly relies on teachers' audible teaching language. Some researchers have examined the characteristics of effective classroom dialog from the perspective of Chinese mathematics classrooms (Zhao et al., 2022). By examining classroom videotapes of expert and novice teachers, they found that the proportion of conversations about the basics was significantly lower, and the ratio of conversations about guessing was considerably higher in expert teachers' classrooms compared to novice teachers. In addition, feedback in the classroom was often achieved through audible teaching language. One study investigated how 47 teachers provided feedback in 172 mathematics lessons in Norwegian junior high schools (Stovner and Klette, 2022). By analyzing the quality of feedback, the amount of feedback, and whether the feedback targeted students' procedural skills, conceptual understanding, or engagement in mathematical practices, the researchers found that teachers spent a significant amount of time providing specific feedback, most of which was directed at procedural skills and less at conceptual feedback. Although there are some studies related to instructional language, these studies have not directly examined the audible teaching language of high school mathematics teachers and have not focused on the vocal qualities of instructional language.

In general, the language of instruction includes both spoken and silent languages. Teacher-student communication in the classroom is not only through spoken language. It is also very important for teachers to understand the non-verbal information they send and receive in the classroom (Miller, 2005). Among them, body language as a silent teaching language has attracted much attention from education researchers. Teachers' body language, such as gestures, eyes, and expressions, plays a vital role in teaching (Woolfolk and Galloway, 1985). This role is directly reflected in that teachers should learn to use natural body movements when speaking in front of students because inappropriate actions can weaken the transmission of knowledge. Therefore, understanding and effectively using body language is also a critical teaching skill (Hale et al., 2017), and many researchers have explored teachers' body language. For example, one study showed that teachers' gestures not only impacted students' learning but also on teaching itself and that different gestures had different effects (Gaythwaite, 2005; Yang et al., 2020). In various situations, gestures tend to be understood faster than using language, so gestures are considered an integral part of cognition (DeLiema et al., 2021).

Teachers' audible language is also important. It has its drawbacks in terms of communication effectiveness (Miller, 2005) because it is non-visible and non-written, and students cannot access the information repeatedly as they can when reading a text. However, teachers' audible language of instruction is more fundamental than body language, and the absence of audible teaching language generally means no instruction. Audible teaching language is at least as crucial as silent instructional language. Both language and the body play a vital role in the foundation of higher-order thinking, with dynamic gestures and speech making individually essential contributions to the formation of mathematical arguments (Pier et al., 2019). The teacher's body language has to coordinate with the verbal language to convey more visual information (Ngo et al., 2022). Therefore, it is also crucial to study teachers' audible language in mathematics classrooms. Because students serve as the focus of classroom instruction, studying the audible teaching language of high school mathematics teachers from the perspective of student satisfaction is valuable for improving mathematics learning satisfaction and teaching quality and will add new knowledge to the field of mathematics teacher education.

Several studies have compared novice teachers with expert teachers from different perspectives. One researcher compared the planning, teaching, and post-lesson reflection of three novice teachers with the three expert teachers they worked with and found that the novices' cognitive patterns and instructional reasoning skills were not as well developed as the expert teachers (Borko and Livingston, 1989). Regarding the organization and content of mathematics teachers' subject matter knowledge, experts categorized problems more finely and deeply, while novices had more horizontal, independent categorization systems (Leinhardt and Smith, 1985). A comparative study of the language of instruction of new and veteran mathematics teachers, on the other hand, noted that novice teachers used too much instructional language to teach knowledge, while veteran teachers were good at using instructional language to guide students' thinking (Ye et al., 2015). These studies suggest that expert teachers are more adept than novice teachers. Therefore, in this study, we hypothesized that expert teachers were more competent than novice teachers in using audible teaching language and that the quality of teachers' audible teaching language could be improved through appropriate training. Learning how expert teachers use the language helped novice and pre-service teachers improve their language ability.

2. Overview of the literature

2.1. Audible teaching language and its research

An audible language is a form of language that is communicated through sound and can help learners connect the information they hear with existing knowledge more quickly (Gaver and Gaver, 1993). Pauses, soothing, and other intonations provide information that is not easily conveyed through text (Carr and Kemmis, 1986). The perception and production of sound are fundamental to human cognition and behavior (Bautista and Roth, 2012). Further, the study of language use can provide insight into popular discussion topics among mathematics teachers and mathematics educators.

Audible teaching language is one of the ways to express the language of instruction. The language of instruction is logical, clear, dominant, humorous, authoritative, and developmental. Based on these characteristics, concise, logical language enhances students' attention, leading to improved student performance (Cogan, 1958), while teachers' clarity, expressiveness, and delivery are significantly correlated and influence students' evaluations (Solomon et al., 1964). The language of instruction is crucial to student development (Wasik and Hindman, 2011). Some researchers have suggested that a teacher's speaking can help students improve their comprehension (Franke et al., 2009). Teachers' audible teaching language is part of classroom discourse, and there has been some research on classroom discourse. The teacher's audible teaching language makes classroom dialog possible. Classroom discourse includes teacher discourse, student discourse, silence, and discussion (Jiang et al., 2018). In general, teachers' discourse dominates classroom conversations, is authoritative, contributes to the transmission of knowledge in the classroom, and influences the structure of authority (Scott, 1998; Ng et al., 2021) and classroom effectiveness (Guo and Xia, 2021), and humorous language contributes to student learning (Do et al., 2022).

Classroom discourse is a significant component of the classroom and carries a great deal of helpful information that plays an essential role in learning (Grifenhagen and Barnes, 2022). Properly managed classroom discourse can allow students to develop their understanding and help them benefit from the ideas of their peers and teachers (Wang et al., 2014).

Classroom dialog can be divided into conversational and questioning types. On the one hand, students learn not in isolation but through dialog (Lee and Kinzie, 2012). Students' engagement in dialogic discourse, such as questioning and connecting ideas, contributes to developing active, analytical, and personal thoughts (Scott, 1998). On the other hand, teacher questioning is a distinctive feature of classroom talk, and focusing on questioning practices helps us better understand the role of teacher questioning in scaffolding instruction. Chin (2007) explored how teachers used questions in classroom discourse to support students' thinking and help them construct knowledge.

The audible teaching language is the teaching language carried by sound. From the above analysis, teachers' audible teaching language is fundamental. Still, there are few achievements in studying teachers' audible teaching language behavior, and further research is needed.

2.2. Comparison of expert and novice mathematics teachers

It is common to distinguish between expert and novice teachers regarding educational and teaching experience and theory knowledge (Sternberg and Horvath, 1995). Cai et al. (2022) treated mathematics teachers with more than 25 years of teaching experience as expert mathematics teachers. Some researchers referred to teachers with an average of 22 years of teaching experience as expert teachers and those with an average of 3 years of teaching experience as novice teachers (Zhao et al., 2022). In China, because there are strict professional standards for evaluating teacher titles, the titles can broadly reflect the professional competence of teachers. In this study, teachers were classified regarding their titles and years of teaching experience. Novice teachers were junior teachers who had taught for less than 5 years. Expert teachers were intermediate and advanced teachers who had taught for over 10 years. Those who have been teaching for more than 5 years and less than 10 years were classified as skilled teachers, regardless of their job title.

There is a richer body of research comparing expert and novice mathematics teachers. About a decade ago, a researcher explored the attention to classroom events of 10 experts and 10 novice teachers in China. This researcher noted that expert teachers focused more on developing students' mathematical thinking and higher-order thinking, as well as the coherence of students' mathematical knowledge, than novice teachers (Huang and Li, 2012). Researchers are still exploring the topic of teacher attention. One study highlighted the cultural dependence on the development of expertise in teacher attention by comparing empirical knowledge of the development of teacher attention from the novice level to the expert level (Bastian et al., 2022). Other researchers have explored more specific issues, such as how secondary school teachers and students impose personal structures on fractional expressions and equations, noting that expert teachers construct a particular fractional expression in various ways (Ruede, 2013).

More recently, some researchers have elaborated on how classroom expertise affects visual perception and mental interpretation by comparing expert and novice teachers' knowledge levels and their decisions to act on classroom events that help clarify differences in teachers' perceptions and representations of events (Wolff et al., 2021). Other researchers examined how expert and novice (pre-service) teachers completed mathematical modeling tasks, respectively, and how they noticed the written work of student thinking completed in response to mathematical modeling tasks. Almost all expert mathematics teachers responded by asking questions, while about one-third of pre-service mathematics teachers directly corrected students' errors, and another third pointed out errors but did not correct them (Cai et al., 2022).

Overall, studies comparing expert and novice mathematics teachers have been numerous and have covered a wide range of topics. However, there have been few studies on the audible teaching language of expert and novice mathematics teachers. The comparative studies that have been conducted on mathematics teachers' instructional language are based on classroom video analyzes, and the instructional language in these studies also includes, for example, body language (Ye et al., 2015) and does not specifically examine audible teaching language. Studying the differences in the audible teaching language of expert and novice mathematics teachers from the perspective of student satisfaction is theoretically and practically necessary.

2.3. Student satisfaction and its evaluation

Student satisfaction is the psychological feeling of satisfaction or dissatisfaction students have when they compare their perceived effectiveness with their desired effectiveness throughout instruction. As addressed in this paper, student satisfaction refers to the extent to which students feel satisfied with the mathematics teacher's audible teaching language. Various factors influence student satisfaction, and we need to distinguish among them to determine their impact on the provision of quality education. Recognizing the specific factors affecting student satisfaction will help us develop strategies to enhance student's learning experience and improve the education provided (Cheng, 2016, pp. 33–45).

Student satisfaction is considered a rather important aspect of educational strategy, and there is a direct, positive, and significant relationship between the quality of education and student satisfaction (Meštrović, 2017). Specifically, student satisfaction effectively influences learning ability (Panyajamorn et al., 2018). Given the influence of student satisfaction on the quality of education, student satisfaction is considered an indicator when assessing instruction content and the level of importance of mathematics subjects (Carlos Ramirez-Cruz et al., 2018). Factors that influence student satisfaction have been widely discussed, including students' learning styles (Kim, 2021; Mahir et al., 2021), gender, students' self-efficacy, teachers' teaching methods and tools (Gudelj et al., 2021), and instructional materials (Lee, 2014).

Students' perception of the classroom learning environment measures student satisfaction. Teacher professionalism affects students' attitudes toward mathematics lessons and perceptions of classroom climate, ultimately affecting students' academic performance and overall satisfaction with the curriculum (Suh et al., 2018). Teacher professional development should include improving teachers' language of instruction. Teachers' classroom discourse can reflect teachers' audible teaching language level to some extent, which is an essential component of the classroom environment. The teacher's central role is to guide and sustain the conversation in the desired direction related to the learning objectives and to reduce gaps in students' performance on the intended learning objectives (Kayima and Mkimbili, 2021).

These studies point out that classroom discussion and dialog are essential avenues of teacher-student interaction, and learning satisfaction is directly influenced by learner interaction, perceived ease of use, and academic performance (Nagy, 2018). How teachers initiate and facilitate discussions when students respond essentially determines classroom interactions (Khoza and Msimanga, 2021). Therefore, facilitating discussions in the mathematics classroom effectively improves students' thinking, reasoning, and problem-solving skills to support their mathematics learning and improve their learning ability and student satisfaction.

The above analysis revealed that student satisfaction is vital for improving the quality of education and that teacher quality is a crucial determinant of educational quality. Promoting teachers' language skills can help to create a positive classroom environment, promote good student-teacher interaction, and increase student satisfaction. Previous research has focused more on the content of the language of instruction and less on the vocal characteristics of the teacher's language. This study helps to advance research in this area.

3. Purpose of study

To understand students' satisfaction with current high school mathematics teachers' audible teaching language and to recognize the difference between expert and novice teachers' use of audible teaching language, this study focused on answering the following questions:

1. How satisfied are students with the audible teaching language used by high school mathematics teachers in their classrooms?
2. What are the similarities and differences in the audible teaching language used by expert, skilled, and novice high school mathematics teachers in their classrooms?

4. Methods and materials

4.1. Framework and tools

Research frameworks and instruments on teachers' audible teaching language are unavailable, but there are several scales for evaluating instructional language. Teachers' discourse actions include engaging in reflective discourse, responding to prior student discourse with neutral restatements, and exploring students' dominant ideas (Soysal and Yilmaz-Tuzun, 2021). Bekiari et al. (2006) focused on the affective expression dimension of teachers' instructional language, examined students' perceptions of teachers' verbal aggression, and investigated students' perceptions of learning activities (doing homework) and the fun nature of the school environment using a classroom satisfaction scale. Smart and Marshall (2013) proposed a structure for measuring instructional language that consisted of questioning level, question complexity, questioning ecology, patterns of engagement, and classroom interaction. The questioning level examines whether the teacher's classroom language is asking questions to students at different levels, and question complexity focuses on whether the questions are focused on one correct answer. In a comparative study of secondary school biology teachers' audible teaching language, researchers constructed a measurement framework that included both form and content elements (Kong, 2018). The form element includes volume, tone, and speech rate; the content element includes imagery, interest, science, conciseness, illumination, relevance, coherence, and education. The framework is the result of the Delphi method, used by eight experts and researchers to discuss the proposed indicators and the weighting and analysis of the proposed indicators, combined with the validation factor analysis after repeated revisions and discussions to form a particular reference value. However, this framework has a crossover between secondary indicators, such as correlation and coherence. Therefore, it is necessary to redesign the survey analysis framework to investigate better and analyze the audible teaching language of high school mathematics teachers from the perspective of student satisfaction. In the absence of instruments to measure teachers' audible teaching language, the researchers developed an instrument based on the work of Kong (2018).

This study first constructed a preliminary framework based on the existing literature and the practical experiences of two skilled high school mathematics teachers. Then the basic framework (framework for analyzing high school mathematics teachers' satisfaction with the audible teaching language) was determined using the Delphi method based on the preliminary framework, as shown in Table 1.

In the framework, the audible teaching language of high school mathematics teachers contains both *Sound* and *Content* attributes. The observed indicators of the *Sound* attribute are *Volume*, *Tone*, *Speed*, and *Quality*. The observed indicators of the *Content* attribute include *Figurativeness*, *Interesting*, *Scientificity*, *Conciseness*, *Inspiring*, *Feedback*, *Sincerity*, and *Adaptability*. The judgment criteria of each refined index are also listed in the table. To observe the exogenous variables, each was measured by several synonymous but differently formulated items. The average value of the synonymous items was taken as the observed value of the exogenous variable.

For example, in the case of *Volume*, "The teacher's voice level is appropriate" and "The teacher can adjust the voice level appropriately to teach the lesson" are the same meaning, and using the average of the scores of these two items as the observation of *Volume*. The mean of these two items was used to observe *Volume* to show the subjects' thoughts more accurately. The researchers developed a 35-item (t1-t35) five-point Likert questionnaire based on the framework. The scores for each option were recorded as 5, 4, 3, 2, and 1 in descending order of satisfaction (approval).

This study was also a scale development study. Therefore, exploratory factor analysis and confirmatory factor analysis were used to explore the suitability of the data of the structure. The McDonald's omega coefficient was a substitute for Cronbach's alpha and could more accurately approach the scale's reliability (Peters, 2014). Therefore, McDonald's omega coefficient was used in this study.

TABLE 1 Framework for analyzing audible teaching language.

Attribute	Indicators	Standard	Items
Sound	Volume	Appropriate loudness of sound.	t1-t2
	Tone	Appropriate use of repetition, stress, and pauses.	t3-t5
	Speed	The speed of speech should be varied.	t6-t8
	Quality	Harmonious sound and a beautiful-sounding voice.	t9-t11
Content	Figurativeness	Flexible use of similes and metaphors to teach.	t12-t14
	Interesting	The teaching is humorous and interesting.	t15-t17
	Scientificity	The teaching language is strict and standardized, with proper use of terminology.	t18-t20
	Conciseness	Concise teaching without complicated language.	t21-t23
	Inspiring	Mobilize students' thinking in all aspects.	t24-t26
	Feedback	Provide timely feedback during teaching activities.	t27-t29
	Sincerity	Sincere and authentic feelings in teaching.	t30-t32
	Adaptability	Adapt to the needs of students' psychology.	t33-t35

4.2. Participants

The participants were 316 students from a key middle school in Changsha, Hunan Province, China. They came from three grades of senior high school. First of all, according to the purpose of the study, the researchers selected a novice, skilled, and expert mathematics teacher (a total of 9 mathematics teachers) from the first, second, and third grades of the senior high school and distributed questionnaires to the students in their class. Because there are no novice mathematics teachers in the third grade of senior high school, the researchers replaced the novice mathematics teachers in the third grade of senior high school with the novice mathematics teachers in the first year of senior high school. Table 2 demonstrates the professional titles and years of teaching experience of the teachers.

4.3. Procedure

This study was divided into five steps. In the first step, the researchers proposed a preliminary research question based on reflection on teaching practice, conducted a literature search and expert consultation, and determined the starting point of this study. In the second step, the researchers clearly proposed a research question based on theoretical research and practical needs, developed a research framework for the research question combined with existing research results, and revised and justified the research framework. In the third step, the researchers designed a questionnaire based on the research framework and conducted expert validation of the validity of the questionnaire. In the fourth step, the researchers recruited the participants according to the research plan, administered the questionnaire to the participants, collected the questionnaire, and transcribed the data. In the fifth step, the researchers analyzed the data, presented the results, and discussed and reflected on them concerning the existing relevant literature, pointing out the limitations of the study and research recommendations.

4.4. Data collection and analyzes

Data processing in this study included five processes: collection, transcription, processing, description, and analysis. For data collection, a questionnaire survey was adopted in this study. The researchers recruited potential participants, issued paper questionnaires, guided them to fill in the questionnaires as required, and collected the questionnaires. For the data transcription, after the quality of the paper questionnaires was preliminarily reviewed, invalid questionnaires were eliminated (those with missing values and those with all the same options were regarded as invalid questionnaires), and the data of valid questionnaires were input into EXCEL and SPSS 22.0. For data description, the researchers conducted descriptive statistical analysis on the primary attributes of all participants and presented preliminary descriptive statistical results. For data analysis, the researchers used exploratory factor analysis and confirmatory factor analysis to explore the suitability of the data of the structure. First, SPSS 22.0 were used to test the reliability and validity of the questionnaire. The researchers present evidence of the model's convergence and discriminant validity and calculate and report the average variance extracted and composite

TABLE 2 Mathematics teacher information.

Teacher	Title	Teaching years
Novice	None	1
Novice	Junior	2
Novice	Junior	3
Skilled	Intermediate	8
Skilled	Intermediate	9
Skilled	Intermediate	8
Expert	Senior	21
Expert	Senior	35
Expert	Senior	37

reliability values (Fornell and Larcker, 1981). Then, AMOS 22.0 was used for confirmatory factor analysis, and the fit of the measurement model was discussed under the maximum likelihood method. Based on the path coefficient significance test, the model was modified and verified by referring to the ideal value of the structural equation model. The final analysis model of high school mathematics teachers' satisfaction with audible teaching language was obtained. Finally, the paper presented the descriptive statistical results of high school students' satisfaction with the audible teaching language of the expert, skilled, and novice mathematics teachers and conducted the difference test.

5. Results

A total of 316 questionnaires were distributed, 291 were returned, and 263 valid questionnaires were obtained after excluding the missing values and selecting all the same options, with an effective rate of 90.4%. The following exploratory factor analysis results are based on the first 131 of the 263 valid data, and confirmatory factor analysis results are based on the last 132 of the 263 valid data.

5.1. Exploratory factor analysis

The researchers counted the scores of 12 exogenous variables (observed variables), and the results showed that the alpha reliability coefficient of the scores of 12 exogenous variables was 0.949, and overall, the reliability of this questionnaire was high. According to the correlation matrix information in Table 3, all the correlation values in the data are greater than 0.3, so the data satisfy the primary conditions for factor analysis.

The validity of the 12 exogenous variables data shows that the KMO value was 0.948, $p = 0.000 < 0.05$, and Bartlett's sphericity test was significant. Combined with the standard that the characteristic root is greater than 1, the explanation rate of the total variance, and the gravel diagram, this study extracted two factors according to the set research framework, see Table 4. The results show that the total variance is 71.849%.

The factor load matrix after rotation is shown in Table 5. All variables can be divided into the *Sound* element and the *Content* element. The *Sound* elements include *Volume*, *Tone*, *Speed*, and *Quality*. The *Content* elements include *Sincerity*, *Inspiring*, *Interesting*, *Adaptability*, *Figurativeness* and *Feedback*, *Scientificity*, and *Conciseness*.

In this study, the McDonald's Omega coefficient was used to describe the reliability of the tool. If this value is higher than 0.8, it means high reliability; if this value is between 0. 7and 0. 8, it is good; if this value is between 0. 6 and 0. 7, it means reliability is acceptable; if this value is less than 0. 6, it means that the reliability is not good. Standardized results showed that the overall McDonald's omega coefficient was 0.951, the McDonald's omega coefficient of sound dimension was 0.867, and the McDonald's omega coefficient of content dimension was 0.938. These values were all greater than 0.8, indicating that the research data have high reliability. In conclusion, the 12 exogenous variables were well-suited for factor analysis, and the validity of the questionnaire was high.

5.2. Confirmatory factor analysis

The validated factor analysis model was fitted using the maximum likelihood method. There was no negative error variance, the factor loading was between 0.5 and 0.95, and there was no large standard error, indicating the good intrinsic quality of the model.

The results of the validated factor analysis are shown in Figure 1. The Chi-square value of the prespecified model is 59.629, with a significance probability value of $0.086 > 0.05$, which does not reach the significance level and accept the null hypothesis.

In terms of model fitness statistical tests, $RMR = 0.020 < 0.050$, $CMIN/DF = 1.296 < 2.000$, $AGFI = 0.885$ (very close to 0.9), $RMSEA = 0.048 < 0.050$, $GFI = 0.932 > 0.900$, $NFI = 0.948 > 0.900$, $RFI = 0.925 > 0.900$, $IFI = 0.988 > 0.900$, $TLI = 0.982 > 0.900$, $CFI = 0.987 > 0.900$, $PRATIO = 0.697 > 0.500$, $PNFI = 0.660 > 0.500$, $PCFI = 0.688 > 0.500$ that meet the criteria for the model to be adaptable. The estimated parameters of the factor loadings all reach a significant level ($p < 0.05$). Therefore, the model has good fitness.

In this study, average variance extraction quantity and combinatorial reliability were used to analyze the convergence validity of the data. The average variance extracted (AVE) was more than 0.5, and the combination reliability (CR) was more than 0.60, which indicates that the results have high combination reliability and convergence validity. The average variance extracted (AVE) and combinatorial reliability (CR) of the latent variable *Sound* were $0.559 > 0.50$ and $0.831 > 0.60$, respectively. The average variance extracted (AVE) and combinatorial reliability (CR) of the latent variable *Content* were $0.597 > 0.50$ and $0.921 > 0.60$, respectively. It showed that the measurement model in this study had high combinatorial reliability and convergence validity.

In summary, the questionnaire designed in this study had good reliability and validity. It could be used to investigate high school students' satisfaction with mathematics teachers' audible teaching language.

5.3. Descriptive statistics of student satisfaction

The statistics of students' satisfaction with mathematics teachers' audible teaching language are shown in Table 6. From the results, it can be seen that students' satisfaction (i.e., the sum of the percentages of basically satisfied and very satisfied) with the use of mathematics

TABLE 3 Correlation matrix for the observed variables.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	1.000											
S2	0.621	1.000										
S3	0.527	0.637	1.000									
S4	0.536	0.717	0.619	1.000								
S5	0.452	0.605	0.639	0.734	1.000							
S6	0.449	0.603	0.589	0.730	0.686	1.000						
S7	0.491	0.539	0.570	0.586	0.493	0.510	1.000					
S8	0.560	0.633	0.622	0.665	0.593	0.562	0.693	1.000				
S9	0.477	0.610	0.663	0.671	0.675	0.737	0.641	0.702	1.000			
S10	0.492	0.612	0.645	0.658	0.611	0.642	0.682	0.667	0.791	1.000		
S11	0.378	0.509	0.545	0.601	0.580	0.669	0.598	0.541	0.697	0.593	1.000	
S12	0.518	0.650	0.627	0.697	0.700	0.706	0.631	0.705	0.769	0.680	0.697	1.000

TABLE 4 Total variance explained.

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative%
1	7.822	65.18	65.18	7.822	65.18	65.18
2	0.800	6.669	71.849	0.800	6.669	71.849
3	0.665	5.544	77.393			
4	0.453	3.773	81.167			
5	0.403	3.358	84.525			
6	0.364	3.033	87.558			
7	0.344	2.865	90.423			
8	0.311	2.595	93.018			
9	0.264	2.204	95.222			
10	0.224	1.867	97.089			
11	0.193	1.604	98.693			
12	0.157	1.307	100			

Extraction method: principal component analysis.

teachers' audible teaching language is high, reaching 95.44%. Students' satisfaction with the *Content* (95.06%) aspect of mathematics teachers' audible teaching language was slightly higher than that of *Sound* (93.54%).

As can be seen from Table 7, overall, students are the most satisfied (98.67%) with the skilled teachers' audible teaching language, followed by the expert teachers (96.71%) and the novice teachers last (91.75%). The satisfaction received by the skilled teachers is 1.96 percentage points higher than the expert teachers, the satisfaction received by the expert teachers is 4.96 percentage points higher than the novice teachers, and the satisfaction received by the skilled teachers is 6.92 percentage points higher than the novice teachers.

As can be seen from Table 8, in terms of *Sound,* students are the most satisfied with the expert teacher's audible teaching language (95.61%), followed by the skilled teacher (93.33%) and the novice teacher last (91.76%). In terms of *Sound* of audible teaching, expert teachers receive 2.28 percentage points higher satisfaction than skilled teachers, skilled teachers receive 1.57 percentage points higher satisfaction than novice teachers, and expert teachers receive 3.85 percentage points higher satisfaction than novice teachers.

Table 9 shows that in terms of *Content*, students' satisfaction with audible teaching language is the highest among skilled teachers (97.33%), followed by expert teachers (96.71%) and novice teachers (90.72%).

In terms of the *Content* of audible teaching, the satisfaction of experienced teachers is 0.62 percentage points higher than that of expert teachers, 5.99 percentage points higher than that of novice teachers, and 6.61 percentage points higher than that of skilled teachers.

5.4. Test of variance of student satisfaction

First, the differences in the audible teaching language of expert, skilled, and novice high school mathematics teachers were analyzed with the Kruskal-Wallis test. As can be seen from the summary of hypothesis testing in Table 10 (Shows asymptotic significance with a

TABLE 5 Rotated component matrix.[a]

Exogenous variable	Component	
	1	2
Sincerity	0.833	
Inspiring	0.814	
Interesting	0.812	
Adaptability	0.763	
Figurativeness	0.726	
Feedback	0.690	
Scientificity	0.558	
Conciseness	0.540	
Volume		0.887
Tone		0.726
Speed		0.577
Quality		0.527

Extraction method: principal component analysis. Rotation method: Varimax with Kaiser normalization. [a]Rotation converged in 3 iterations.

significance level of 0.05), it is found that $p < 0.05$ in terms of *Tone* and *Adaptability* of the audible teaching language and reject the original hypothesis. It can be concluded that the differences in *Tone* and *Adaptability* of expert, skilled, and novice high school mathematics teachers are statistically significant.

Table 11 reflects the situation of the rank of *Tone* and *Adaptability* in the audible teaching language. In terms of *Tone*, the rank means of novice, skilled, and expert mathematics teachers are increasing in order (118.84 < 130.29 < 147.43). Thus, students' satisfaction with the *Tone* of the audible teaching language of novice, skilled, and expert high school mathematics teachers increased sequentially. Regarding *Adaptability*, the rank means of novice, skilled, and expert mathematics teachers increase in descending order (115.80 < 135.53 < 146.53). So students' satisfaction with the *Adaptability* of novice, skilled, and expert high school mathematics teachers' audible teaching language increased in descending order.

Table 12 gives the results of the pairwise comparison of the rank means of expert, skilled, and novice high school mathematics teachers in terms of *Tone*.

From the results, it is clear that the rank mean of novice teachers minus the rank mean of expert teachers is−28.594 and the difference after the test is statistically significant ($p < 0.05$), indicating that students are more satisfied with expert teachers than with novice teachers in terms of *Tone*.

In terms of *Adaptability*, Table 13 gives the results of the pairwise comparison of the rank means of expert, skilled, and novice high school mathematics teachers. It is clear that the rank mean of novice teachers minus the rank mean of expert teachers is-30.548, and the difference after the test is statistically significant ($p < 0.05$), indicating that in terms of *Adaptability*, students are more satisfied with expert teachers than with novice teachers.

6. Discussion and enlightenment

This study explored students' satisfaction with the audible teaching language of high school mathematics teachers in the classroom. According to the research results, students were very satisfied with the audible teaching language of high school mathematics teachers. As the school of the participants is one of the best in the province and the professional quality of the teachers is very high, such results are expected. Overall, the students had the highest satisfaction with the skilled teachers' audible teaching language, followed by expert and novice teachers. Some research results show that novice and skilled math teachers have no significant differences in their understanding of problem solving and teaching as a whole but significant differences in individual aspects (Jiang et al., 2022). This study showed that the above characteristics of novice and non-novice teachers also existed in the audible teaching language of mathematics teachers. Although students' satisfaction with expert, skilled, and novice math teachers decreased successively in terms of the *Sound* of audible teaching language, and students' satisfaction with skilled, expert, and novice math teachers also decreased successively in terms of the *Content* of audible teaching language. There was no significant difference in the overall satisfaction obtained by expert, skilled, and novice high school math teachers, there were only differences in individual observed variables. For example, in terms of *Tone* and *Adaptability*, students were significantly more satisfied with the audible teaching language of expert math teachers than novice math teachers.

Skilled teachers received higher satisfaction than expert teachers, and expert teachers received higher satisfaction than novice teachers. It is easy to understand why expert teachers outperformed novice teachers in audible teaching language, but why did skilled teachers outperform expert teachers? It may be related to the evaluation and promotion of teachers, as skilled teachers face a critical period of promotion and are under pressure from family and career, so they will try to do their best in all aspects. It also verifies the research hypothesis that proper training can improve the quality of teachers' audio teaching language. Expert teachers, on the other hand, are under relatively less pressure. However, research findings have shown that the curriculum reform process has produced many regressive expert teachers who combine some typical novice elements and some typical expert elements in their professional practice (Liberman et al., 2012). It suggests that the motivation of expert teachers is necessary and that educational administrators should create opportunities for them to continue learning.

There was no significant difference between the expert and skilled teachers on each of the observed variables, nor was there a significant difference between skilled and novice teachers on each of the observed variables. In terms of *Tone*, there was a significant difference between expert and novice mathematics teachers, with expert teachers outperforming novice teachers. It is a significant result showing that novice mathematics teachers can improve the quality of the audible teaching language by improving the *Tone* of the audible teaching language in their teaching. In terms of *Adaptability*, expert mathematics teachers are more aware of students' needs and are able to use appropriate audible teaching language in their teaching. It may be because expert mathematics teachers are more student-centered and not overly focused on themselves. As noted, expert teachers always respond to students by asking questions, but novice teachers point out students' errors directly (Cai et al., 2022).

Teacher professional development is a complex, long-term process (Jin et al., 2021). Mature teachers have their own distinctive audible language styles, and these styles are not incompatible with each other, nor are they superior or inferior. In light of the above analysis, novice mathematics teachers (or pre-service mathematics

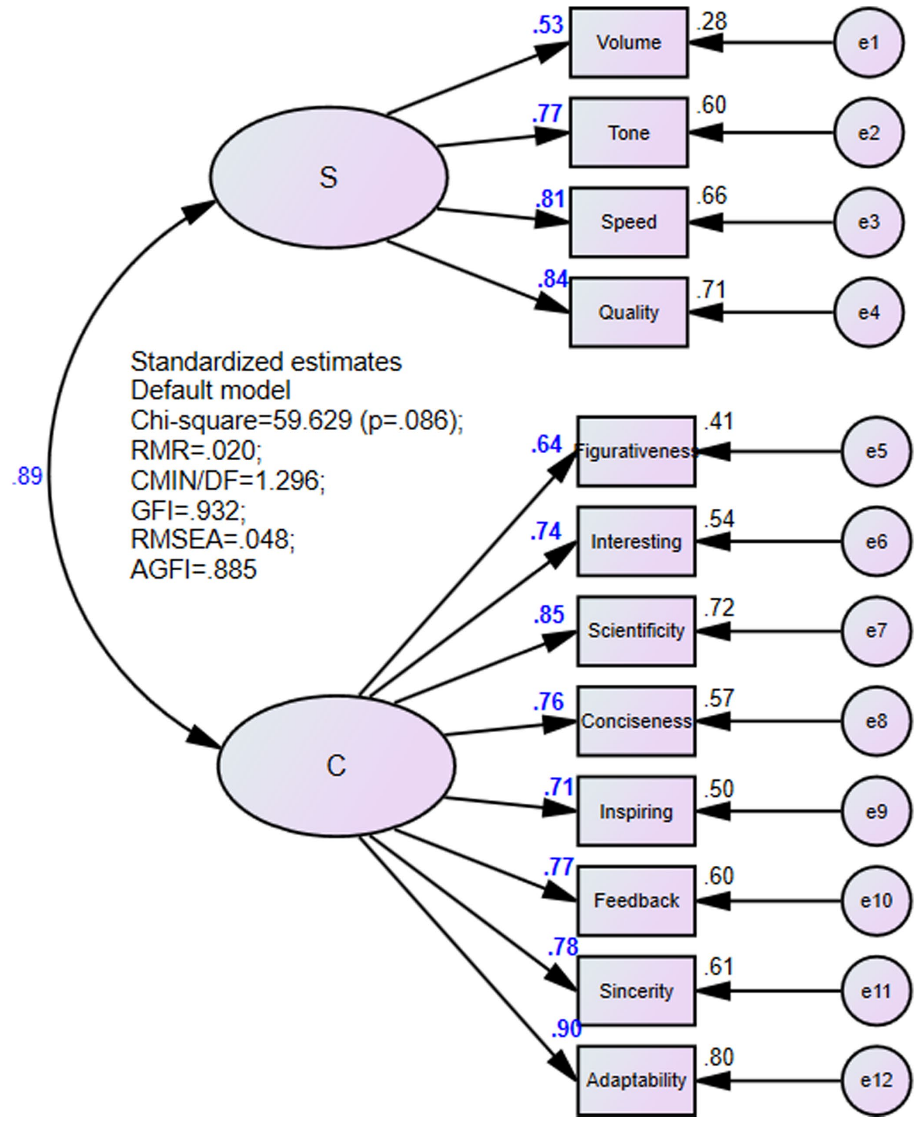


FIGURE 1
The adaptation result of the measurement model.

TABLE 6 Statistics of students' total satisfaction (*N*=263).

Satisfaction	Total		Sound		Content	
	N	Percentage	*N*	Percentage	*N*	Percentage
Very satisfied (4,5)	171	65.02%	165	62.74%	172	65.40%
Satisfied (3,4)	80	30.42%	81	30.80%	78	29.66%
Unsatisfied (1,3)	12	4.56%	17	6.46%	13	4.94%

TABLE 7 Students' satisfaction with expert, skilled, and novice teachers (*N*=263).

Satisfaction	Expert		Skilled		Novice	
	N	Percentage	*N*	Percentage	*N*	Percentage
Very satisfied (4,5)	67	73.63%	47	62.67%	57	58.76%
Satisfied (3,4)	21	23.08%	27	36.00%	32	32.99%
Unsatisfied (1,3)	3	3.30%	1	1.33%	8	8.25%

TABLE 8 Satisfaction with expert, skilled, and novice teachers in terms of *Sound* (*N*=263).

Satisfaction	Expert		Skilled		Novice	
	N	Percentage	*N*	Percentage	*N*	Percentage
Very satisfied (4,5)	65	71.43%	42	56.00%	60	61.86%
Satisfied (3,4)	22	24.18%	28	37.33%	29	29.90%
Unsatisfied (1,3)	4	4.40%	5	6.67%	8	8.25%

TABLE 9 Satisfaction with expert, skilled, and novice teachers in terms of *Content* (*N*=263).

Satisfaction	Expert		Skilled		Novice	
	N	Percentage	*N*	Percentage	*N*	Percentage
Very satisfied (4,5)	67	73.63%	45	60.00%	57	58.76%
Satisfied (3,4)	21	23.08%	28	37.33%	31	31.96%
Unsatisfied (1,3)	3	3.30%	2	2.67%	9	9.28%

TABLE 10 Kruskal-Wallis test for independent samples (expert, skilled, and novice teachers).

Original hypothesis	Sig.	Decision
The distribution of volume is the same	0.345	Retain the original hypothesis
The distribution of tone is the same	0.027	Reject the original hypothesis
The distribution of speed is the same	0.348	Retain the original hypothesis
The distribution of quality is the same	0.368	Retain the original hypothesis
The distribution of sound is the same	0.132	Retain the original hypothesis
The distribution of figurativeness is the same	0.173	Retain the original hypothesis
The distribution of interesting is the same	0.062	Retain the original hypothesis
The distribution of scientificity is the same	0.090	Retain the original hypothesis
The distribution of conciseness is the same	0.127	Retain the original hypothesis
The distribution of inspiring is the same in	0.472	Retain the original hypothesis
The distribution of feedback is the same in	0.078	Retain the original hypothesis
The distribution of sincerity is the same in	0.076	Retain the original hypothesis
The distribution of adaptability is the same	0.013	Reject the original hypothesis
The distribution of content is the same	0.137	Retain the original hypothesis
The distribution of total is the same	0.157	Retain the original hypothesis

TABLE 11 The rank of *Tone* and *Adaptability*.

	Group	*N*	Mean rank
Tone	1	97	118.84
	2	75	130.29
	3	91	147.43
Adaptability	1	97	115.80
	2	75	135.53
	3	91	146.35

TABLE 12 Paired comparison results (*Tone*).

Sample 1-Sample 2	Test statistic	Std. error	Std. test statistic	Sig.	Adj. sig
0–1	−11.453	11.278	−1.016	0.310	0.930
0–2	−28.594	10.704	−2.671	0.008	0.023
1–2	−17.141	11.439	−1.498	0.134	0.402

Each row tests the original hypothesis: sample 1 and sample 2 are equally distributed. Show asymptotic significance (2-sided test), and the significance level is 0.05.

teachers) should be trained to focus on the *Tone* and *Adaptability* of their audible teaching language. Pre-service teacher training programs should pay appropriate attention to the vocal aspects of instructional language. Teacher educators should provide training courses on the *Tone* and *Adaptability* of the audible teaching language so those novice mathematics teachers (or pre-service mathematics teachers) are genuinely concerned with the needs of their students, are student-centered, and teach by learning.

This study expands the knowledge in the field of teacher education and illustrates some characteristics of the audible teaching language of expert, skilled, and novice high school mathematics teachers. The study is also of great practical value,

TABLE 13 Paired comparison results (*Adaptability*).

Sample1-Sample2	Test statistic	Std. error	Std. test statistic	Sig.	Adj. Sig
0–1	−19.729	11.083	−1.780	0.075	0.225
0–2	−30.548	10.519	−2.904	0.004	0.011
1–2	−10.818	11.241	−0.962	0.336	1.000

Each row tests the original hypothesis: sample 1 and sample 2 are equally distributed. Show asymptotic significance (2-sided test), and the significance level is 0.05.

pointing out how differences should be overcome to improve the quality of the audible teaching language. However, the fact that the selected participants were from a top provincial high school, though exemplary, limits the generalization of the findings. Future studies should select a less affluent sample because the participants for this study were affluent students who attended a wealthy school.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Hunan Normal University. Written informed consent from the participants' legal guardian/next of Kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

PJ designed the study, developed the research tools, and wrote the paper. XZ collected data for this research. XR and ZF analyzed and processed the data. BX entirely directed this study and completed the correspondence. YJ assisted in communication correspondence and data analysis. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by a grant from the Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice 18dz2271000.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aziza, M. (2021). A teacher questioning activity: the use of Oral open-ended questions in mathematics classroom. *Qual. Res. Educ.* 10, 31–61. doi: 10.17583/qre.2021.6475

Bastian, A., Kaiser, G., Meyer, D., Schwarz, B., and König, J. (2022). Teacher noticing and its growth toward expertise: an expert-novice comparison with pre-service and in-service secondary mathematics teachers. *Educ. Stud. Math.* 110, 205–232. doi: 10.1007/s10649-021-10128-y

Bautista, A., and Roth, W. (2012). Conceptualizing sound as a form of incarnate mathematical consciousness. *Educ. Stud. Math.* 79, 41–59. doi: 10.1007/s10649-011-9337-y

Bekiari, A., Kokaridas, D., and Sakellariou, K. (2006). Associations of students' self-reports of their teachers' verbal aggression, intrinsic motivation, and perceptions of reasons for discipline in Greek physical education classes. *Psychol. Rep.* 98, 451–461. doi: 10.2466/pr0.98.2.451-461

Borko, H., and Livingston, C. (1989). Cognition and improvisation: differences in mathematics instruction by expert and novice teachers. *Am. Educ. Res. J.* 26, 473–498. doi: 10.3102/00028312026004473

Cai, J., LaRochelle, R., Hwang, S., and Kaiser, G. (2022). Expert and preservice secondary teachers' competencies for noticing student thinking about modelling. *Educ. Stud. Math.* 109, 431–453. doi: 10.1007/s10649-021-10071-y

Carlos Ramirez-Cruz, J., Marcos Lopez-Mojica, J., and Ake, L. P. (2018). Importance of mathematics in special education degree in their initial training. *ATENAS* 3, 100–114.

Carr, W., and Kemmis, S.. (1986). *Becoming Critical: Education, Knowledge and Action Research*. Philadephia: Falmer Press, 249.

Cheng, M. (2016). "Student satisfaction and quality education" in *Quality in Higher Education*, ed. Ming Cheng (Rotterdam: Sense Publishers), 33–45.

Chin, C. (2007). Teacher questioning in science classrooms: approaches that stimulate productive thinking. *J. Res. Sci. Teach.* 44, 815–843. doi: 10.1002/tea.20171

Cogan, M. L. (1958). The behavior of teacher and the productive behavior of their pupils: II. "Trait" analysis. *J. Exp. Educ.* 27, 107–124.

DeLiema, D., Enyedy, N., Steen, F., and Danish, J. A. (2021). Integrating viewpoint and space: how lamination across gesture, body movement, language, and material resources shapes learning. *Cogn. Instr.* 39, 328–365. doi: 10.1080/07370008.2021.1928133

Do, H. D., Wen, J. M., and Huang, S. K. (2022). Impact of humorous Chinese-teaching videos and Echo method on teaching Chinese language: a case study of Vietnamese students. *Asia Pac. Educ. Res.* 31, 725–738. doi: 10.1007/s40299-021-00622-5

Fornell, C., and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.1177/002224378101800104

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., and Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *J. Teach. Educ.* 60, 380–392. doi: 10.1177/0022487109339906

Gaver, J. J., and Gaver, K. M. (1993). Additional evidence on the association between the investment opportunity set and corporate financing, dividend, and compensation policies. *J. Account. Econ.* 16, 125–160. doi: 10.1016/0165-4101(93)90007-3

Gaythwaite, E. S. (2005). Didn't you see what I meant? *Curric. Teach. Dialogue* 7, 97–108.

Grifenhagen, J. F., and Barnes, E. M. (2022). Reimagining discourse in the classroom. *Read. Teach.* 75, 739–748. doi: 10.1002/trtr.2108

Gudelj, A., Ligere, J., Zaitseva-Pärnaste, I., and Załęska-Fornal, A. (2021). Survey of maritime student satisfaction: a case study on the international student survey to identify the satisfaction of students in mathematical course. *Педагогика* 93, 9–23. doi: 10.53656/ped21-6s.01sur

Guo, X., and Xia, L.. (2021). Discussion on the Innovative Approach to the Teaching Discourse System of Ideological and Political Academics. Atlantis Press: In *1st*

International Conference on Education: Current Issues and Digital Technologies (ICECIDT 2021), pp. 348–352.

Hale, A. J., Freed, J., Ricotta, D., Farris, G., and Smith, C. C. (2017). Twelve tips for effective body language for medical educators. *Med. Teach.* 39, 1–6. doi: 10.1080/0142159X.2017.1324140

Høynes, S.-M., Klemp, T., and Nilssen, V. (2019). Mentoring prospective mathematics teachers as conductors of whole class dialogues–using video as a tool. *Teach. Teach. Educ.* 77, 287–298. doi: 10.1016/j.tate.2018.10.014

Huang, R., and Li, Y. (2012). What matters Most: a comparison of expert and novice teachers' noticing of mathematics classroom events. *Sch. Sci. Math.* 112, 420–432. doi: 10.1111/j.1949-8594.2012.00161.x

Jiang, D., Chen, Y., and Garg, A. (2018). A hybrid method for overlapping speech detection in classroom environment. *Comput. Appl. Eng. Educ.* 26, 171–180. doi: 10.1002/cae.21855

Jiang, P., Zhang, Y., Jiang, Y., and Xiong, B. (2022). Preservice mathematics teachers' perceptions ofmathematical problem solving and its teaching: a case from China. *Front. Psychol.* 13:998586. doi: 10.3389/fpsyg.2022.998586

Jin, X., Li, T., Meirink, J., van der Want, A., and Admiraal, W. (2021). Learning from novice-expert interaction in teachers' continuing professional development. *Prof. Dev. Educ.* 47, 745–762. doi: 10.1080/19415257.2019.1651752

Kayima, F., and Mkimbili, S. T. (2021). How do chemistry teachers deal with students' incorrect/undesired responses to oral classroom questions? Exploring effective feedback practices. *Res. Sci. Educ.* 51, 647–668. doi: 10.1007/s11165-019-09876-y

Khoza, H. C., and Msimanga, A. (2021). Understanding the nature of questioning and teacher talk moves in interactive classrooms: a case of three South African teachers. *Res. Sci. Educ.* 52, 1717–1734. doi: 10.1007/s11165-021-10024-8

Kim, H. K. (2021). A case study on the satisfaction of mathematics online class and its relationship with mathematical learning in Corona-19. *Commun. Math. Educ.* 35, 341–358.

Kong, F. (2018). *A Comparative Study of Secondary School Biology Teachers' Voiced Teaching Language Behaviors*. Liaoning Normal University, Dalian, 10–11.

Lee, J. (2014). An exploratory study of effective online learning: assessing satisfaction levels of graduate students of mathematics education associated with human and design factors of an online course. *Int. Rev. Res. Open Distrib. Learn.* 15, 111–132. doi: 10.19173/irrodl.v15i1.1638

Lee, Y., and Kinzie, M. (2012). Teacher question and student response with regard to cognition and language use. *Instr. Sci.* 40, 857–874. doi: 10.1007/s11251-011-9193-2

Leinhardt, G., and Smith, D. A. (1985). Expertise in mathematics instruction: subject matter knowledge. *J. Educ. Psychol.* 77, 247–271. doi: 10.1037/0022-0663.77.3.247

Lester, F. (2013). Thoughts about research on mathematical problem-solving instruction. *Math. Enthus.* 10, 245–278. doi: 10.54870/1551-3440.1267

Liberman, N., Kolikant, Y. B.-D., and Beeri, C. (2012). "Regressed experts" as a new state in teachers' professional development: lessons from computer science teachers' adjustments to substantial changes in the curriculum. *Comput. Sci. Educ.* 22, 257–283. doi: 10.1080/08993408.2012.721663

Liu, Y., Wang, Y., Liu, R.-D., Ding, Y., Wang, J., and Mu, X. (2022). How classroom environment influences academic enjoyment in mathematics among Chinese middle school students: moderated mediation effect of academic self-concept and academic achievement. *Psychol. Res. Behav. Manage.* 15, 2035–2048. doi: 10.2147/PRBM.S371092

López, V., Salgado, M., and Berkowitz, R. (2022). The contributions of school and classroom climate to mathematics test scores: a three-level analysis. *Sch. Eff. Sch. Improv.* 1-22, 1–22. doi: 10.1080/09243453.2022.2096645

Mahir, N., Fikret, E. R., Demir, B., Erdogan, N. K., Sonmez, H., and Yilmaz, R. (2021). Satisfaction of open education students about the learning materials of mathematics. *Turk. Online J. Dist. Educ.* 22, 94–111. doi: 10.17718/tojde.906813

Meštrović, D. (2017). Service quality, students' satisfaction and behavioural intentions in STEM and IC higher education institutions. *Interdisciplinary Descr. Complex Syst.* 15, 66–77. doi: 10.7906/indecs.15.1.5

Miller, P. W. (2005). Body language in the classroom. *Techniques* 80, 28–30.

Mok, S. Y., Hämmerle, C. S., Rüede, C., and Staub, F. C. (2022). How do professional development programs on comparing solution methods and classroom discourse affect students' achievement in mathematics? The mediating role of students' subject matter justifications. *Learn. Instr.* 82, 101668–101613. doi: 10.1016/j.learninstruc.2022.101668

Nagy, J. T. (2018). Evaluation of online video usage and learning satisfaction: an extension of the technology acceptance model. *Int. Rev. Res. Open Distrib. Learn.* 19:2886. doi: 10.19173/irrodl.v19i1.2886

Ng, O.-L., Cheng, W. K., Ni, Y., and Shi, L. (2021). How linguistic features and patterns of discourse moves influence authority structures in the mathematics classroom. *J. Math. Teach. Educ.* 24, 587–612. doi: 10.1007/s10857-020-09475-z

Ngo, T., Unsworth, L., and Herrington, M. (2022). Teacher orchestration of language and gesture in explaining science concepts in images. *Res. Sci. Educ.* 52, 1013–1030. doi: 10.1007/s11165-021-10011-z

Ormond, C. A. (2021). Teaching classroom mathematics: linking two pedagogical models for promoting student engagement and conceptual connections. *Aust. J. Teach. Educ.* 46, 35–62. doi: 10.14221/ajte.2021v46n4.3

Panyajamorn, T., Suanmali, S., Kohda, Y., Chongphaisal, P., and Supnithi, T. (2018). Effectiveness of e-learning design in Thai public schools. *Malays. J. Learn. Instr.* 15, 1–34.

Peters, G. Y. (2014). The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha. *Eur. Health Psychol.* 16:576.

Pier, E. L., Walkington, C., Clinton, V., Boncoddo, R., Williams-Pierce, C., Alibali, M. W., et al. (2019). Embodied truths: how dynamic gestures and speech contribute to mathematical proof practices. *Contemp. Educ. Psychol.* 58, 44–57. doi: 10.1016/j.cedpsych.2019.01.012

Ruede, C. (2013). How secondary level teachers and students impose personal structure on fractional expressions and equations–an expert-novice study. *Educ. Stud. Math.* 83, 387–408. doi: 10.1007/s10649-012-9462-2

Scott, P. (1998). Teacher talk and meaning making in science classrooms: a Vygotskian analysis and review. *Stud. Sci. Educ.* 32, 45–80. doi: 10.1080/03057269808560127

Smart, J. B., and Marshall, J. C. (2013). Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science. *J. Sci. Teach. Educ.* 24, 249–267. doi: 10.1007/s10972-012-9297-9

Solomon, D., Rosenberg, L., and Bezdek, W. E. (1964). Teacher behavior and student learning. *J. Educ. Psychol.* 55, 23–30. doi: 10.1037/h0040516

Soysal, Y., and Yilmaz-Tuzun, O. (2021). Relationships between teacher discursive moves and middle school students' cognitive contributions to science concepts. *Res. Sci. Educ.* 51, 325–367. doi: 10.1007/s11165-019-09881-1

Sternberg, R. J., and Horvath, J. A. (1995). A prototype view of expert teaching. *Educ. Res.* 24, 9–17. doi: 10.3102/0013189X024006009

Stovner, R. B., and Klette, K. (2022). Teacher feedback on procedural skills, conceptual understanding, and mathematical practices: a video study in lower secondary mathematics classrooms. *Teach. Teach. Educ.* 110, 103593–103512. doi: 10.1016/j.tate.2021.103593

Suh, H., Bae, Y., Lee, J. S., and Han, S. (2018). The impacts of teacher education on students' academic achievement and satisfaction in mathematics lessons. *Math. Educ.* 57, 393–412.

Wang, Z., Pan, X., Miller, K. F., and Cortina, K. S. (2014). Automatic classification of activities in classroom discourse. *Comput. Educ.* 78, 115–123. doi: 10.1016/j.compedu.2014.05.010

Wasik, B., and Hindman, A. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *J. Educ. Psychol.* 103:455. doi: 10.1037/a0023067

Wolff, C. E., Jarodzka, H., and Boshuizen, H. P. A. (2021). Classroom management scripts: a theoretical model contrasting expert and novice teachers' knowledge and awareness of classroom events. *Educ. Psychol. Rev.* 33, 131–148. doi: 10.1007/s10648-020-09542-0

Woolfolk, A. E., and Galloway, C. M. (1985). Nonverbal communication and the study of teaching. *Theory Pract.* 24, 77–84. doi: 10.1080/00405848509543150

Yang, J., Zhu, F., Guo, P., and Pi, Z. (2020). Instructors' gestures enhance their teaching experience and performance while recording video lectures. *J. Comput. Assist. Learn.* 36, 189–198. doi: 10.1111/jcal.12397

Ye, L., Li, Y., and Si, H. (2015). Comparative study of the new and experienced teachers' differences about their language of instruction in the mathematics classroom. *J. Math. Educ.* 24, 40–43.

Zhao, W., Ma, J., and Cao, Y. (2022). What is effective classroom dialog? A comparative study of classroom dialog in Chinese expert and novice mathematics teachers' classrooms. *Front. Psychol.* 13:964967. doi: 10.3389/fpsyg.2022.964967

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 06 March 2023
DOI 10.3389/fpsyg.2023.1132184

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Slavoljub Jaroslav Hilcenko,
Vsovsu, Serbia
Yipeng Tang,
East China Normal University, China
Li Tong,
Chongqing Normal University, China

*CORRESPONDENCE
Tianzhuo Jiang
jiangtz674@nenu.edu.cn
Shuwen Li
lisw937@nenu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 27 December 2022
ACCEPTED 20 February 2023
PUBLISHED 06 March 2023

CITATION
Jiang T and Li S (2023) Secondary school students' use and perceptions of textbooks in mathematics learning: A large-scale investigation in China.
Front. Psychol. 14:1132184.
doi: 10.3389/fpsyg.2023.1132184

Secondary school students' use and perceptions of textbooks in mathematics learning: A large-scale investigation in China

Tianzhuo Jiang* and Shuwen Li*

School of Mathematics and Statistics, Northeast Normal University, Changchun, China

Students' use of textbooks is the key link of students engaged and learned curriculum and has received much attention recently. However, existing studies were mainly case studies or small-scale investigations and few addressed the context of China. Hence, this study provided a general overview of mathematics textbook use by Chinese secondary students through a large-scale investigation. Using a mixed-method approach, we collected the quantitative data from 2,145 students in eight provinces through a questionnaire survey and the qualitative data from 20 students and 8 teachers by the interviews. The results revealed that (1) Chinese students relied heavily on mathematics textbooks and pointedly used a portion of components in textbooks, mainly kernels, examples, and exercises; (2) Chinese students used mathematics textbooks for various but typical reasons, particularly to understand basic knowledge and skills, and showed self-regulation and teacher-mediation behind their use; and (3) Chinese students had a positive view about textbook use in mathematics learning, especially in developing mathematical knowledge, skills, and abilities. Furthermore, there were significant differences in mathematics textbook use among different students in terms of school regions, grade levels, and teachers' demographic variables. Finally, explanations and implications of the results were discussed.

KEYWORDS

textbook use, mathematics textbooks, mathematics learning, curriculum resources, secondary school students, Chinese mathematics education

1. Introduction

As the main vehicles of curriculum content and the key resources for teaching and learning, textbooks have always been a hot topic in educational research. In recent decades, various issues related to mathematics textbooks have been researched and discussed, including their composition, use, and history (Schubring and Fan, 2018). Among the issues, research on mathematics textbook use has received increasing attention and become an important part of topic study groups of the 13th (Fan et al., 2018) and 14th International Congress on Mathematical Education and one of the hot topics of the 2nd (Schubring et al., 2018), 3rd (Rezat et al., 2019) and 4th International Conference on Mathematics Textbook Research and Development.

As educational researchers have increasingly recognized the important role that textbooks play in mathematics teaching and learning, how mathematics textbooks are incorporated into teaching and learning has been explored. With regard to teachers' use of mathematics textbooks, numerous studies have been conducted from various perspectives, including but not limited to offering a framework for characterizing and studying teachers' interactions with curriculum materials (Remillard, 2005), examining how teachers used their curriculum resources to teach new mathematics standards (Polly, 2017), and so on (Nicol and Crespo, 2006; Grave and Pepin, 2015; Olsher and Cooper, 2021). On the contrary, studies on students' use of mathematics textbooks have not been paid adequate attention, with the results being largely discrete and non-inclusive (Wang and Fan, 2021, p. 2). Meanwhile, it is clear that most of previous studies were carried out in the broad international context but in a small scale (Fan et al., 2013; Rezat, 2013) and little research focused Chinese context. Moreover, researchers have discussed some factors that influence students' textbook use, but there is a lack of comprehensive research regarding the demographic and contextual factors influencing students' use of mathematics textbooks.

In this study, we attempt to provide an overview of Chinese secondary students' use and perceptions of mathematics textbooks and its factors through a large-scale investigation. We believe that focusing on Chinese students' use of mathematics textbooks can make a distinctive and important contribution to research on students' use of textbooks. First, China is the most populous developing country with a large number of students in junior high schools (Ministry of Education of the People's Republic of China, 2022b) in the world. Second, the tradition of Chinese culture, especially Confucian culture, has played an important role in modern mathematics education. Third, the impacts of textbook use on students' mathematics learning might, to some extent, offer an explanation for the well-known fact that Chinese students have outperformed their western counterparts in the international comparative studies of mathematics achievements such as the Trends in International Mathematics and Science Study and the Program for International Student Assessment (Lin et al., 2018). Thus, this study proposed the following questions:

(1) What is the current status of Chinese students' use of mathematics textbooks?
(2) Are there differences among Chinese students in terms of school regions, student genders, grade levels, and teachers' demographic variables in mathematics textbook use?

2. Literature review

2.1. Students' use of mathematics textbooks

Students' use of mathematics textbooks is common, but the possible relationship between students and mathematics textbooks is complex and dynamic. According to existing research, text-reading theory, reader-oriented theory, and activity theory have generally been used to analyze students' use of mathematics textbooks. Based on text-reading theory, the "use" is understood as reading, a transaction between mathematics textbooks (written curriculum) and students (Berger, 2019). In reader-oriented theory, the student is viewed as actively constructing meaning from mathematics textbooks through the reading process, which is shaped and constrained by the intentions of the author, the beliefs of the reader, and the qualities the text requires the reader to possess (Weinberg and Wiesner, 2011). From the perspective of activity system, students' textbook use refers to the activities that are primarily associated with textbooks in students' mathematics learning, such as reading and practicing (Rezat and Sträßer, 2013; Wang and Fan, 2021). To capture current situation of Chinese students' use of mathematics textbooks, this study employed the definition of students' textbook use based on activity theory. Activity theory is a descriptive psychological theory that studies and explains the emergence and development of human psychology with activity as its logical starting point and central category. Activity theory has its roots in the classical philosophy of Kant and Hegel, the contemporary philosophy of Marx and Engels, and the cultural-historial psychology of Vygotsky, Leontev, and Luria (Jonassen and Rohrer-Murphy, 1999). As a powerful psychological framework rather than a methodology, the six elements in activity theory: subject, artifact, object, community, rules, and division of labor directed this study. Specifically, we focused on how students interact with mathematics textbooks in social contexts by situating students, textbooks, and mathematics into the subject-artifacts-object triangle (Vygotsky, 1978). Meanwhile, we considered the influence of the region (social contexts), students' gender and grade level (subject), and teachers' variables (community, rules, and division of labor) on students' use of mathematics textbooks.

As mentioned earlier, several studies have been conducted from different facets to investigate how students used mathematics textbooks. Fan et al. (2004) focused on the following aspects of students' textbook use, including (1) the frequency and timing of textbook use, (2) how students used different textual parts, (3) to what extent students thought that textbook use was important for learning mathematics, and (4) whether and why students changed their ways of textbook use compared to the last semester. They found that textbooks were students' main learning resources for both in-class exercises and homework. Besides, Rezat's series of studies dealt with students' use of mathematics textbooks at school level. Rezat (2011) presented five self-regulated learning activities that involved textbooks in learning mathematics: (1) task and problem solving, (2) practice, (3) acquisition of new knowledge, (4) interest-driven activities, and (5) meta-cognitive learning activities. Meanwhile, he (2013) also conducted a study on how 74 students in two German secondary schools used their mathematics textbooks for practice. In Weinberg et al.'s 2012 study, students in introductory mathematics courses were surveyed to answer which parts of the texts they used, and when and why they used textbooks. Moreover, Thomas (2013) examined how students in two classrooms taught by the same teacher used print and digital formats of the Algebra 1 textbook. The results indicated that most students used a small portion of the resources and features in the textbook, tended to view the textbook primarily as a source for homework, and rarely bothered to develop examples and texts in class. From a comparative perspective, Wang and Fan (2021) proposed seven indicators to investigate the use of mathematics textbooks by students in Shanghai and England, including the frequency, duration, timing, purpose, and motivation of textbook use, access to textbooks, and the influence of textbook use on mathematics learning.

In summary, most studies focused on how students used mathematics textbooks in English-speaking countries and few addressed the context of China. Moreover, existing studies were mainly case studies or small-scale investigations. Meanwhile, various indicators related to the word "use" were proposed to embody students' use of mathematics textbooks. Against this background, this study aims to conduct a large-scale investigation of Chinese secondary students' use and perceptions of mathematics textbooks. According to Jonassen and Rohrer-Murphy's (1999) process for applying activity theory, we focused on when and why activities occur, what are used to perform activities, and what are the outcomes of activities. And based on previous studies, we established a conceptual framework (**Table 1**) consisted of four subdimensions with seven indicators to capture how Chinese students use mathematics textbooks.

The first four indicators are proposed to understand to what extent students rely on mathematics textbooks in their learning. Two of the indicators are the days of textbook use in a week and the duration of textbook use in a day. Meanwhile, the frequency of textbook use at different timing refers to the frequency of students' textbook use before, in, and after class and before the examination. The frequency of using different components refers to the frequency of using "introductions," "exploratory tasks," "kernels (definitions, theorems, and formulas)," "worked examples," "tips," "exercises and problems," "summaries," "mathematics activities," and "reading materials" which were divided based on textbook editors' suggestions.

Furthermore, this study defines the purpose of textbook use as specific learning activities related to mathematics textbooks, which include "preview," "revision," "doing homework," "doing extra exercises," "in-class learning and exercises," "looking up examples, answers, and references," and "looking up definitions, theorems, and formulas" (Weinberg et al., 2012; Rezat, 2013; Wang and Fan, 2021). Moreover, the indicator "motivation" is proposed to understand the reasons behind textbook use, which was drawn upon Amabile et al.'s (1994) scale, Ryan and Deci's (2000) definition and structure, and Wang and Fan's (2021) construct. The motivation consists of intrinsic motivation (enjoyment and challenge) and extrinsic motivation, which is composed of two aspects: external regulation (teacher-mediation and parent-supervision) and self-regulation (students' recognition of the impacts of textbook use on "grades," "knowledge, skills, and abilities," and "thinking methods, activity experience, and emotions and values").

TABLE 1 Subdimensions and indicators of current situation of students' textbook use.

Subdimensions	Indicators
Length of time of textbook use	The days of textbook use in a week
	The duration of textbook use in a day
Frequency of textbook use	The frequency of textbook use at different timing
	The frequency of using different components
Reasons of textbook use	The purposes of textbook use
	The motivation of textbook use
Perceptions of textbook use	The recognition of the impacts of textbook use on students' mathematics learning

In addition, students' perceptions of textbook use are adopted to understand to what extent students recognize the impacts of textbook use on mathematics learning. Previous research has explored the effects of curriculum resources as instruments on students' achievement (Van den Ham and Heinze, 2018; Sievert et al., 2021), conceptual understanding (Rezat, 2021), beliefs (Moyer et al., 2018; Kersey, 2019), identities (Macintyre and Hamilton, 2010), and levels of participation (Ewing, 2006). Also based on the aims of Mathematics Curriculum Standards for Compulsory Education (Ministry of Education of the People's Republic of China, 2011, 2022a), this study reflects the influence of textbook use on mathematics learning in whether textbook use helps students improve mathematics grades, master mathematics knowledge, enhance mathematical skills, develop mathematical abilities, understand mathematical thinking methods, gain mathematical activity experience, and shape mathematical emotions and values.

2.2. Factors on students' use of mathematics textbooks

Students' use of mathematics textbooks is a learning activity which is influenced by both internal and external factors of activity. Internal factors of activity consists of student factors and textbook factors. For student factors, Fan et al. (2004) found the majority of students (78%) have changed their ways of textbook use from the first year to the second year in junior high schools. One of the main reasons for the change was that students realized mathematics more important for them. Students' grade levels and beliefs of mathematics become important factors which influence students' use of mathematics textbooks. Regarding textbook factors, students' use of mathematics textbooks is influenced by content, structure and formats of textbooks (Jukić Matić and Glasnović Gracin, 2016). Österholm (2006) recruited 61 secondary and 34 university students to compare their reading comprehension of one historical text and two mathematical texts, both of which presented basic concepts of group theory, but one did it using mathematical symbols, whereas the other only used natural language. The vertical comparison of the students' prior knowledge with their results in the reading test revealed a similarity in reading comprehension between the mathematical text without symbols and the historical text as well as a difference between the two mathematical texts. Similarly, Rezat's (2013) study showed how textbook users were influenced by the way mathematics was presented in the textbooks. Besides, Thomas (2013) identified there were differences between digital and print textbook use by students in two classrooms taught by the same teacher.

Among external factors of activity, country or region factors and teacher factors are the important factors behind students' use of mathematics textbooks. In their comparative study, Wang and Fan (2021) found that there were significant differences between Shanghai and England regarding the role that textbooks played as curriculum resources in students' mathematics learning. Meanwhile, Fan et al. (2004) found there were significant differences in students' use of mathematics textbooks, especially the frequency of using different components, between two regions in China. Furthermore, teachers' mediated intervention plays an important role in students' use of mathematics textbooks (Griesel and Postel, 1983; Rezat, 2006, 2009). Although many instructors might not clearly tell their students how to use the textbook, students reported that they used it more productively when they believed they had been asked to do so (Weinberg et al., 2012). Teachers as the mediators decided which textbooks to use; when and where the textbook was to be used; which sections of the textbook to use; the sequencing of topics in the textbook; the ways in which students engaged with the text; the level and type of teacher intervention between students and textbooks; and so on (Pepin and Haggarty, 2001). Rezat (2012) summarized the conceptualization of six different ways teachers mediate textbook use in matrix and stated that all three dimensions are intertwined in a concrete mediation of textbook use.

Summarily, students' use of mathematics textbooks is affected by various factors, including students' grade levels, text formats, school regions, and teachers. However, little research has taken the influence of teachers' demographic variables (teachers' gender, education level, title, teaching experience, and experience in teaching with textbooks) on students' use of mathematics textbooks into consideration. Hence, this study intends to present a systematic investigation of the factors that influence Chinese secondary students' textbook use in learning mathematics.

TABLE 2 The profiles of students surveyed.

Variables		Categories	N	Percentage
Region		East	728	33.94%
		Middle	653	30.44%
		West	764	35.62%
Gender		Male	1106	51.56%
		Female	1039	48.44%
Grade level		7th grade	747	34.83%
		8th grade	902	42.05%
		9th grade	496	23.12%
Students taught by different teacher groups	Gender	Male	285	13.29%
		Female	1860	86.71%
	Education level	Bachelor's degrees or lower	1468	68.44%
		Master's degrees or higher	677	31.56%
	Title	Primary	1169	54.50%
		Middle	390	18.18%
		Senior	586	27.32%
	Teaching experience	≤5 years	1230	57.34%
		6–15 years	143	6.67%
		>15 years	772	35.99%
	Experience in textbook teaching	≤5 years	1230	57.34%
		>5 years	915	42.66%

TABLE 3 Basic information of teachers interviewed.

	Gender	Level of education	Title	Experience of teaching	Experience of teaching with textbooks
T1	Male	Bachelor	Primary	≤5 years	≤5 years
T2	Female	Master	Primary	≤5 years	≤5 years
T3	Female	Bachelor	Primary	≤5 years	≤5 years
T4	Female	Master	Middle	6–15 years	>5 years
T5	Male	Master	Middle	6–15 years	>5 years
T6	Male	Master	Senior	>15 years	>5 years
T7	Male	Bachelor	Senior	>15 years	>5 years
T8	Female	Bachelor	Senior	>15 years	>5 years

3. Materials and methods

3.1. Instruments

In this study, mixed methods were used to collect data on students' textbook use in learning mathematics through a questionnaire survey and the interviews with students and teachers.

3.1.1. Questionnaire

The questionnaire, designed on the basis of the conceptual framework, consisted of five parts with 27 questions. The first part related to students' demographic information, including region, gender and grade level. The second part was about the days of textbook use in a week and the duration of textbook use in a day. The options for two questions were intervals. The third part contained 13 questions, four of which referred to the frequency of textbook use before, in, and after class and before the examination and another nine of which referred to the frequency of using "introductions," "exploratory tasks," "kernels," "worked examples," "exercises and problems," "tips," "summaries," "mathematics activities," and "reading materials." A five-point Likert scale was used to capture students' options (e.g., never, seldom, sometimes, often, always). In the fourth part, two questions focused on the reasons for textbook use: one related to specific activities students engaged in when learning mathematics with textbooks and the other involved possible motivations behind textbook use. Students were asked to select all options that applied to their purposes and motivation of textbook use. Finally, the remaining seven questions addressed the impacts of textbook use on improving mathematics grades, mastering mathematics knowledge, improving mathematical skills, developing mathematical abilities, understanding mathematical thinking methods, gaining mathematical activity experience, and shaping mathematical emotions and values. A five-point Likert scale was employed to gather students' recognition of the impacts of textbook use on their mathematics learning (e.g., not helpful, not very helpful, slightly helpful, helpful, very helpful). After the questionnaire was drafted, a panel of mathematics education researchers was invited to review it and they were highly positive about the instrument, indicating reasonable validity. Meanwhile, the reliability test of the questionnaire determined using Cronbach's alpha yielded a value of 0.950, indicating high reliability.

3.1.2. Interview outline

Furthermore, we conducted interviews with students and teachers to complement the questionnaire data and gather more in-depth details about how students use mathematics textbooks. The outline of student interview was in line with the four subdimensions of the questionnaire. The first part included two questions: (1) How frequently and when do you use textbooks in your mathematics learning? What other curriculum resources do you use to learn mathematics? (2) Do you think that textbook use is helpful in learning mathematics? And why? The second part consisted of two different situations: in class and out of class (mainly at home), but with the same three questions: (1) Do you use the textbook at the request of teachers/parents or on your own initiative? (2) Which parts of the textbook do you use? (3) For what purposes do you use the text components? According to Jukić Matić and Glasnović Gracin (2016) questions and Rezat's (2012) conceptualization, the protocol of teacher interview was developed with three questions: (1) How do you usually prepare a mathematics lesson? (2) What proportion of your teaching content in class comes from mathematics textbooks? And why are the contents in textbooks added, deleted and adjusted? (3) Whether you ask students to use textbooks inside and outside mathematics classrooms? Which parts and why are students asked to use?

3.2. Data collection and analysis

3.2.1. Data collection

After obtaining ethical approval to conduct the research, students who participated in the questionnaire were selected through multi-stage sampling. In the first stage, according to stratified sampling, three provinces were separately selected from eleven provinces in the east and twelve provinces in the west, and two provinces were selected from eight provinces in the midlands. Secondly, one city from each sampled province was selected by cluster sampling. In the third phase, one school was randomly selected in each city. Finally, at least one class of each grade level was selected in sampled schools. At the same time, we also collected the basic information of mathematics teachers who taught the sampled students. We distributed 2,300 questionnaires to all participating students and obtained 2,145 valid questionnaires, which was a response rate of 93.3%. **Table 2** shows the profiles of the participating students.

For an in-depth understanding of the quantitative data, we randomly selected 20 students not involved in the questionnaire to conduct focus group interviews online and selected 8 their mathematics teachers by purposive sampling to conduct individual interviews on line or by phone calls. The interviewed students and teachers covered different comparison groups to ensure representativeness. Specifically, **Table 3** shows the basic information of teachers interviewed and among the 20 students, 8 were males and 12 were females; 7 were from the east, 7 were from the midlands, and 6 were from the west; 8 were from the 7th grade, 9 were from the 8th grade, and 3 were from the 9th grade. During the whole process of survey and interviews, we keep the data strictly confidential and anonymous to fully protect the privacy of all participating students and teachers.

3.2.2. Data analysis

The questionnaire data were analyzed by descriptive and inferential statistics, and presented numerically and graphically. We first calculated the percentage or mean and standard deviation for each item in the questionnaire. To explore factors influencing students' use of mathematics textbooks, we distinguished several comparison groups of students in terms of school regions, student genders, grade levels, and teachers' demographic variables. Then, we conducted T-tests, One-way analysis of variance (ANOVA), or Chi-square tests on each item to examine whether there were statistically significant differences in textbook use among students of different groups.

For the qualitative data from the interviews, we first coded the interviewed students as S1 to S20 and interviewed teachers as T1 to T8 to protect their privacy. Then, two researchers conducted independent transformation and interpretation of interview transcripts by identifying different indicators of textbook use to obtain more information about students' textbook use.

4. Results

4.1. Current situation of students' textbook use

4.1.1. Students' length of time of mathematics textbook use

Figure 1 shows the days students used mathematics textbooks in a week and the duration students used mathematics textbooks in a day. Nearly half (48.9%) of students used textbooks at least 5 days per week, 40% used textbooks 1–4 days per week, and a minority (11.1%) didn't use textbooks. Meanwhile, the majority (63.5%) of students used textbooks more than 15 min per day and 25.4% used textbooks within 15 min per day. During the interviews, 14 students reported that they used textbooks more than 5 days per week and another 6 students used textbooks 1–4 days per week. Meanwhile, 3 students reported that they used textbooks more than 30 min per day, 9 students used textbooks 16–30 min per day, and the remaining used textbooks less than 15 min. The results revealed that Chinese students used mathematics textbooks for many days per week and longtime per day, indicating that they relied heavily on mathematics textbooks.

4.1.2. Students' frequency of mathematics textbook use

As shown in Figure 2, about 15% of students never used mathematics textbooks before, in, and after class and before the examination. And about half of students at least often used textbooks in class (51.7%) and before the examination (47.2%), while more than half of students seldom or sometimes used textbooks before (56.0%) and after (57.8%) class. In the interviews, 12 students reported that they at least often used textbooks in class. Meanwhile, 16 students stated that they sometimes or often used textbooks before and after class, which appeared more frequent use than the questionnaire data. In contrast to the data surveyed, only seven students reported that they sometimes or often used textbooks before the examination. Merely one student (S4) said: "I always used textbooks before the final and midterm examinations and weekly tests." The results revealed that Chinese students used mathematics textbooks more frequently in class and before the examination than before and after class.

According to Figure 3, about 10% of students always used introductions, exploratory tasks, tips, mathematics activities, and reading materials, and 15–20% of students never used these parts. However, about half of students at least often used kernels, worked examples, exercises and problems, and summaries, and about 15% of students never used these parts. During the interviews, all students stated that they at least sometimes used kernels, worked examples, and exercises and problems. Meanwhile, exploratory tasks, tips, summaries, and reading materials were rarely mentioned (less than 5 students, respectively), and introduction and mathematics activities were not mentioned at all. The results revealed that Chinese students used core content in mathematics textbooks more frequently than content with guidance and auxiliary.

4.1.3. Students' reasons of mathematics textbook use

As shown in Figure 4, mathematics textbooks were typically used by more than half of Chinese students to preview, revise, do homework, learn and exercise in class, and look up definitions, theorems, and formulas. Consistent with the data surveyed, all interviewees mentioned the above five purposes of textbook use. Besides, 11 interviewed students used textbooks to look up examples and 3 interviewed students used textbooks to do extra exercises. The four typical motivations behind textbook use in turn were their perceptions of textbook's helpfulness for "knowledge, skills, and abilities" and "thinking methods, activity experience, and emotions and values," teacher-mediation, and challenges. In the

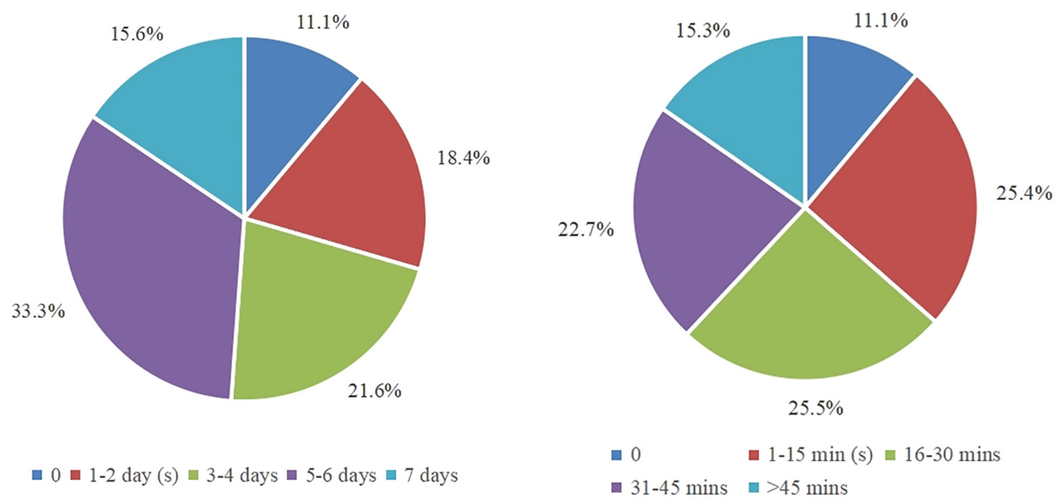


FIGURE 1
Percentage distributions of the days and duration of textbook use.

FIGURE 2
Percentage distributions of the frequency of textbook use at different timing.

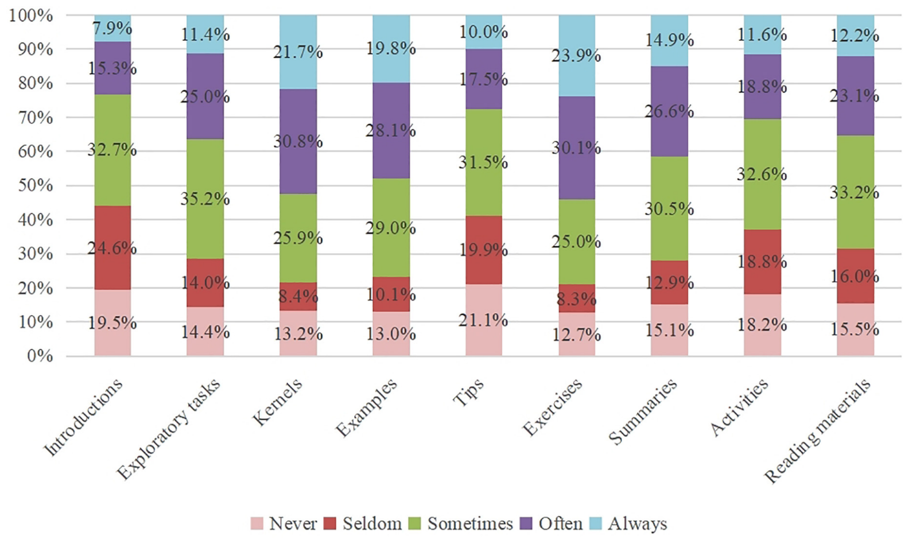


FIGURE 3
Percentage distributions of the frequency of using different components.

interviews, all students pointed out that textbook use is helpful for their mathematics learning and their teachers asked them to use textbooks in class, but no one mentioned that they would feel fulfilled when they solved difficult problems in textbooks. Besides, six students reported that their parents asked them to use textbooks. For example, S18 said: "My mother told me 'I was too busy to tutor you, so you should use textbooks more because there were all you have ever studied in textbooks.'" The results revealed that Chinese students used mathematics textbooks for various but typical purposes and motivation.

4.1.4. Students' perceptions of mathematics textbook use

Table 4 shows the means, standard deviation, and rank of the influence of textbook use on mathematics learning. The mean scores of seven items ranged from 3.502 to 3.992, which were greater than 3.500, implying positive impacts of textbook use on mathematics learning. Textbook use is perceived to be most helpful for mathematics knowledge, followed by mathematical skills, abilities, thinking methods, activity experience, values and emotions, and grades. During the interviews, all students stated that textbook use is helpful for developing mathematical knowledge, skills, and abilities. Besides, 4 students said that "When I met the problems that I could not solve, I would read the examples and exercises to find the methods." One student (S3) emphasized that "I would lose marks because of incomplete steps, so I would often read and imitate the steps of the examples in mathematics textbooks, which could help me reduce the loss of marks." While, no one explicitly mentioned that textbook use is helpful for gaining activity experience and shaping emotions and values. The results revealed

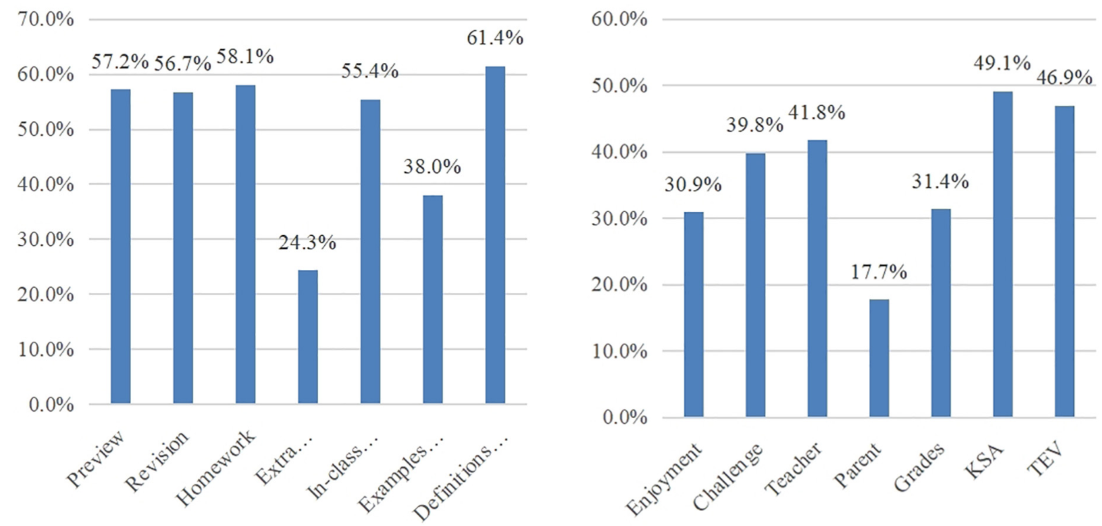


FIGURE 4
Percentage distributions of the purposes and motivation of textbook use.

TABLE 4 The M, SD, and rank of the influences of textbook use on mathematics learning.

	Grades	Knowledge	Skills	Abilities	Methods	Experience	Emotions
M	3.502	3.992	3.595	3.570	3.546	3.539	3.509
SD	1.0112	0.9622	1.0364	1.0780	1.0424	1.0525	1.0940
Rank	7	1	2	3	4	5	6

that Chinese students had a positive view about textbook use in mathematics learning, especially in developing basic knowledge, skills, and abilities.

4.2. Analysis of differences across student groups

To examine the effects of different factors on students' use of mathematics textbooks (see Table 5), Chi-square tests were conducted for the length of time and reasons of textbook use among different comparison groups. In frequency and perceptions of textbook use, *T*-tests were conducted across students' gender, teachers' gender, education level and experience in teaching with textbooks and ANOVA tests were conducted across school region, grade level, and teachers' title and teaching experience. Meanwhile, the interviews with students and teachers were conducted to offer explanations for the differences in students' textbook use across different comparison groups.

4.2.1. School region: Access to and types of other curriculum resources

The test results showed that there were statistically significant differences in the length of time, frequency, reasons, and perceptions of textbook use among the three regions. Specifically, students from the midlands used textbooks significantly more time, more frequently, and more for typical reasons and thought significantly more highly of textbooks in mathematics learning than those from the west, whereas students from the west were significantly greater than those from the east in four dimensions of textbook use. During the interviews, we found that the region differences in textbook use were related to access to and types of other curriculum resources. According to students' responses, other curriculum resources could be classified into information and communication technologies (ICTs), supplementary educational books, and school-based learning materials and their deciders could be the publisher, local teaching and research section, school, teacher, parent, or student (see Table 6). The interview data revealed that students from the east had more access to other more curriculum resources, so that they used textbooks less time, less frequently, and less for typical reasons and thought less highly of textbooks in mathematics learning than those from the west, whereas students from the west were less than those from the midlands in four dimensions of textbook use.

4.2.2. Student genders

The test results showed that there were no statistically significant differences in the length of time and frequency of textbook use between the two genders, except that boys used textbooks significantly more minutes per day than girls and used introductions in mathematics textbooks significantly more frequently than girls. Also, there were no statistically significant differences in reasons and perceptions of textbook use between the two genders, except that it is significantly more enjoyable for boys to use textbooks than girls. Similarly, we have not found that there were evident differences in textbook use between the two genders in the interviews.

4.2.3. Grade level: Curriculum content and students' mathematics knowledge base

The test results showed that there were statistically significant differences in the length of time, frequency, reasons, and perceptions of textbook use among the three grade levels. Specifically, the seventh graders used textbooks significantly more time, more frequently, and more for typical reasons and thought significantly more highly of textbooks in mathematics learning than the eighth graders, whereas the eighth graders were significantly greater than the ninth graders in four dimensions of textbook use. According to the interviews, we thought the grade differences in textbook use were due to curriculum content and students' mathematics knowledge base.

For example, S9 said: "The content in the seventh grade is relatively simple and some have been learned in primary school. When in the eighth grade, the amount and difficulty of curriculum content are grater. So I often used the textbook to look up definitions, theorems, and formulas." Also, S1 and S14, the eighth graders, expressed the same views. Meanwhile, S10 said: "The content in the ninth grade is more compositive. When learned new content or solved the problems, I need more prior knowledge than before. So I often used the textbook to review."

S5, S6, and S8 said: "When learned parallel lines, congruent triangles, and parallelogram, I used the textbook more frequently to look up examples. Because I had to imitate the steps of proof to do homework and further understand reasoning."

S15 and S17 said: "Compared to the last semester, I think I have understood and mastered the content after learning in class and doing homework (not in the textbook) after class. So I hardly used the textbook and occasionally used the textbook only when I cannot deal with the problems well."

The results of the above interviews revealed that breadth and difficulty of curriculum content, specific content areas, and students' mathematics knowledge base were the factors that affected students' textbook use. Specifically, as the grade changed, students used textbooks more time, more frequently, and more for typical reasons and thought more highly of textbooks in mathematics learning with curriculum content broader and more difficult, but students used textbooks less time, less frequently, and less for typical reasons and thought less highly of textbooks in mathematics learning with the foundation of mathematics knowledge more solid. Moreover, when it came to figures and geometry, students used textbooks more frequently to understand methods and imitate process of proof. Overall, the phenomenon of "students' use of mathematics textbooks declined with higher grade level" was prominent.

4.2.4. Teachers' demographic variables: Teachers' textbook use and interventions with students

The test results showed that there were statistically significant differences in the length of time and frequency of textbook use among students taught by teachers with different demographic variables, except the days of textbook use in a week, the frequency of textbook use before and after class, and the frequency of using examples and exercises in textbooks between students taught by teachers with different levels of education. Regarding the reasons of textbook use, there were no statistically significant differences among students taught by teachers with different demographic

TABLE 5 Significant differences among different students in mathematics textbook use.

		Length of time of textbook use (X^2)		Frequency of textbook use at different timing (F/t)			
		Days/week	Min/day	Before class	In class	After class	Before the exam
Region		178.494***	213.551***	65.383***	138.744***	49.633***	63.890***
Gender			11.790*				
Grade level		304.491***	295.354***	65.049***	76.174***	71.828***	109.397***
Students taught by different teacher groups	Gender	101.287***	101.023***	−6.579***	−7.487***	−7.850***	−8.996***
	Education level		11.790*		5.669***		2.732**
	Title	83.835***	84.073***	20.520***	38.101***	19.673***	27.334***
	Teaching experience	44.292***	44.916***	5.738**	12.303***	6.513**	16.571***
	Experience in textbook teaching	27.594***	26.827***	−2.284*	−4.189***	−2.961**	−5.063***

(Continued)

TABLE 5 (Continued)

		Frequency of using different components (F/t)								
		Introductions	Exploratory tasks	Kernels	Worked examples	Tips	Exercises	Summaries	Mathematics activities	Reading materials
Region		93.453***	71.440***	44.199***	48.498***	86.122***	47.318***	77.717***	88.473***	90.636***
Gender		2.461*								
Grade level		66.435***	99.824***	105.036***	101.806***	41.867***	124.357***	76.431***	68.733***	84.459***
Students taught by different teacher groups	Gender	−7.262***	−8.023***	−7.785***	−5.903***	−5.590***	−7.558***	−6.625***	−7.665***	−7.310***
	Education level	3.687***	2.727**	2.172*		2.856**		3.556***	2.924**	3.767***
	Title	25.964***	27.778***	22.129***	23.697***	21.712***	26.103***	32.293***	31.667***	39.953***
	Teaching experience	12.813***	16.137***	14.465***	11.926***	7.754***	13.528***	11.754***	15.970***	17.035***
	Experience in textbook teaching	−4.540***	−4.381***	−4.033***	−2.630**	−2.597**	−3.476**	−4.275***	−4.390***	−4.989***
		Reasons of textbook use (X^2)		Perceptions of textbook use (F/t)						
		Purpose	Motivation	Grades	Knowledge	Skills	Abilities	Methods	Experience	Emotions
Region		60.026***	83.796***	133.537***	34.292***	102.836***	98.243***	90.797***	88.878***	104.515***
Gender			30.467***							
Grade level		38.400***	44.766***	39.815***	16.753***	27.523***	37.529***	37.167***	37.880***	33.920***
Students taught by different teacher groups	Gender			−5.380***	−4.553***	−4.107***	−4.056***	−4.767***	−5.521***	−5.026***
	Education level	17.117**		5.538***	3.224**	4.169***	4.183***	3.976***	2.358*	3.922***
	Title		26.564**	34.482***	12.660***	36.632***	33.792***	35.138***	26.102***	29.821***
	Teaching experience			11.284***	5.911**	8.962***	10.056***	13.555***	7.969***	6.723**
	Experience in textbook teaching		14.587*	−4.161***	−2.055*	−2.711**	−3.667***	−4.080***	−2.136*	−2.582*

*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

TABLE 6 The classifications of other curriculum resources and their deciders.

	Students from the east	Students from the midlands	Students from the west
ICTs	Smartphone (parent/student), computer (parent/student), website (student), PowerPoint (teacher), Bilibili (student), Geometer's Sketchpad (student)	Smartphone (parent/student), website (student), PowerPoint (teacher)	Smartphone (parent/student), website (student), PowerPoint (teacher), Bilibili (student)
Supplementary educational books (SEB)	Workbook (publisher), SEB1 (school), SEB2 (student), SEB3 (student)	Workbook (publisher), SEB4 (school)	Workbook (publisher), SEB5 (school), SEB6 (school)
School-based learning materials	Class notes (teacher), guiding case (teacher), Workbook (local teaching and research section)	Class notes (teacher)	Class notes (teacher), guiding case (teacher)

variables, except the purposes of textbook use between students taught by teachers with different levels of education and the motivation of textbook use among students taught by teachers with different titles and with different experiences in teaching with textbooks. And for the perceptions of textbook use, there were statistically significant differences among students taught by teachers with different demographic variables.

Specifically, students taught by female teachers used textbooks significantly more time, more frequently, and more for typical reasons and thought significantly more highly of textbooks in mathematics learning than students taught by male teachers. Similarly, students taught by teachers with bachelor's degrees or lower used textbooks significantly more time, more frequently, and more for typical reasons and thought significantly more highly of textbooks in mathematics learning than students taught by teachers with master's degrees or higher. In contrast, students taught by teachers with middle titles or with 6–15 years of experience in teaching mathematics used textbooks significantly more time, more frequently, and more for typical reasons and thought significantly more highly of textbooks in mathematics learning than students taught by teachers with primary and senior titles or with less than 5 years and more than 15 years of experience in teaching mathematics. Meanwhile, students taught by teachers with more than 5 years of experience in teaching with textbooks used textbooks significantly more time, more frequently, and more for typical reasons and thought significantly more highly of textbooks in mathematics learning than students taught by teachers with less than 5 years of experience in teaching with textbooks. In the interviews, we found these results were related to teachers' ways of textbook use and ways of intervention with students. According to teachers' responses, teachers' ways of textbook use involved direct use, indirect use, and absence of use (Jukić Matić and Glasnović Gracin, 2016) and teachers' ways of intervention with students were intertwined in direct (specific/general)/indirect and obligatory/voluntary (Rezat, 2012), as shown in **Table 7**.

From the results of interviews, female teachers and teachers with bachelor's degrees or lower were more likely to use mathematics textbooks directly and mediate students directly and specifically. For example, T3 said: "The examples, exercises, and homework were directly from textbooks... and I always asked students to read the kernels in class." Thus, their students used textbooks more frequently and thought more highly of textbooks in mathematics learning. However, teachers with middle titles, teachers with 6–15 years of experience in teaching mathematics, and teachers with more than 5 years of experience in teaching with textbooks were more likely to use mathematics textbooks less directly and more indirectly and mediate students more generally and indirectly. For example, T4 said: "I would adjust the amount and sequence of exploratory tasks in textbooks when I think they are not suitable for students... and I would also advise students to use textbooks when need help for tasks." But their students still used textbooks more frequently and thought more highly of textbooks in mathematics learning.

5. Discussion

5.1. Mathematics textbooks: Traditional but important curriculum resources for students' mathematics learning

Due to rapid technological, cultural, and economic development in China over the past 20 years, ICTS, supplementary educational books and school-based learning materials have become increasingly available in Chinese classrooms and families. As traditional curriculum resources, mathematics textbooks are still important tools for Chinese secondary students to learn mathematics which is consistent with previous studies (Fan et al., 2004; Wang and Fan, 2021). Specifically, Chinese students relied heavily on mathematics textbooks. Quite a few students used mathematics textbooks both in school and at home and not only on weekdays but also on weekends. Meanwhile, they used mathematics textbooks at different timing, especially in class and before the examination. And they pointedly used a portion of components in textbooks, mainly kernels, examples, and exercises. This finding is in line that of with existing studies (Weinberg et al., 2012; Thomas, 2013). Besides, Chinese students used mathematics textbooks for various but typical reasons. They used mathematics textbooks particularly to preview, revise, do homework, learn and exercise in class, and look up definitions, theorems, and formulas, which was largely affected by the Confucian tradition of 'learning the new by repeating the old' and the two-basic teaching of basic knowledge and basic skills. And they showed self-regulation (Wang and Fan, 2021) and teacher-mediation behind their use. Furthermore, Chinese students had a positive view about textbook use in mathematics learning, especially in developing mathematical knowledge, skills, and abilities. But, they were less positive about the impact of textbook use on grades, mainly because they had more access to other curriculum resources and ignored that using textbooks could improve their grades by developing their basic knowledge and basic skills.

TABLE 7 Teachers' ways of textbook use and ways of interventions with students.

	Teachers' ways of textbook use			Teachers' ways of intervention with students
	Direct use	Indirect use	Absence of use	
T1	80%	15%	5%	Direct, specific, obligatory, and voluntary
T2	75%	15%	10%	Indirect and voluntary
T3	85%	15%		Direct, specific, obligatory
T4	70%	25%	5%	Direct, specific and general, obligatory
T5	65%	25%	10%	Indirect and voluntary
T6	65%	30%	5%	Direct, general, obligatory, and voluntary
T7	70%	25%	5%	Direct, general, obligatory, and voluntary
T8	75%	20%	5%	Direct, specific, obligatory, and voluntary

5.2. Differences in students' use of mathematics textbooks across demographic factors

The results revealed that there existed significant differences in students' use of mathematics textbooks in terms of school regions, grade levels, and teachers' demographic variables, except student genders. Regarding school regions, students from the east relied significantly less on textbooks, used textbooks significantly less for typical reasons, and had significantly less positive views about textbook use in mathematics learning than students from the west, who were less than students from the midlands. This finding is different from Fan et al.'s (2004) conclusions that students from Fuzhou in the east used textbooks and their components more frequently than students from Kunming in the west. We think this contradiction is partly related to access to and types of other curriculum resources, behind that is local economic and educational development level. In fact, China is divided into the east, midlands, and west according to economic development level of every province ranging from high to low and geographical location. In the early 20th century, students from two regions had less access to other curriculum resources, which was not the main factor affecting students' use of mathematics textbooks at that time. Whereas after 20 years of economic and educational rapid development, students from three regions have more access to other various curriculum resources, which has a huge impact on textbook use. Further studies should be conducted to explore how economic development affects students' textbook use.

With regard to student genders and grade levels, there were fewer significant differences between gender groups, which revealed that student genders might not be the main factor affecting students' textbook use. In terms of grade levels, the higher graders relied significantly less on textbooks, used textbooks significantly less for typical reasons, and had significantly less positive views about textbook use in mathematics learning than the lower graders. From the interviews, we found this result was related to breadth and difficulty of curriculum content, specific content areas, and students' mathematics knowledge base, which was both consistent with and complementary to the existing study (Fan et al., 2004). Fan et al. (2004) found the reasons why students have changed their use of mathematics textbooks compared to the previous semester were mainly breadth and difficulty of curriculum content, students' beliefs of mathematics, editors of textbooks, teachers, and parents. Curriculum content broader and more difficult with the grade level increasing is understandable because curriculum content in Chinese mathematics textbooks are organized in spiral sequence form. Meanwhile, there are more reasoning and proof in figures and geometry than numbers and algebra and statistics and probability (Ministry of Education of the People's Republic of China, 2011) and examples in textbooks provide students scientific ideas, authoritative process, and standard steps of the proof to imitate and understand reasoning, which were the reasons why students used textbooks more frequently and had more positive views about textbook use when they learned figures and geometry. In addition, students' developmental levels represented by students' mathematics knowledge are also the factors that influence students' textbook use. As grade levels changed, students with more solid foundation of mathematics knowledge and more mature beliefs could more effectively use textbooks to integrate new knowledge with prior experience, so that they used textbooks less frequently and had less positive views about textbook use. Under the influence of student factors and content factors, students' use of mathematics textbooks declined with higher grade level. Whether student factors play a more important role in students' textbook use than content or textbook factors needs to be further examined.

Regarding teachers' demographic variables, students taught by female teachers, by teachers with bachelor's degrees, by teachers with middle titles, by teachers with 6–15 years of experience in teaching mathematics, or by teachers with more than 5 years of experience in teaching with textbooks relied significantly more on textbooks and had significantly more positive views about textbook use in mathematics learning than students taught by male teachers, by teachers with master's degrees, by teachers with primary and senior titles, by teachers with 0–5 and more than 15 years of experience in teaching mathematics, or by teachers with 0–5 year of experience in teaching with textbooks, respectively. But there were fewer differences in reasons of textbook use among students taught by teachers with different demographic variables. This finding was related to teachers' ways of textbook use and ways of intervention with students, behind that were teachers' knowledge and beliefs about mathematics textbooks (Rezat, 2013; Jukić Matić and Glasnović Gracin, 2016). According to the interviews, female teachers and teachers with bachelor's degrees or lower, who were more likely to regard textbooks as the decisive tools to complete the teaching task, would tend to use mathematics textbooks faithfully and explicitly ask students to use specific section. Although teachers with more than 5 years of experience in teaching with textbooks have more indirect use of textbooks and general mediation on students, experienced textbook users who had wealthy knowledge and mature beliefs about textbooks would urge students to read the main body in textbooks more (Fan et al., 2004). Therefore, these explained why their students used textbooks more frequently and thought more highly of textbooks in mathematics learning. But compared to novice and experienced teachers, knowledge and beliefs about mathematics textbooks of teachers with middle titles and teachers with 61–5 years of experience in teaching mathematics are in the stage of development and formation. They were more likely to use mathematics textbooks less directly and more indirectly and mediate students more generally and indirectly. Why their students used textbooks more frequently and thought more highly of textbooks in mathematics learning could not be well explained in this study. This might be due to the small number of interviewed teachers with middle titles and with 6–15 years of experience in teaching mathematics and no classroom observations. Further research should be conducted to explore what cause such a difference.

5.3. Implications

Although this study focused on Chinese context, some results and discussions may have general implications. From the perspective of the subject and artifact in activity system, mathematics textbooks should meet students' individual development and diverse needs. On the one hand, students have more access to more various curriculum resources. On the other hand, different groups of students use mathematics textbooks in different ways. These mean that mathematics textbooks will be easily at a disadvantage in students' resource system. Therefore, it is a key issue to monitor the quality of curriculum resources (Wang and Fan, 2021), further to make mathematics textbooks play a more important role in the resource system from students' different learning needs.

From the perspective of the community in activity system, students should be taught to use mathematics textbooks autonomously and creatively to improve learning quality through teachers' mediation. Teachers also proved to be important in the students' use of mathematics textbooks. Teachers' knowledge and beliefs about textbooks strongly shape their practice and decide when, where, and which sections of the textbook were to be used by students. Thus, it is necessary for teachers to help students realize the importance and significance of mathematics textbooks and textbook use. It can promote students to know long-term and short-term goals for textbook use and develop the awareness and ability to regulate their textbook use.

5.4. Limitations and future directions

Finally, we should point out that the data collected were mainly based on self-report and no classroom observations were employed due to the unexpected impact of the COVID-19 pandemic, which are the limitations of this study. According to this study, Chinese students use a variety of curriculum resources to learn mathematics and teachers' knowledge and beliefs about textbooks behind teachers' demographic variables are the potential factors that affect students' textbook use. Hence, future research should explore how students in different economic, social, and cultural contexts incorporate various curriculum resources in mathematics learning. In fact, many researchers expressed similar concerns about this issue. For example, Glasnović Gracin and Jukić Matić (2021) have explored teachers' and students' use of textbooks and other resources during the process of educational reform in Croatia and Fan et al. (2022) have studied Shanghai students' access, use, and perceptions of ICTs in learning mathematics. Moreover, examining the relationship between teachers' knowledge and beliefs about textbooks and students' textbook use could be a direction for further studies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Northeast Normal University Academic Ethics Committee. The participants provided their written informed consent to participate in this study.

Author contributions

TJ contributed to the literature review and theoretical background, collected and analyzed the data, and wrote the original draft of the manuscript. SL contributed to the construction of concepts and index framework, questionnaire preparation, introduction, and methodology, and revised the manuscript. Both authors contributed to the article and approved the submitted version.

Acknowledgments

We thank the students, teachers, and experts for participating in and assisting with this study, especially thank Prof. Li Jiang, Northeast Normal University, for her assistance at the revision stage of this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Amabile, T. M., Hill, K. G., Hennessey, B. A., and Tighe, E. M. (1994). The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *J. Pers. Soc. Psychol.* 66, 950–967. doi: 10.1037/0022-3514.66.5.950

Berger, M. (2019). Different reading styles for mathematics text. *Educ. Stud. Math.* 100, 139–159. doi: 10.1007/s10649-018-9871-y

Ewing, B. (2006). "Go to the page and work it from there": Young people's experiences of learning mathematics from a text. *Aus. Sen. Math. J.* 20, 8–14.

Fan, L., Chen, J., Zhu, Y., Qiu, X., and Hu, J. (2004). "Textbook use within and beyond mathematics classrooms: A study of 12 secondary schools in Kunming and Fuzhou of China," in *How Chinese learn mathematics: Perspectives from insiders*, eds L. Fan, N.-Y. Wong, J. Cai, and S. Li (Singapore: World Scientific), 228–261. doi: 10.1142/9789812562241_0009

Fan, L., Luo, J., Xie, S., Zhu, F., and Li, S. (2022). Chinese students' access, use and perceptions of ICTs in learning mathematics: Findings from an investigation of Shanghai secondary schools. *ZDM Math. Educ.* 54, 611–624. doi: 10.1007/s11858-022-01363-5

Fan, L., Xiong, B., Zhao, D., and Niu, W. (2018). How is cultural influence manifested in the formation of mathematics textbooks? A comparative case study of resource book series between Shanghai and England. *ZDM Math. Educ.* 50, 787–799. doi: 10.1007/s11858-018-0976-7

Fan, L., Zhu, Y., and Miao, Z. (2013). Textbook research in mathematics education: Development status and directions. *ZDM Math. Educ.* 45, 633–646. doi: 10.1007/s11858-013-0539-x

Glasnović Gracin, D., and Jukić Matić, L. (2021). Use of textbooks and other resources in curriculum reform. A longitudinal case study. *ZDM Math. Educ.* 53, 1373–1385. doi: 10.1007/s11858-021-01271-0

Grave, I., and Pepin, B. (2015). Teachers' use of resources in and for mathematics teaching. *Nordic Stud. Math. Educ.* 20, 199–222.

Griesel, H., and Postel, H. (1983). Zur Theorie des Lehrbuchs-Aspekte der Lehrbuchkonzeption. *ZDM Math. Educ.* 83, 287–293.

Jonassen, D. H., and Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educ. Tech. Res.* 47, 61–79. doi: 10.1007/BF02299477

Jukić Matić, L., and Glasnović Gracin, D. (2016). The use of the textbook as an artefact in the classroom. *J. Math. Didakt.* 37, 349–374. doi: 10.1007/s13138-016-0091-7

Kersey, S. (2019). Student perceptions on teaching and learning using open educational resources in college Calculus. *J. Comput. Math. Sci. Teach.* 38, 249–265.

Lin, F. L., Wang, T. Y., and Yang, K. L. (2018). Description and evaluation of a large-scale project to facilitate student engagement in learning mathematics. *Stud. Educ. Eval.* 58, 178–186. doi: 10.1016/j.stueduc.2018.03.001

Macintyre, T., and Hamilton, S. (2010). Mathematics learners and mathematics textbooks: A question of identity? Whose curriculum? Whose mathematics? *Curriculum J.* 21, 3–23. doi: 10.1080/09585170903558224

Ministry of Education of the People's Republic of China (2011). *Mathematics curriculum standards for compulsory education*. Beijing: Beijing Normal University Press.

Ministry of Education of the People's Republic of China (2022b). *National statistical bulletin on the development of education*. Available online at: http://www.gov.cn/xinwen/2022-09/15/content_5710039.htm (accessed September 15, 2022)

Ministry of Education of the People's Republic of China (2022a). *Mathematics curriculum standards for compulsory education*. Beijing: Beijing Normal University Press.

Moyer, J. C., Robison, V., and Cai, J. (2018). Attitudes of high-school students taught using traditional and reform mathematics curricula in middle school: A retrospective analysis. *Educ. Stud. Math.* 98, 115–134. doi: 10.1007/s10649-018-9809-4

Nicol, C. C., and Crespo, S. M. (2006). Learning to teach with mathematics textbooks: How preservice teachers interpret and use curriculum materials. *Educ. Stud. Math.* 62, 331–355. doi: 10.1007/s10649-006-5423-y

Olsher, S., and Cooper, J. (2021). Aspects of the teacher-textbook relationship: What can we learn about teachers when they tag didactic metadata? *ZDM Math. Educ.* 53, 1347–1358. doi: 10.1007/s11858-021-01251-4

Österholm, M. (2006). Characterizing reading comprehension of mathematical texts. *Educ. Stud. Math.* 63, 325–346. doi: 10.1007/s10649-005-9016-y

Pepin, B., and Haggarty, L. (2001). Mathematics textbooks and their use in English, French and German classrooms. *ZDM Math. Educ.* 33, 158–175. doi: 10.1007/BF02656616

Polly, D. (2017). Elementary school teachers' uses of mathematics curricular resources. *J. Curriculum Stud.* 49, 132–148. doi: 10.1080/00220272.2016.1154608

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Rev. Educ. Res.* 75, 211–246. doi: 10.3102/00346543075002211

Rezat, S. (2006). "A model of textbook use," in *Proceedings of the 30th conference of the international group for the psychology of mathematics education*, (Prague: Psychology of Mathematics Education), 409–416.

Rezat, S. (2009). "The utilization of mathematics textbooks as instruments for learning," in *Proceedings of the 6th congress of the European society for research in mathematics education*, eds V. Durand-Guerrier, S. Soury-Lavergne, and F. Arzarello (Westerburg: INRP), 1260–1269.

Rezat, S. (2011). Wozu verwenden Schüler ihre Mathematikschulbücher? Ein Vergleich von erwarteter und tatsächlicher Nutzung. *J. Math. Didakt.* 32, 153–177. doi: 10.1007/s13138-011-0028-0

Rezat, S. (2012). "Interactions of teachers' and students' use of mathematics textbooks," in *From text to 'lived' resources*, eds G. Gueudet, B. Pepin, and L. Trouche (New York: Springer), 231–245.

Rezat, S. (2013). The textbook-in-use: Students' utilization schemes of mathematics textbooks related to self-regulated practicing. *ZDM Math. Educ.* 45, 659–670. doi: 10.1007/s11858-013-0529-z

Rezat, S. (2021). How automated feedback from a digital mathematics textbook affects primary students' conceptual development: Two case studies. *ZDM Math. Educ.* 53, 1433–1445. doi: 10.1007/s11858-021-01263-0

Rezat, S., and Sträßer, R. (2013). "Methodologies in Nordic research on mathematics textbooks," in *Nordic research in didactics of mathematics: Past, present, and future*, eds B. Grevholm, P. S. Hundeland, K. Juter, K. Kislenko, and P.-E. Persson (Oslo: Cappelen Damm Akademisk), 469–482.

Rezat, S., Fan, L., Hattermann, M., Schumacher, J., and Wuschke, H. (2019). *Proceedings of the 3rd international conference on mathematics textbook research and development*. Paderborn: Universitätsbibliothek Paderborn.

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020

Schubring, G., and Fan, L. (2018). Recent advances in mathematics textbook research and development: An overview. *ZDM Math. Educ.* 50, 765–771. doi: 10.1007/s11858-018-0979-4

Schubring, G., Fan, L., and Giraldo, V. (2018). *Proceedings of the 2nd International conference on mathematics textbook research and development*. Rio de Janeiro: Federal University of Rio de Janeiro.

Sievert, H., Van den Ham, A.-K., and Heinze, A. (2021). The role of textbook quality in first graders' ability to solve quantitative comparisons: A multilevel analysis. *ZDM Math. Educ.* 53, 1417–1431. doi: 10.1007/s11858-021-01266-x

Thomas, A. (2013). *A study of algebra 1 students' use of digital and print textbooks. [Doctoral dissertation]*. Missouri: University of Missouri-Columbia.

Van den Ham, A.-K., and Heinze, A. (2018). Does the textbook matter? Longitudinal effects of textbook choice on primary school students' achievement in mathematics. *Stud. Educ. Eval.* 59, 133–140. doi: 10.1016/j.stueduc.2018.07.005

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Wang, Y., and Fan, L. (2021). Investigating students' perceptions concerning textbook use in mathematics: A comparative study of secondary schools between Shanghai and England. *J. Curriculum Stud.* 53, 675–691. doi: 10.1080/00220272.2021.1941265

Weinberg, A., and Wiesner, E. (2011). Understanding mathematics textbooks through reader-oriented theory. *Educ. Stud. Math.* 76, 49–63. doi: 10.1007/s10649-010-9264-3

Weinberg, A., Wiesner, E., Benesh, B., and Boester, T. (2012). Undergraduate students' self-reported use of mathematics textbooks. *PRIMUS* 22, 152–175. doi: 10.1080/10511970.2010.509336

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 06 March 2023
DOI 10.3389/fpsyg.2023.1050259

OPEN ACCESS

EDITED BY
Ahmed Mohamed Fahmy Yousef,
Fayoum University,
Egypt

REVIEWED BY
Shan Li,
Lehigh University,
United States
Xin Tang,
Shanghai Jiao Tong University,
China

*CORRESPONDENCE
Xiang Hu
✉ xianghu@ruc.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 21 September 2022
ACCEPTED 13 February 2023
PUBLISHED 06 March 2023

CITATION
Guo M, Cao Y and Hu X (2023) Chinese students' filial piety beliefs and procrastination in mathematics learning: The mediating role of academic emotions.
Front. Psychol. 14:1050259.
doi: 10.3389/fpsyg.2023.1050259

Chinese students' filial piety beliefs and procrastination in mathematics learning: The mediating role of academic emotions

Meng Guo[1], Yiming Cao[1] and Xiang Hu[2]*

[1]School of Mathematical Sciences, Beijing Normal University, Beijing, China, [2]School of Education, Renmin University of China, Beijing, China

This study examined the relations between Chinese students' filial piety beliefs and mathematics procrastination and the mediating role of academic emotions in the relations. Analysis of data on 1,476 primary school students in China with structural equation modeling revealed that students' reciprocal and authoritarian filial piety beliefs were positively related to academic enjoyment and anxiety, respectively. Students' procrastination in mathematics learning was positively related to anxiety and authoritarian filial piety beliefs and had negative associations with enjoyment and reciprocal filial piety beliefs. The bootstrap analysis results confirmed the mediating role of anxiety in the relation between authoritarian filial piety beliefs and procrastination. Reciprocal filial piety beliefs had negative indirect relationship with procrastination *via* enjoyment. The results were explained from a socio-cultural perspective. The theoretical contributions and practical implications are discussed.

KEYWORDS

filial piety beliefs, enjoyment, anxiety, procrastination, mathematics learning

Introduction

Academic procrastination is a problematic behavior in student learning associated with a series of maladaptive learning outcomes such as poor academic performance (Moon and Illingworth, 2005; Kim and Seo, 2015) and less use of cognitive learning strategies (Wolters, 2003). Although academic procrastination has been extensively studied, extant research has overwhelmingly focused on Western populations (Klassen et al., 2008; Burnam et al., 2014). Preliminary studies found cross-cultural differences of students' absolute level of academic procrastination and its consequences (e.g., Klassen et al., 2010), calling for more research of academic procrastination among non-Western samples such as East Asians (Jin et al., 2019; Yang et al., 2019).

Given the maladaptive role of procrastination in student learning, much attention has been paid to examine what contributes to students' academic procrastination. Parent–child interaction has been found to be a critical factor affecting students' procrastination behaviors (Ma et al., 2011; Zakeri et al., 2013; Chen et al., 2022). Filial piety beliefs (guiding children to repay, honor, and care for their parents, and unconditionally obey parents' requirements; Ho, 1994) reflect Chinese students' perceptions and beliefs of parent–child relationship and interaction (Yeh and Bedford, 2004; Chen, 2014). In Chinese societies, filial piety means that children should work hard to pursue for academic success to repay their parents and honor families (Tao and Hong, 2014), which may decrease students' procrastination in learning process. However, few empirical studies have been conducted to investigate the relation between students' filial piety beliefs and procrastination behaviors. This is important to understand cultural and family antecedents of students' academic procrastination in Chinese societies.

Filial piety beliefs have been found to influence students' academic outcomes, such as view of learning (Chen and Wong, 2014) and achievement (Chen, 2014). Academic emotions may also be shaped by students' filial piety beliefs. According to control-value theory (Pekrun, 2019), students' academic emotions are influenced by appraised and broader environmental antecedents. Filial piety beliefs reflect parent–child interaction from children's perspectives, which may serve as a family antecedent of academic emotions. Children who tend to obey to parents' authority (e.g., authoritarian filial piety) may have higher level of academic anxiety. As mathematics is a critical school subject for students' future academic and career success in China, parents place high emphasis on their children's mathematics learning and have high expectation on children's mathematics performance (Cao et al., 2007). Thus, filial Chinese students may tend to internalize parents' beliefs and endorse the value and importance of mathematics, and try to fulfill parents' expectation, which may affect their emotions in mathematics learning. In other words, students' filial piety beliefs may shape their appraisal for and emotions in mathematics.

Given the potential relationship between filial piety and emotions (Yeh, 2006) and between emotions and procrastination (Rahimi and Vallerand, 2021), academic emotions may serve as a mediator in the relations between filial piety beliefs and procrastination. In the Chinese context, filial students tend to value schools and engage more positive emotions in learning, which may create a protective mechanism to avoid procrastination. On the other hand, children who focus on family honor and obedience to parents are likely to experience more anxiety (Yeh, 2006), and thus engage in academic procrastination (Pychyl et al., 2000). In other words, students' filial piety beliefs might influence their academic emotions and further shape their procrastination behaviors. However, there is a lack of empirical evidence testing this relation.

To fill in the aforementioned research gaps, this study aims to empirically examine the relations between Chinese students' filial piety beliefs and their procrastination in mathematics learning and the mediational role of mathematics emotions in the relations.

Literature review

Filial piety beliefs

Filial piety is an important concept in Confucian culture, describing how children should interact with their parents (Ho and Lee, 1974; Yeh and Bedford, 2003). Filial piety requires children to love their parents by supporting and caring for parents, maintaining family honor, obeying parents' decisions, and achieving parents' wishes (Ho, 1994). Some research showed that filial piety was associated with maladaptive outcomes, including cognitive conservatism (Ho, 1996), neuroticism (Zhang and Bond, 1998), and less creativity (Ho, 1994). In contrast, other studies identified a positive role of filial piety, as it facilitated harmonious intergenerational relationships (Sung, 1995) and family cohesion (Cheung et al., 1994).

Given the contradictory findings on the role of filial piety, Yeh (2003) proposed the dual filial piety model (DFPM), which further divides filial piety into two dimensions: reciprocal and authoritarian filial piety. DFPM posits that reciprocal filial piety is associated with adaptive outcomes, while authoritarian filial piety is associated with maladaptive outcomes (Yeh, 2003). Reciprocal filial piety comprises "emotionally and spiritually attending to one's parents out of gratitude for their efforts in having raising one, and physical and financial care for one's parents as they age and when they die for the same reason," while authoritarian filial piety encompasses "suppressing one's own wishes and complying with one's parents' wishes because of their seniority in physical, financial or social terms, as well as continuing the family lineage and maintaining one's parents' reputation because of the force of role requirements" (Yeh and Bedford, 2003, p. 216). Existing literature has largely support DFPM in that reciprocal filial piety is associated with desirable outcomes such as openness (Yeh and Bedford, 2003) and better interpersonal relationships (Yeh and Bedford, 2004), whereas authoritarian filial piety is related to maladaptive outcomes, like performance-avoidance goals (Chen, 2016), neuroticism (Yeh and Bedford, 2003), and anxiety (Yeh, 2006). Therefore, this study followed DFPM (Yeh, 2003) to conceptualize filial piety as comprising both reciprocal and authoritarian filial piety.

Filial piety and academic procrastination

Procrastination refers to the unnecessary delay in taking action, despite unavoidably undesirable results (Steel, 2007). Academic procrastination is seen as an undesirable learning tendency, as it was always related to such maladaptive learning outcomes as high levels of anxiety and stress (Chu and Choi, 2005; Spada et al., 2006), less use of cognitive learning strategies (Wolters, 2003), and poor academic performance (Moon and Illingworth, 2005; Kim and Seo, 2015). Given the maladaptive role of procrastination in student learning, it is essential to identify potential antecedents of procrastination (Steel and Klingsieck, 2016; Rahimi and Vallerand, 2021). Previous studies have found that procrastination was influenced by individual personality traits (e.g., conscientiousness; Steel and Klingsieck, 2016) and motivation (e.g., goal orientations; Howell and Watson, 2007), teacher-related factors (e.g., teachers' clear expectations; Schraw et al., 2007), and so forth.

Researchers have found that East Asian culture shapes students' learning process (Leung, 2006; Hu et al., 2018; Guo et al., 2022). However, informative as the extant literature is, few studies have empirically examined whether and how cultural background might influence academic procrastination. Further, it is noteworthy that research on academic procrastination has mostly surveyed college students (Rabin et al., 2011; Kim and Seo, 2015), while younger students' academic procrastination has been under-investigated (Xue et al., 2021). To fill in the research gaps, this study focused on one potential cultural antecedent of academic procrastination—students' filial piety beliefs—and examined the relation between filial piety beliefs and procrastination in mathematics learning among primary school students in China.

Exiting evidence on parental influence on children's academic procrastination provides support for our hypothesis on the relation between filial piety and academic procrastination. Previous studies have found that authoritarian parenting style positively predicted children's academic procrastination (Zakeri et al., 2013; Soysa and Weiss, 2014; Chen et al., 2022). For instance, by investigating 743

Chinese college students, Chen et al. (2022) identified that students who perceived authoritarian parenting style were more inclined to focus on mistakes and thus procrastinated learning tasks. Given authoritarian filial piety beliefs reflect students' perception of authoritarian parent–child interaction (Chen, 2014), it is also likely to increase children's academic procrastination. Specifically, students with authoritarian filial piety tend to restrain their own wishes and obey their parents' academic requirements (Yeh, 2006), which may hurt their individual autonomy (Sun et al., 2019) in the learning process, and further lead to more academic procrastination. Therefore, students' authoritarian filial piety is hypothesized to increase their procrastination in mathematics learning.

Regarding reciprocal filial piety, there is also a lack of research on its association with academic procrastination in mathematics learning. However, existing literature have confirmed its beneficial role in student personal development and academic learning (Yeh, 2006). A study of 750 junior high school students revealed that students' reciprocal filial belief positively predicted their autonomy need satisfaction and academic performance in mathematics and reading (Zhou et al., 2020). Indeed, studying hard to achieve academic success is regarded as an important way to repay parents in China (Leung, 2006; Tao and Hong, 2014). Students who have the desire to repay their parents (i.e., reciprocal filial piety) tend to be hardworking at school and less academically procrastinated. Therefore, we hypothesize that reciprocal filial piety will negatively predict academic procrastination.

The mediating role of academic emotions

The importance of academic emotions students experienced in school settings has long been recognized by educators. Academic emotions refer to "emotions tied directly to achievement activities or achievement outcomes" (Pekrun, 2006, p. 317). Pekrun's (2006) control-value theory (CVT) posits that students' academic emotions are influenced by appraised and broader environmental antecedents, and influence learning outcomes such as achievement, motivation to learn, and self-regulation. Filial piety, as an essential cultural aspect in China, might influence students' academic emotions, which in turn affects procrastination in mathematics learning. That is, academic emotions may play a mediating role on the relation between filial piety and academic procrastination. However, there are, to the best of our knowledge, few studies have empirically tested the mediating effect of academic emotions on the relation, which is a focus of our study. Notably, we specifically focus on two academic emotions (i.e., enjoyment and anxiety) as they have been widely studied and demonstrated importance in Chinese students' learning process (Jiang and Dewaele, 2019; Guo et al., 2020).

The relation between academic emotions and academic procrastination has been well-documented in literature (Spada et al., 2006; Rahimi and Vallerand, 2021). Positive academic emotions (e.g., enjoyment) can serve as a protective factor for academic procrastination (Rahimi and Vallerand, 2021), while negative emotions were found to trigger more academic procrastination (Pychyl et al., 2000; Rahimi and Vallerand, 2021). Rahimi (2019) conducted a three-phase longitudinal study and found a reciprocal positive relationship of enjoyment and anxiety with academic procrastination. Therefore, we hypothesize that mathematics enjoyment and anxiety will have a positive and negative association with mathematics procrastination, respectively.

The relation of filial piety with academic enjoyment and anxiety has been under-investigated. Preliminary evidence, however, support the role of filial piety in students' emotional process (Yeh, 2006). By investigating senior and junior high school students in Chinese Taipei, Yeh (2006) found that students with reciprocal filial piety perceived less level of depression and anxiety, while students' authoritarian filial piety triggered their experience of anxiety. Thus, we hypothesize that reciprocal and authoritarian filial piety beliefs negatively and positively predict mathematics anxiety, respectively. Further, extant evidence on the relation between filial piety and academic motivation might shed light on the relation between filial piety and academic enjoyment. For instance, Chen (2016) found that reciprocal filial piety positively predicted Chinese students' mastery goals. Students with mastery goals always experienced more positive emotions (e.g., enjoyment; Lee et al., 2003; Pekrun et al., 2009). Thus, this study hypothesizes that students' reciprocal filial piety will positively predict enjoyment. Authoritarian filial piety beliefs require children to unconditionally follow parents' arrangements and obey parental authority, which may hinder students' individual autonomy in the learning process (Sun et al., 2019). According to self-determination theory, autonomy serves as an important factor facilitating the development of intrinsic motivation (Niemiec and Ryan, 2009). Researchers have identified the positive relationship between intrinsic motivation and enjoyment (Gråstén et al., 2012). Therefore, authoritarian filial piety is hypothesized to negatively predict mathematics enjoyment.

To sum up, extant studies have identified associations between filial piety and academic emotions (Yeh, 2006) and between academic emotions and procrastination (Rahimi and Vallerand, 2021), indicating a potential mediating role of academic emotions in the relation between filial piety and procrastination.

The present study

Drawing on the control-value theory of academic emotions, we advanced an indigenous psychological perspective to examine the relations between Chinese students' filial piety—an essential cultural belief for Chinese people—with their academic emotions (i.e., math enjoyment and anxiety) and procrastination in mathematics learning. Based on previous studies on filial piety, emotions, and academic procrastination (Yeh, 2006; Chen, 2016; Rahimi and Vallerand, 2021), this study hypothesizes that students' filial piety has a significant relationship with academic emotions and further links to procrastination in mathematics learning. The specific hypothesis in the study is presented below.

> *Hypothesis 1*: Authoritarian filial piety beliefs will be related to a higher level of mathematics anxiety and less enjoyment.
>
> *Hypothesis 2*: Reciprocal filial piety beliefs are expected to be positively and negatively associated with mathematics enjoyment and anxiety, respectively.
>
> *Hypothesis 3*: Mathematics anxiety and enjoyment will have a positive and negative association with academic procrastination, respectively.

> *Hypothesis 4*: Reciprocal and authoritarian filial piety beliefs will be negatively and positively related to academic procrastination, respectively.

Methods

Participants and procedure

This study involved 1,476 Fourth to Sixth Grade primary school students (Mage = 10.82 years, SD = 0.96) in a city of Guizhou Province, China. Of these students, 48.0% were boys, 51.5% were girls, and 0.5% were missing gender information. After consulting with local experienced teachers, we invite each school with high, medium, and low teaching quality, respectively, in the city to participated into the study, resulting in totally 3 participating schools. With the help of teachers in the schools, the first author administered the questionnaire to students in class. Before data collection, ethics approval was acquired. Students who agreed to take part in this study filled in a consent form and then completed a self-reported questionnaire on filial piety and academic emotions.

Measures

Filial piety

The filial piety scale (Yeh, 2003) was used to measure both reciprocal filial piety (4 items; α =0.813; e.g., "Children should support their parents to make them live better") and authoritarian filial piety (4 items; α =0.769; e.g., "Children should do what their parents ask"), on a 5-point Likert scale.

Academic emotions

Measures of mathematics enjoyment and anxiety were adapted from the enjoyment and anxiety subscale in the Achievement Emotions Questionnaire (Pekrun et al., 2011) and further revised to focus on mathematics learning. Four and three items were employed to measure mathematics enjoyment (α = 0.846; 4 items; e.g., "Mathematics class makes me feel happy") and anxiety (α = 0.795; 3 items; e.g., "I am worried that mathematics class will be very difficult"), respectively.

Procrastination

Four items were used to measure student procrastination in mathematics learning (α = 0.854; 4 items, e.g., "I often put off finishing my math homework"), which were adapted from the procrastination scale in the study of Wolters (2003).

Background variables

Student gender (boy = 0; girl = 1), grade (Grade 4 = 1; Grade 5 = 2; Grade 6 = 3) and parent highest educational level (primary school = 1; secondary School = 2; high school = 3; university = 4; master's degree or above = 5) were considered as control variables in the analysis.

Data analysis

Before formal analyses, we imputed missing values of included variables using expectation maximization (EM) algorithm in SPSS software (Version 25.0). Structural equation modeling (SEM) was used to model the relation between students' filial piety beliefs, academic emotions, and procrastination. The comparative fit index (CFI) > 0.90 (Bentler, 1990), the Tucker-Lewis index (TLI) > 0.90, and the root mean squared error of approximation (RMSEA) < 0.08 (Browne and Cudeck, 1992) indicated acceptable model fit. Bootstrap methods with 5,000 resamples were conducted to test the mediation effects of academic emotions in the relation between filial piety beliefs and procrastination (Hayes, 2013) using Mplus software (Version 8.0).

Results

Preliminary analysis

Results of descriptive statistics and correlations among variables are shown in Table 1. Authoritarian filial piety was

TABLE 1 Descriptive statistics, correlations, and internal consistency of the main variables in this study.

	RFP	AFP	En	An	Pr	PE	Gr	SG
RFP	1							
AFP	0.224**	1						
En	0.358**	0.109**	1					
An	0.009	0.064*	−0.126**	1				
Pr	−0.227**	0.067**	−0.260**	0.236**	1			
PE	0.040	−0.162**	0.094**	−0.094**	−0.080**	1		
Gr	−0.032	−0.224**	−0.091**	0.140**	0.009	−0.066*	1	
SG	0.034	−0.021	−0.040	0.103**	−0.108**	−0.075**	0.038	1
M	4.595	3.547	3.967	3.345	1.646	2.540	1.957	0.517
SD	0.525	0.876	0.780	1.079	0.769	0.956	0.802	0.500
α	0.813	0.769	0.846	0.795	0.854	/	/	/

RFP, Reciprocal filial piety; AFP, Authoritarian filial piety; Enjoyment, En; Anxiety, An; Procrastination, Pr; PE, parental highest educational level; Gr, Grade; SG, Student gender. *p < 0.05, **p < 0.01.

positively correlated with reciprocal filial piety ($r = 0.224$, $p < 0.01$). Mathematics enjoyment had positive correlations with reciprocal ($r = 0.358$, $p < 0.01$) and authoritarian filial piety ($r = 0.109$, $p < 0.01$). Mathematics anxiety was positively related to authoritarian filial piety ($r = 0.064$, $p < 0.05$). Procrastination had positive correlations with authoritarian filial piety ($r = 0.067$, $p < 0.01$) and anxiety ($r = 0.236$, $p < 0.01$) and had negative correlations with reciprocal filial piety ($r = -0.227$, $p < 0.01$) and enjoyment ($r = -0.260$, $p < 0.01$).

Structural equation modeling

The SEM results showed that the hypothesized model fit the data well [$X^2(df) = 529.101$ (184), $X^2/df = 2.876$, $RMSEA = 0.036$ with 90% CI [0.032, 0.039], $CFI = 0.969$, $TLI = 0.962$]. The factor loading of each item was significant, ranging from 0.493 to 0.818. The variance explained by the model was 0.165 for enjoyment, 0.056 for anxiety, and 0.211 for academic procrastination, respectively.

As shown in Figure 1, reciprocal filial piety was positively associated with enjoyment ($\beta = 0$ .370, $p < 0.001$). Authoritarian filial piety was positively related to anxiety ($\beta = 0$ .126, $p < 0.001$). Academic procrastination was negatively predicted by both enjoyment ($\beta = -0.181$, $p < 0.001$) and reciprocal filial piety ($\beta = -0.229$, $p < 0.001$). Authoritarian filial piety ($\beta = 0$ .140, $p < 0.001$) and anxiety ($\beta = 0$ .254, $p < 0.001$) had a positive association with academic procrastination.

Mediation tests

We checked the potential mediating roles of academic emotions in the relation between filial piety and procrastination *via* bootstrap methods (see Table 2). The results indicated that reciprocal filial piety had a direct negative relationship with student academic procrastination ($\beta = -0.229$, $p < 0.001$; 95%CI [−0.300, −0.159]) and indirect effects *via* enjoyment ($\beta = -0.067$, $p < 0.001$; 95%CI [−0.098, −0.040]). The authoritarian filial piety was positively related to procrastination *via* anxiety ($\beta = 0$ .032, $p = 0.002$; 95% CI [0.014, 0.054]). The direct positive effect of authoritarian filial piety on procrastination was also significant ($\beta = 0$ .140, $p < 0.001$; 95%CI [0.076, 0.203]).

Discussion

The aim of this study was to examine the relations between students' filial piety and academic procrastination in mathematics learning and the mediating role of mathematics enjoyment and anxiety in the relations. Results were all consistent with our hypotheses. We found that reciprocal filial piety had a direct positive relationship with procrastination. A plausible explanation is that students with reciprocal filial piety are more inclined to focus on mastering knowledge and improving competence (Chen, 2016), which can further minimize procrastination behaviors (Wolters, 2004). This is consistent with King and McInerney's (2019) findings that students who had the desire to repay and support their parents tended to have less disengagement. Another possibility is that students who are grateful to their parents have a higher level of self-esteem (Yan and Chen, 2018), which can prevent procrastination behaviors (Hajloo, 2014).

As hypothesized, authoritarian filial piety was positively related to procrastination in mathematics learning. Researchers have found that students who endorse authoritarian filial piety tend to hold an entity view of learning (i.e., intelligence is fixed and unchangeable; Chen and Wong, 2014), which may in turn lead to more procrastination (Rickert et al., 2014). Another plausible explanation is that students who focus on gaining family honor and meeting parents' requirements may

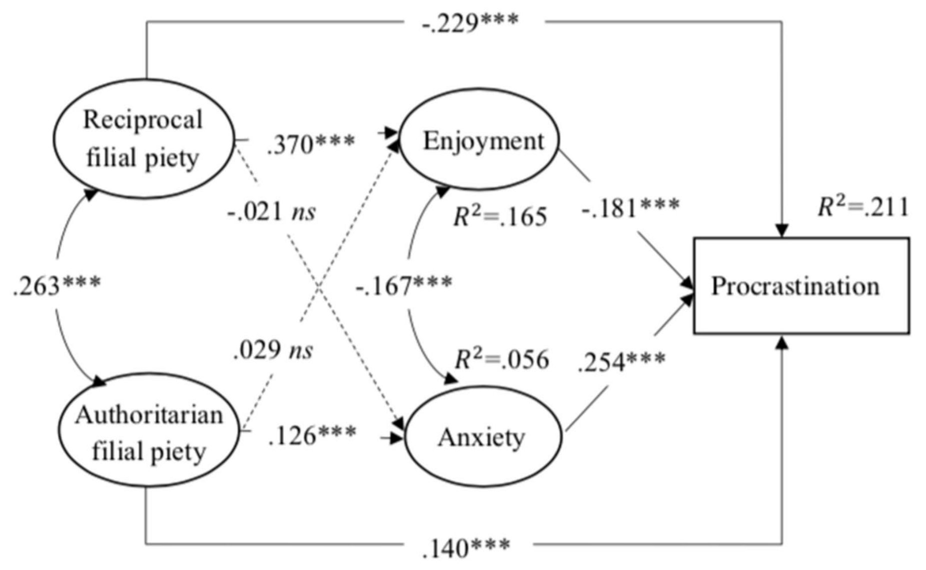


FIGURE 1
The relationship between Chinese students' filial piety beliefs, academic emotions and academic procrastination. Non-significant paths and covariables were not included in the figure for clarity. ***p<0.001.

TABLE 2 Standardized total effects, direct effect, and indirect effect on procrastination.

	Point estimate			Bootstrapping BC 95%CI	
	Standardized β	S.E.	*p*	Lower	Upper
Effects from RFP to Pr					
Direct effects	−0.229***	0.036	0.000	−0.300	−0.159
Indirect effects					
RFP-En-Pr	−0.067***	0.014	0.000	−0.098	−0.040
RFP-An-Pr	−0.005	0.008	0.494	−0.020	0.010
Total effects	−0.301***	0.034	0.000	−0.369	−0.234
Effects from AFP to Pr					
Direct effects	0.140***	0.032	0.000	0.076	0.203
Indirect effects					
AFP-En-Pr	−0.005	0.007	0.453	−0.021	0.007
AFP-An-Pr	0.032**	0.010	0.002	0.014	0.054
Total effects	0.167***	0.033	0.000	0.101	0.230

RFP, Reciprocal filial piety; AFP, Authoritarian filial piety; Enjoyment, En; Anxiety, An; Procrastination, Pr. $^{**}p<0.01$, $^{***}p<0.001$.

worry about possible academic failure and adopt performance-avoidance goals (Chen, 2016), and thus have more academic procrastination (Wolters, 2004). Also, authoritarian filial piety may be detrimental to the development of students' individual autonomy (Sun et al., 2019), which may hurt their intrinsic motivation (Ryan and Deci, 2000) and thus lead to procrastination in the learning process (Senécal et al., 1995).

We found that reciprocal filial piety beliefs were positively related to students' enjoyment of mathematics learning, indicating that students who had the desire to repay and love parents experienced more academic enjoyment. Students with reciprocal filial piety tend to have higher levels of self-esteem (Yan and Chen, 2018), which may further increase their academic enjoyment (Cheng and Furnham, 2003). Previous studies have found that students' reciprocal filial piety can promote their endorsement of mastery goals (Chen, 2016), which further trigger their enjoyment of learning (Pekrun et al., 2009). Another possible explanation is that students who endorse reciprocal filial piety tend to have a close and intimate relationship with their parents. These students may perceive more relatedness with their parents. Based on self-determination theory, relatedness can foster students' intrinsic motivation (Niemiec and Ryan, 2009) and further increase students' enjoyment experience in the learning process (Gråstén et al., 2012).

Authoritarian filial piety beliefs were found to have a positive relationship with students' academic anxiety, which is consistent with previous studies (Yeh, 2006; Rahimi and Vallerand, 2021). It suggests that students who emphasize family reputation and obey parental authority tend to experience more anxiety. These students may worry about the possible negative effects of academic failure on their families (e.g., losing their parents' faces, and failing to meet parental expectations), which may lead to a higher level of anxiety in learning. Previous studies have confirmed that authoritarian filial piety increased students' adoptions of performance-avoidance goals (Chen, 2016), which can trigger students' experience of anxiety (Pekrun et al., 2009).

Our findings revealed that students' mathematics enjoyment was negatively related to mathematics procrastination. If students experienced enjoyment in the learning process, they may tend to continue pursuing the positive experience in the learning tasks, which further facilitates academic engagement and avoid academic procrastination. The negative relationships between positive emotions and procrastination were also found by some previous researchers (Balkis and Duru, 2016; Rahimi, 2019). For instance, Rahimi and Vallerand (2021) identified that students with positive emotions in their studies tended to have less procrastination in their academic assignments.

As hypothesized, students who experienced a higher level of mathematics anxiety were more inclined to procrastinate in the learning process. The close relationship between negative emotions and academic procrastination was also confirmed in many previous studies (Pychyl et al., 2000; Spada et al., 2006; Balkis and Duru, 2016; Rahimi, 2019; Rahimi and Vallerand, 2021). When students perceived anxiety in the learning process, they may delay the completion of academic tasks to escape from the negative emotions evoked by the assignment.

This study confirmed the mediational role of academic emotions between filial piety beliefs and procrastination in mathematics learning. Students who have gratitude to parents and hope to repay parents tended to experience more enjoyment in the learning process and further had less academic procrastination, while students who focused on obedience to parents and maintaining family honor were more likely to experience anxiety and thus engaged more in academic procrastination.

Our findings have significant implications for educational practices. Parents are suggested to establish intimate relationships and affection with their children, which can develop children's reciprocal filial piety (Yeh et al., 2013) and further promote the experience of positive emotions and reduce undesirable learning outcomes. Based on the negative nature of authoritarian filial piety in students' mathematics learning, parents need to avoid authoritarian parent–child interactions and reduce parental control in family contexts. Parents are suggested to create a democratic and harmonious family climate, which can minimize children's negative emotions and avoid them from engaging in academic procrastination.

Some limitations should be noted in the present study. First, this study used a cross-sectional design, so the results cannot reveal causality. Thus, caution should be made in explaining the findings in this study. Longitudinal research design is needed in future studies to check the causal relationship. Second, this study was only conducted with a sample of primary school students from three schools in China. The relationship between filial piety and academic emotions and procrastination may vary across different regions and educational levels. Further studies can check and expand our findings by investigating students in other countries and regions and other grade levels (e.g., secondary school and college). Third, this study only collected students' self-reported questionnaire data. Qualitative research methods (e.g., interview and observation) are needed to deeply explore the influence of filial piety on students' mathematics learning process and the potential reasons behind their relationship. Fourth, the control variables in this study only include student gender, grade, and parent educational level. In fact, students' prior mathematics achievement may significantly affect their emotion and procrastination. However, the present study did not control the influence of students' mathematics performance when examining the relationship between filial piety, emotions, and procrastination. Future studies need to take students' mathematics performance as a control variable in the relationship.

Conclusion

This study confirmed the association between students' filial piety beliefs and mathematics procrastination and the mediating role of mathematics enjoyment and anxiety in the association. The results suggested that students with reciprocal filial piety tended to have fewer procrastination behaviors. The reciprocal filial piety triggered more academic enjoyment, which may establish a protective mechanism against procrastination. We also identified that students who endorsed authoritarian filial piety were inclined to have more procrastination in mathematics learning. Students with authoritarian filial piety experienced more anxiety, which further led to more academic procrastination. These results expand our understanding of the relation between filial piety beliefs, mathematics emotions, and mathematics procrastination, which helps to unpackage the role of filial piety and academic emotions in Chinese students' mathematics learning process.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Faculty of Education, The University of Hong Kong. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

MG, YC, and XH contributed to conception and design of the study. MG collected the data and performed the statistical analysis. MG and XH wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Beijing Normal University Postdoctoral Research Fund (no. 110321001) and China Postdoctoral Science Foundation (no. 2022M710418).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Balkis, M., and Duru, E. (2016). Procrastination, self-regulation failure, academic life satisfaction, and affective well-being: Underregulation or misregulation form. *Eur. J. Psychol. Educ.* 31, 439–459. doi: 10.1007/s10212-015-0266-5

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238

Browne, M. W., and Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociol. Methods Res.* 21, 230–258. doi: 10.1177/0049124192021002005

Burnam, A., Komarraju, M., Hamel, R., and Nadler, D. R. (2014). Do adaptive perfectionism and self-determined motivation reduce academic procrastination? *Learn. Individ. Differ.* 36, 165–172. doi: 10.1016/j.lindif.2014.10.009

Cao, Z., Bishop, A., and Forgasz, H. (2007). Perceived parental influence on mathematics learning: a comparison among students in China and Australia. *Educ. Stud. Math.* 64, 85–106. doi: 10.1007/s10649-006-9033-5

Chen, W. W. (2014). The relationship between perceived parenting style, filial piety, and life satisfaction in Hong Kong. *J. Fam. Psychol.* 28, 308–314. doi: 10.1037/a0036819

Chen, W. W. (2016). The relations between filial piety, goal orientations and academic achievement in Hong Kong. *Educ. Psychol.* 36, 898–915. doi: 10.1080/01443410.2015.1008404

Chen, W. W., and Wong, Y. L. (2014). What my parents make me believe in learning: the role of filial piety in Hong Kong students' motivation and academic achievement. *Int. J. Psychol.* 49, 249–256. doi: 10.1002/ijop.12014

Chen, W. W., Yang, X., and Jiao, Z. (2022). Authoritarian parenting, perfectionism, and academic procrastination. *Educ. Psychol.* 42, 1145–1159. doi: 10.1080/01443410.2021.2024513

Cheng, H., and Furnham, A. (2003). Personality, self-esteem, and demographic predictions of happiness and depression. *Personal. Individ. Differ.* 34, 921–942. doi: 10.1016/S0191-8869(02)00078-8

Cheung, C. K., Lee, J. J., and Chan, C. M. (1994). Explicating filial piety in relation to family cohesion. *J. Soc. Behav. Pers.* 9, 565–580.

Chu, A. H. C., and Choi, J. N. (2005). Rethinking procrastination: positive effects of "active" procrastination behavior on attitudes and performance. *J. Soc. Psychol.* 145, 245–264. doi: 10.3200/SOCP.145.3.245-264

Gråstén, A., Jaakkola, T., Liukkonen, J., Watt, A., and Yli-Piipari, S. (2012). Prediction of enjoyment in school physical education. *J. Sports Sci. Med.* 11, 260–269. PMID: 24149199

Guo, M., Hu, X., and Leung, F. K. (2022). Culture, goal orientations, and mathematics achievement among Chinese students. *Int. J. Sci. Math. Educ.* 20, 1225–1245. doi: 10.1007/s10763-021-10202-0

Guo, M., Leung, F. K., and Hu, X. (2020). Affective determinants of mathematical problem posing: the case of Chinese Miao students. *Educ. Stud. Math.* 105, 367–387. doi: 10.1007/s10649-020-09972-1

Hajloo, N. (2014). Relationships between self-efficacy, self-esteem and procrastination in undergraduate psychology students. *Iran. J. Psychiatry Behav. Sci.* 8, 42–49. PMID: 25780374

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York: Guilford.

Ho, D. Y. F. (1994). Filial piety, authoritarian moralism, and cognitive conservatism in Chinese societies. *Genet. Soc. Gen. Psychol. Monogr.* 120, 349–365.

Ho, D. Y. F. (1996). "Filial piety and its psychological consequences," in *The handbook of Chinese psychology*. ed. M. H. Bond (New York, NY: Oxford University Press), 155–165.

Ho, D. Y. F., and Lee, L. Y. (1974). Authoritarianism and attitudes toward filial piety in Chinese teachers. *J. Soc. Psychol.* 92, 305–306. doi: 10.1080/00224545.1974.9923111

Howell, A. J., and Watson, D. C. (2007). Procrastination: associations with achievement goal orientation and learning strategies. *Personal. Individ. Differ.* 43, 167–178. doi: 10.1016/j.paid.2006.11.017

Hu, X., Leung, F. K., and Teng, Y. (2018). The influence of culture on students' mathematics achievement across 51 countries. *Int. J. Sci. Math. Educ.* 16, 7–24. doi: 10.1007/s10763-018-9899-6

Jiang, Y., and Dewaele, J. M. (2019). How unique is the foreign language classroom enjoyment and anxiety of Chinese EFL learners? *System* 82, 13–25. doi: 10.1016/j.system.2019.02.017

Jin, H., Wang, W., and Lan, X. (2019). Peer attachment and academic procrastination in Chinese college students: a moderated mediation model of future time perspective and grit. *Front. Psychol.* 10:2645. doi: 10.3389/fpsyg.2019.02645

Kim, K. R., and Seo, E. H. (2015). The relationship between procrastination and academic performance: a meta-analysis. *Personal. Individ. Differ.* 82, 26–33. doi: 10.1016/j.paid.2015.02.038

King, R. B., and McInerney, D. M. (2019). Family-support goals drive engagement and achievement in a collectivist context: integrating etic and emic approaches in goal research. *Contemp. Educ. Psychol.* 58, 338–353. doi: 10.1016/j.cedpsych.2019.04.003

Klassen, R. M., Ang, R. P., Chong, W. H., Krawchuk, L. L., Huan, V. S., Wong, I. Y., et al. (2010). Academic procrastination in two settings: motivation correlates, behavioral patterns, and negative impact of procrastination in Canada and Singapore. *Appl. Psychol.* 59, 361–379. doi: 10.1111/j.1464-0597.2009.00394.x

Klassen, R. M., Krawchuk, L. L., and Rajani, S. (2008). Academic procrastination of undergraduates: low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemp. Educ. Psychol.* 33, 915–931. doi: 10.1016/j.cedpsych.2007.07.001

Lee, F. K., Sheldon, K. M., and Turban, D. B. (2003). Personality and the goal-striving process: the influence of achievement goal patterns, goal level, and mental focus on performance and enjoyment. *J. Appl. Psychol.* 88, 256–265. doi: 10.1037/0021-9010.88.2.256

Leung, F. K. S. (2006). "Mathematics education in East Asia and the west: does culture matter?" in *Mathematics education in different cultural traditions-a comparative study of East Asia and the west*. eds. F. K. S. Leung, K. D. Graf and F. J. Lopez-Real (Boston, MA: Springer), 21–46.

Ma, X. Y., Ling, H., Zhang, J. R., Xiong, L., and Li, X. L. (2011). Related study on procrastination and parental rearing patterns in college students. Chinese. *J. Clin. Psychol.* 19, 675–680. doi: 10.16128/j.cnki.1005-3611.2011.05.025

Moon, S. M., and Illingworth, A. J. (2005). Exploring the dynamic nature of procrastination: a latent growth curve analysis of academic procrastination. *Personal. Individ. Differ.* 38, 297–309. doi: 10.1016/j.paid.2004.04.009

Niemiec, C. P., and Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: applying self-determination theory to educational practice. *Theory Res. Educ.* 7, 133–144. doi: 10.1177/1477878509104318

Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educ. Psychol. Rev.* 18, 315–341. doi: 10.1007/s10648-006-9029-9

Pekrun, R. (2019). "Achievement emotions: a control-value theory perspective" in *Emotions in late modernity*. eds. R. Patulny, A. Bellocchi, R. E. Olson, S. Khorana, J. McKenzie and M. Peterie (New York, NY: Routledge), 142–157.

Pekrun, R., Elliot, A. J., and Maier, M. A. (2009). Achievement goals and achievement emotions: testing a model of their joint relations with academic performance. *J. Educ. Psychol.* 101, 115–135. doi: 10.1037/a0013383

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., and Perry, R. P. (2011). Measuring emotions in students' learning and performance: the achievement emotions questionnaire (AEQ). *Contemp. Educ. Psychol.* 36, 36–48. doi: 10.1016/j.cedpsych.2010.10.002

Pychyl, T. A., Lee, J. M., Thibodeau, R., and Blunt, A. (2000). Five days of emotion: an experience sampling study of undergraduate student procrastination. *J. Soc. Behav. Pers.* 15, 239–245.

Rabin, L. A., Fogel, J., and Nutter-Upham, K. E. (2011). Academic procrastination in college students: the role of self-reported executive function. *J. Clin. Exp. Neuropsychol.* 33, 344–357. doi: 10.1080/13803395.2010.518597

Rahimi, S. (2019). Understanding academic procrastination: A longitudinal analysis of procrastination and emotions in undergraduate and graduate students. unpublished doctoral dissertation. Montreal, QC: McGill University.

Rahimi, S., and Vallerand, R. J. (2021). The role of passion and emotions in academic procrastination during a pandemic (Covid-19). *Personal. Individ. Differ.* 179:110852. doi: 10.1016/j.paid.2021.110852

Rickert, N. P., Meras, I. L., and Witkow, M. R. (2014). Theories of intelligence and students' daily self-handicapping behaviors. *Learn. Individ. Differ.* 36, 1–8. doi: 10.1016/j.lindif.2014.08.002

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020

Schraw, G., Wadkins, T., and Olafson, L. (2007). Doing the things we do: a grounded theory of academic procrastination. *J. Educ. Psychol.* 99, 12–25. doi: 10.1037/0022-0663.99.1.12

Senécal, C., Koestner, R., and Vallerand, R. J. (1995). Self-regulation and academic procrastination. *J. Soc. Psychol.* 135, 607–619. doi: 10.1080/00224545.1995.9712234

Soysa, C. K., and Weiss, A. (2014). Mediating perceived parenting styles–test anxiety relationships: academic procrastination and maladaptive perfectionism. *Learn. Individ. Differ.* 34, 77–85. doi: 10.1016/j.lindif.2014.05.004

Spada, M. M., Hiou, K., and Nikcevic, A. V. (2006). Metacognitions, emotions, and procrastination. *J. Cogn. Psychother.* 20, 319–326. doi: 10.1891/jcop.20.3.319

Steel, P. (2007). The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychol. Bull.* 133, 65–94. doi: 10.1037/0033-2909.133.1.65

Steel, P., and Klingsieck, K. B. (2016). Academic procrastination: psychological antecedents revisited. *Aust. Psychol.* 51, 36–46. doi: 10.1111/ap.12173

Sun, P., Fan, X., Sun, Y., Jiang, H., and Wang, L. (2019). Relations between dual filial piety and life satisfaction: the mediating roles of individuating autonomy and relating autonomy. *Front. Psychol.* 10:2549. doi: 10.3389/fpsyg.2019.02549

Sung, K. T. (1995). Measures and dimensions of filial piety in Korea. *Gerontologist* 35, 240–247. doi: 10.1093/geront/35.2.240

Tao, V. Y., and Hong, Y. Y. (2014). When academic achievement is an obligation: perspectives from social-oriented achievement motivation. *J. Cross Cult. Psychol.* 45, 110–136. doi: 10.1177/0022022113490072

Wolters, C. (2003). Understanding procrastination from a self-regulated learning perspective. *J. Educ. Psychol.* 95, 179–187. doi: 10.1037/0022-0663.95.1.179

Wolters, C. A. (2004). Advancing achievement goal theory: using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *J. Educ. Psychol.* 96, 236–250. doi: 10.1037/0022-0663.96.2.236

Xue, X., Wang, Y., Li, H., Gao, J., and Si, J. (2021). The association between mathematical attitudes, academic procrastination and mathematical achievement among primary school students: the moderating effect of mathematical metacognition. *Curr. Psychol.*, 1–12. doi: 10.1007/s12144-021-02133-4

Yan, J., and Chen, W. W. (2018). "The relationships between filial piety, self-esteem, and life satisfaction among emerging adults in Taiwan" in *Close relationships and happiness across cultures*. eds. M. Demir and N. Seumer (Cham: Springer), 151–164.

Yang, Z., Asbury, K., and Griffiths, M. D. (2019). An exploration of problematic smartphone use among Chinese university students: associations with academic anxiety, academic procrastination, self-regulation and subjective wellbeing. *Int. J. Ment. Heal. Addict.* 17, 596–614. doi: 10.1007/s11469-018-9961-1

Yeh, K. H. (2003). "The beneficial and harmful effects of filial piety: an integrative analysis," in *Progress in Asian social psychology: Conceptual and empirical contributions*. eds. K. S. Yang, K. K. Hwang, P. S. Pederson and I. Daibo (Santa Barbara, CA: Greenwood), 67–82.

Yeh, K. H. (2006). The impact of filial piety on the problem behaviours of culturally Chinese adolescents. *J. Psychol. Chin. Soc.* 7, 237–257.

Yeh, K. H., and Bedford, O. (2003). A test of the dual filial piety model. *Asian J. Soc. Psychol.* 6, 215–228. doi: 10.1046/j.1467-839X.2003.00122.x

Yeh, K. H., and Bedford, O. (2004). Filial belief and parent-child conflict. *Int. J. Psychol.* 39, 132–144. doi: 10.1080/00207590344000312

Yeh, K. H., Yi, C. C., Tsao, W. C., and Wan, P. S. (2013). Filial piety in contemporary Chinese societies: a comparative study of Taiwan, Hong Kong, and China. *Int. Sociol.* 28, 277–296. doi: 10.1177/0268580913484345

Zakeri, H., Esfahani, B. N., and Razmjoee, M. (2013). Parenting styles and academic procrastination. *Procedia Soc. Behav. Sci.* 84, 57–60. doi: 10.1016/j.sbspro.2013.06.509

Zhang, J., and Bond, M. H. (1998). Personality and filial piety among college students in two Chinese societies: the added value of indigenous constructs. *J. Cross Cult. Psychol.* 29, 402–417. doi: 10.1177/0022022198293002

Zhou, J., Guo, Q., and Xu, R. (2020). Reciprocal filial piety facilitates academic success via autonomy: generalizing findings in Chinese society to a global context. *Front. Psychol.* 11:69. doi: 10.3389/fpsyg.2020.00069

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 07 March 2023
DOI 10.3389/fpsyg.2023.1126047

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Lianchun Dong,
Minzu University of China, China
Wenjun Zhao,
Sichuan Normal University, China

*CORRESPONDENCE
Dingliang Tan
✉ tandingliang@njnu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 17 December 2022
ACCEPTED 20 February 2023
PUBLISHED 07 March 2023

CITATION
Shi L, Dong L, Zhao W and Tan D (2023) Improving middle school students' geometry problem solving ability through hands-on experience: An fNIRS study. *Front. Psychol.* 14:1126047. doi: 10.3389/fpsyg.2023.1126047

Improving middle school students' geometry problem solving ability through hands-on experience: An fNIRS study

Licheng Shi[1†], Linwei Dong[2†], Weikun Zhao[3] and Dingliang Tan[4*]

[1]School of Education Science, Nantong University, Nantong, China, [2]Jiangsu Institute of Education Sciences, Nanjing, China, [3]Yulong Road Experimental School, Yancheng, China, [4]School of Education Science, Nanjing Normal University, Nanjing, China

Hands-on learning is proposed as a prerequisite for mathematics learning in kindergarten and primary school. However, it remains unclear that whether hands-on experience aids understanding of geometry knowledge for middle school students. We also know little about the neural basis underlying the value of hands-on experience in math education. In this study, 40 right-handed Chinese students (20 boys and 20 girls) with different academic levels were selected from 126 seventh-grade students in the same school, who learnt "Axisymmetric of an Isosceles Triangle" in different learning style (hands-on operation vs. video observation). Half of them operated the concrete manipulatives while the other half watched the instructional videos. The learning-test paradigm and functional near-infrared spectroscopy (fNIRS) technique were used to compare the differences in geometry reasoning involved in solving well-structured problems and ill-structured problems. Behavioral results showed that hands-on experience promoted students' performances of geometry problem-solving. Students with lower academic level were more dependent on hands-on experience than those with higher academic level. The fNIRS results showed that meaningful hands-on experience with concrete manipulatives related to learning contents increased reactivation of the somatosensory association cortex during subsequent reasoning, which helped to improve the problem-solving performance. Hands-on experience also reduced students' cognitive load during the well-structured problem-solving process. These findings contribute to better understand the value of hands-on experience in geometry learning and the implications for future mathematics classroom practices.

KEYWORDS

problem-solving, geometry learning, hands-on experience, embodied cognition, educational neuroscience, fNIRS

Introduction

Problem solving is a process whereby learners use the knowledge and skills acquired in the past to seek solutions in order to adapt to the needs of the situation (Kahney, 1993). According to Mayer (1998), problem-solving requires students to integrate problem information and maintain mental images of the problem in working memory, and then estimate a reasonable answer and check to make sure everything is accurate. Improving problem-solving performance from the perspective of cognitive psychology has become a research problem that many scholars and mathematics educators pay attention to (Montague et al., 2014). Geometry curriculums in middle school help students develop geometric intuition and promote spatial reasoning ability. Success in geometry problem-solving is highly correlated with students' math academic achievement in school and beyond (Krawec, 2014).

According to Compulsory Education Mathematics Curriculum Standards (Ministry of Education of the People's Republic of China, 2012, 2011th Edition) in China, students should have enough time to experience observation, experiment, guess, calculation, reasoning, verification, and other activities in math learning. Mathematical experience needs to be gradually accumulated in the process of doing and thinking. The Compulsory Education Mathematics Curriculum Standards (2022 Edition) point out that it is important to improve teaching methods by mathematical experiments and visualize abstract mathematical knowledge to promote students' understanding of mathematical concepts and construction of mathematical knowledge. There is accumulating evidence to suggest that hands-on experience plays an important role in math performance. Theories of embodied cognitive hold that cognitive processes are deeply rooted in the body's interactions with the physical world (Wilson, 2002). According to the theory of embodied mathematics (Giardino, 2018), touching concrete manipulatives contributes to the formation of enriched mental representations, which are beneficial for mathematical learning. In other words, instruction should be more focused on perception that emerge from dynamic interactions, such as folding, drawing, measuring, or manipulating objects. Teachers should encourage students to link these experiences to abstract ideas, rather than concentrate purely on abstract rules (Pouw et al., 2014; Novak and Schwan, 2021). From the viewpoint of perceptual symbol system, the combination of visual, auditory and tactual modalities enriches mental representation (Hutmacher and Kuhbandner, 2018). Memory traces are better understood in terms of sensorimotor encoding (Giardino, 2018; Palmiero et al., 2019), and thus may support the development of more complex understandings (Novack et al., 2014; Kontra et al., 2015; Stull et al., 2018).

Hands-on experience is usually inspired by manipulative materials (Swan and Marshal, 2010; Abrahamson and Sánchez-García, 2016; Nathan and Walkington, 2017). For example, origami/paper folding, the number board game, counters, Cuisenaire rods, Unifix cubes, paper money, pattern blocks, base-10 blocks etc., are often used in the math classroom. A recent study by Freina et al. (2018) found that digital games supported the development and consolidation of visuospatial abilities in students of the last 2 years of the primary school, and such a training would have a positive impact on their mathematics performance. Similar results were obtained by Otten et al. (2020). Fifth-grader students in primary school who worked with the physical balance model more often used advanced algebraic strategies and made a larger improvement in their algebraic reasoning. Hands-on experience blended within instruction was beneficial to understand mathematics concepts and solve real-world problems. Additionally, students who attending the hands-on activities were more likely to recall their experience from a first-person perspective instead of a third-person observer. Shifting between perspectives helped students enrich mental representations and develop a deeper understanding, leading to greater learning gains (Smith, 2018).

However, some researchers dispute the presence of physicality as a requirement and claim that hands-on experience is not always beneficial in mathematics learning (Carbonneau et al., 2013; Rau and Herder, 2021). Thus, despite the ever-increasing researches on embodied learning, there continues to be a need for understanding children's hands-on experience as it exists in classroom settings.

Firstly, do all students need hands-on experience in mathematical learning? Referring to Boakes (2009) and Burte et al. (2017), students in elementary schools and middle schools took part in hands-on tasks, such as folding and cutting paper, to develop spatial thinking from two-dimensional to three-dimensional. Results showed that all students in elementary schools improved performance on real-world math problems and older students (Grades 5 and 6) improved performance on visual and spatial math problems, while students in middle schools did as well as they did without it. The effectiveness of hands-on activities couldn't be guaranteed, as age differences should be taken into consideration. According to Piagetian theory, students have the ability to form mental images of abstract concepts and process mental operations of those images to solve mathematical problems given after the age of 11. Some researchers hold the view that hands-on experience wouldn't enhance math problem-solving at least from the fourth grade onward, for students who have entered the formal operational stage and had the ability of abstract thinking (for an overview see Brown et al., 2009; Carbonneau et al., 2013). However, other researchers provide evidence that hands-on experience brought by physical manipulatives promotes math learning, regardless of the age of the student (Novak and Schwan, 2021). To fill this research gap, the present study is designed to explore whether it is necessary to understand geometric knowledge with the help of hands-on experience for students entering the formal operational stage.

Secondly, does hands-on experience promote or hinder mathematical learning? Researchers hold different views. For mathematics learning that emphasizes abstract thinking training, perceptual richness obtained by manipulatives has been identified as a potential deterrent (McNeil and Uttal, 2009). For example, 11-year-olds and undergraduate students transfer mathematics knowledge more successfully from abstract and symbolic representations than from multiple concrete examples (Kaminski et al., 2009). The perceptual experience that students construct when manipulating objects is often specific to the learning situations, making it difficult to transfer to other contexts or to more abstract knowledge (Brown et al., 2009). However, some researchers pick up the opposite view. As has been discussed above, physical manipulation also can induce psychological simulation of learning materials, which helps to establish a relational link between perceptual experience and abstract symbols, thus promote the development of mathematical thinking with content "visualization" (Novak and Schwan, 2021) and facilitate knowledge retention and transfer (Rau and Herder, 2021). Physicality becomes a cognitive anchor for comprehending abstract knowledge during math learning (Pouw et al., 2014).

Thirdly, how hands-on experience affect geometry problem-solving? Based on the theory of embodied cognition, prior researches concerned about students' conceptual learning (Pexman, 2017; Rau and Herder, 2021), science learning (Zacharia and Olympiou, 2011), chemistry learning (Stull et al., 2018), physics learning (Kontra et al., 2015), and discussed the mechanism underlying the value of gestures or body movements in problem-solving (Chu and Kita, 2011; Vallotton et al., 2015; Smith, 2018). Few studies have focused on the impact of hands-on experience on student's geometric reasoning. Existing studies using functional magnetic resonance imaging (fMRI) to examine the state of the cerebral cortex during physical manipulation have found that, compared with observational learning and mental simulation, physical manipulation activates the sensorimotor areas of the frontoparietal cortex to a greater extent, such as primary motor Cortex, premotor cortex, somatosensory cortex, somatosensory association cortex (Bellebaum et al., 2013; Kontra et al., 2015). Whether this is the case for the learning of geometric concepts is unclear. At the same time, meta-analysis confirmed that the neural mechanism of mathematical learning depends on the coordination function of the left prefrontal cortex and parietal cortex area (Artemenko et al., 2019; Molina et al., 2019). If the subject is right-handed, physical manipulation of materials by the right hand and arm consequently results in increased left hemisphere activation (Casasanto, 2009, 2016; Jin et al., 2020). Therefore, in this study, we focused on the frontal and parietal areas of the left hemisphere to explore the neural basis of hand-on experience affects the geometry problem solving.

Functional near-infrared spectroscopy (fNIRS) is a neuroimaging technique that uses near-infrared light to monitor metabolic/hemodynamic changes related to neural activity. Compare to other neuroimaging techniques, fNIRS is portable, lightweight, less sensitive to motion artifacts (Bahnmueller et al., 2014; Quaresima and Ferrari, 2019), which represents a good compromise in terms of spatial and temporal resolution. Moreover, recent development of wireless fNIRS devices has opened the way for new applications in educational neuroscience research, which can be more sensitive and accurate in assessing cognitive function in real-world tasks (Pinti et al., 2015; Soltanlou et al., 2018).

Most of the educational neuroscience research on the effects of hands-on experience use electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). fNIRS imposes fewer physical constraints than EEG and allows free movement in a more natural environment than fMRI, which has great advantages in brain imaging studies that require motor participation in children and is suitable for use in real classroom settings (Piper et al., 2014; Scarapicchia et al., 2017; Soltanlou et al., 2018).

Research objectives and research questions

In the present study, we aim to investigate whether hands-on experience is a prerequisite for geometry learning in middle school. By hands-on, we talk about the actual and active touch of concrete material related to learning contents under the guidance of math teachers. The opinions on the effect of hands-on experience on children's math learning are inconsistent, especially for the research focused on students entering the formal operational stage. At the same time, math ability is bound to affect mathematical behavior and cognitive neural processes (Artemenko et al., 2019). But existing studies have different views on the structure of math ability, and viewpoints on the assessment of math ability are also inconsistent. Academic level can reflect math ability to a certain extent (Dupeyrat et al., 2011; Weissgerber et al., 2022), so the present study takes academic level as an important factor when explores the impact of hands-on experience on geometry learning in more detail (Artemenko et al., 2019; Rivella et al., 2021).

In summary, three questions direct this study.

Research Question 1: Does hands-on experience aid middle school students' understanding of geometry knowledge?

Research Question 2: Is hands-on experience more conducive to geometry problem solving for high-achieving students or low-achieving students?

Research Question 3: What is the neural basis underlying the value of hands-on experience in math education?

Materials and methods

Participants

A total of 40 right-handed Chinese students (20 boys and 20 girls, *M*age = 13.58 years, *SD* = 0.32) with different academic levels were recruited from 126 seventh-grade students in the same school in China, who had no prior instruction on related contents. Existing studies have not been consistent in the division of academic levels, such as the top 50% and bottom 50% of students in the class (Chang and Hsin, 2021), or the top 25% and bottom 25% of students (Sermier and Bless, 2013), or the top 20% and bottom 20% of students (Chen et al., 2017) are defined as high and low academic levels, respectively. In order to eliminate the experimental error caused by students with extreme grades, the average of the last two math scores was converted into the standard score in the class. The top 10–27% were defined as high academic level (standard scores = 1.28, *SD* = 0.21), while the bottom 10–27% of the rankings were defined as low academic level (standard scores = −1.24, *SD* = 0.22) in our study. A total of 40 students were selected and arranged into four conditions: A group of 10 students with high academic level under operation condition (operation + H condition), a group of 10 students with low academic level under operation condition (operation + L condition), a group of 10 students with high academic level under observation condition (observation + H condition), and a group of 10 students with low academic level under operation condition (observation + L condition). All students had normal or corrected-to-normal vision and no brain disease. The students and their parents in this study were provided written informed consent before participating in this research and were awarded a present (about $10) for their participation. The present study was carried out in accordance with the ethical standards and requirements established by the local ethical committee of our institute.

Experimental design

We used a 2 × 2 between-subjects design with the factors hands-on experience (operation/observation) and academic level (high/low). The students of different academic level were randomly assigned to operational condition and observational condition. Dependent variables were the accuracy and reaction time for the problem-solving tasks, as well as the changes of oxyhemoglobin (HbO2) concentration in the left frontal and parietal cortex during the learning process and problem-solving process.

Experimental tasks

Taking "Axisymmetric of an Isosceles Triangle" as learning content, participants in the operation group turned and folded the paper along the bisector of the top angle to explore the properties of an isosceles triangle, and finally summarized the main points of knowledge in 3-min. Participants in the observation group watched a 3-min video lecture, without operating the concrete manipulatives.

In educational psychology, a problem is a situation in which there is some obstacle between the initial state and the goal state that needs to be overcome. It can be divided into well-structured problem and ill-structured problem (Reed, 2016). Well-structured problems can be represented by a problem space consisting of well-defined initial states and goal states that are connected by legal moves. In contrast, the initial states, the goal states and the intermediate states of ill-structured problems are incompletely specified. In this study, well-structured problems included 10 judgment tasks, of which two were practice tasks and eight were formal tasks. For example, in ΔABC, if $AB = BC$, $AD \perp BC$, then $\angle BAD = \angle CAD$, is it right? Students solved the problems by interpreting and reasoning according to the properties of triangles. Stimuli were randomly presented. Ill-structured problem emphasized the application of knowledge in real life, which tested the participants' abilities to transfer. The problem was as follows: A worker encountered a problem when building a house. He wanted to know whether the beams of the house were horizontal? An old man solved the problem using an isosceles triangle and a plumb bob tied with a string. The worker didn't know the reason. How about you?

The well-structured tasks and ill-structured tasks were developed by math teachers in middle school. The well-structured tasks were selected from after-school practices consistent with the learning content. The ill-structured tasks came from the mathematics textbook of eighth grade. The materials were drawn using Auto-CAD software (see **Figure 1**).

Experimental procedure

The participants were tested individually in a dimly lit and noise-free room in the middle school. The lighting is particularly important, given that bright light can affect fNIRS signals (Shadgan et al., 2010). Each participant was seated approximately 70 cm in front of a 21-inch computer monitor (refresh rate 60 Hz, 1,920 × 1,080 resolution) and worn an fNIRS helmet throughout the tasks.

The learning process lasted 3 min. Instructions for the operation group were: "Welcome to our experiment. For the next 3 min, you can fold the isosceles triangle paper beside your hands to explore the nature of an isosceles triangle. If you're ready, and we'll start." The instructions for the observation group were: "Welcome to this activity. For the next 3 min, please watch the video lecture carefully, with no gesture or body movement. If you're ready, and we'll start."

After a 5-min rest, the participants solved the problems. As to the well-structured problem, each trial began with a red fixation cross in the center of the screen for 1 s, followed by the stimulus. When the response was given, the next trial started automatically. There were eight trials in total. As to the ill-structured problem, participants thought for 40 s, then answered verbally. In order to exclude the influence of the experimental sequence, half of the participants solved the well-structured problems first and then the ill-structured problems, while the other half of the participants did the opposite.

This research lasted about 30 min for each participant. The chart of the experimental procedure was shown in **Figure 2**.

Data recording

Behavioral data were captured using the E-Prime 3.0 (Psychology Software Tools).[1] The software tool recorded information on accuracy and response times in well-structured tasks. The answers of ill-structured tasks were recorded with a voice recorder. Further, three mathematics teachers in middle school were asked to independently rate each subject's answer with a score ranging from 0 to 10. The average score determined the subject's final score on the task. The scorer reliability was 0.86.

We used LIGHTNIRS system (Shimadzu Corp., Kyoto, Japan) with three wavelengths of near-infrared light (780, 805, and 830 mm) to record the absorption changes of oxyhemoglobin (HbO_2), deoxyhemoglobin (HbR) and total-hemoglobin (total-Hb). A total of 20 measurement channels (eight sources, eight detectors, source-detector distance was 25.00 mm on average) in a fNIRS helmet covered the prefrontal and parietal cortex in the left hemisphere (see **Figure 3** and **Table 1**).

In order to determine the anatomical areas underneath the channels, the 3D localizer (FASTRAK, Polhemus, Colchester, VT, USA) were utilized to confirm the positions of CZ, NZ, Al, AR, and the probes. The coordinate of the channels was positioned and aligned with the automatic anatomical labeling (AAL) atlas and Brodmann's area (BA) in SPM software. See **Table 1** for details.

Data analysis

Two-way analysis of variance (ANOVA) was conducted on the accuracy (ACC) and response time (RT), to determine whether there were significant differences between the four different conditions in well-structed problem solving tasks. As to ill-structed problem, ACC was analyzed.

1 http://www.pstnet.com/eprime.cfm

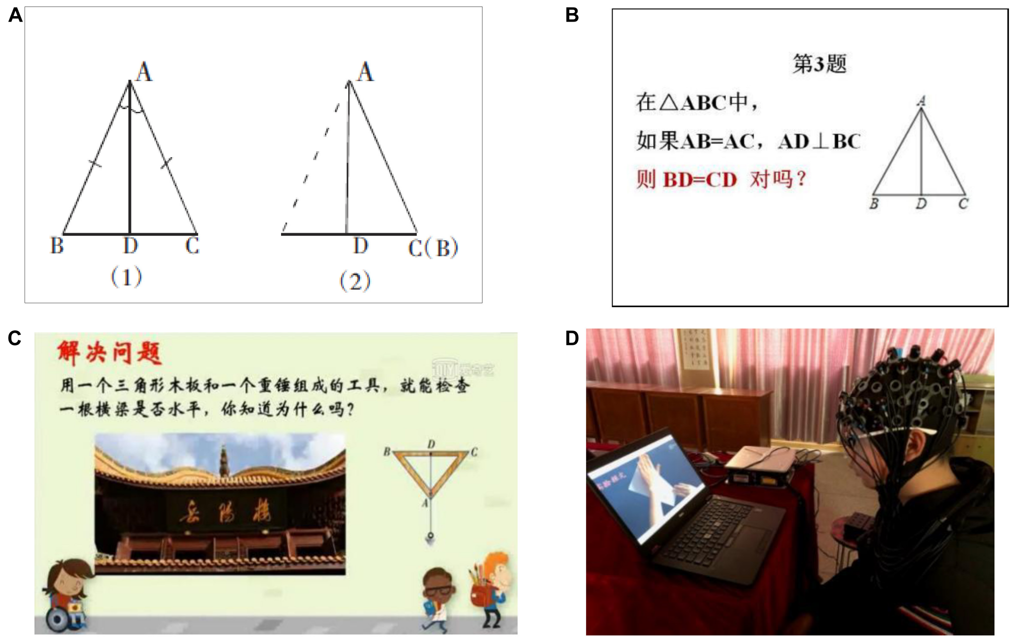


FIGURE 1
(A) Learning content, axisymmetric of an isosceles triangle. **(B)** An example of well-structured problem. **(C)** Ill-structured problem. **(D)** Experimental scenario. See the text for additional details.

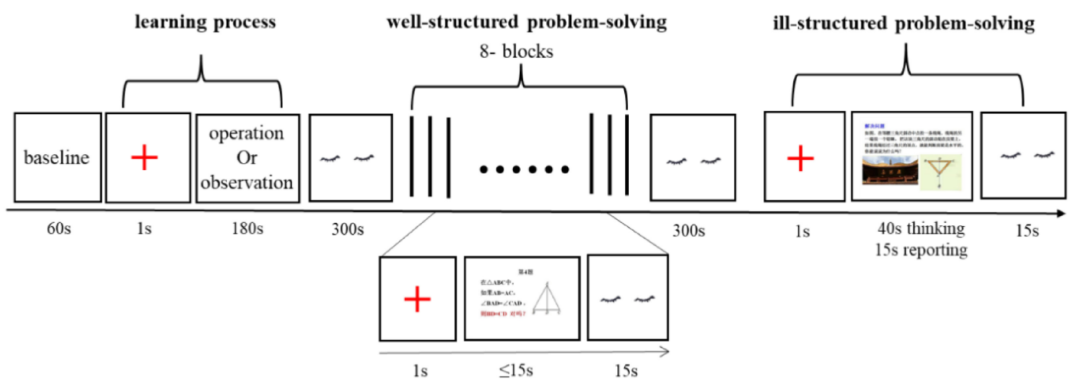


FIGURE 2
Experimental procedure.

Functional near-infrared spectroscopy data analysis was carried out using NIRS_SPM software (Ye et al., 2009) and matlab2014a (MathWorks, Natick, Ma, USA). Modified Beer-Lambert law was used to transform measured optical densities into hemoglobin concentration. For each channel, the original fNIRS data were low-pass filtered and high-pass filtered through wavelet-based methods of hemodynamic response function (HRF) and Discrete Cosine Transform (DCT) with a cutoff period of 128 s to remove motion artifacts and physiological noise induced by heartbeat, breathing cycle and low frequency oscillations of blood pressure (Ye et al., 2009; Tak and Ye, 2014). The mean changes in HbO_2 and HbR concentration were obtained using the last 30 s of the resting state before the beginning of the task as a baseline. Finally, the general linear model (GLM) was used to calculate the individual β values for each channel, participant, and task. We focused on HbO_2, which has the highest sensitivity to changes in cerebral blood flow (Okamoto et al., 2004; Hoge et al., 2005), to assess the participants' brain activation. All the results were corrected using the false discovery rate (FDR), and the adjusted significant level of *p*-value was set at 0.05 (Singh and Dan, 2006). With the help of EasyTopo toolbox (Tian et al., 2013), two-dimensional plane images were output to display the location of activated brain regions.

Results

Considering that behavior results need to correspondence to fNIRS data, two participants had to be excluded because they didn't follow the instructions properly or because of technical problems in response recording. Math standard scores were analyzed by 2 × 2 ANOVA. There were significant differences between academic

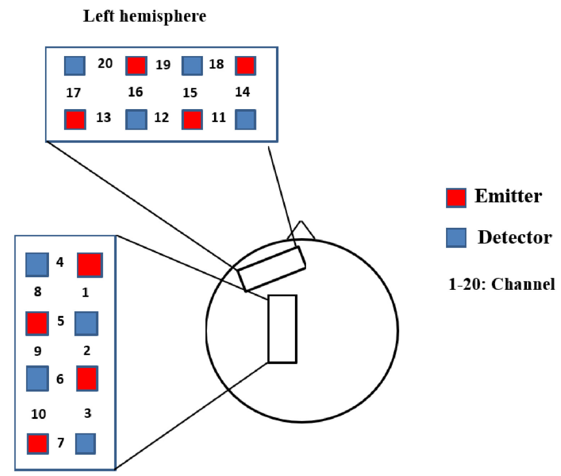


FIGURE 3
Optical probe positions.

level [$F_{(1,34)}$ = 613.00, p < 0.001, η^2_p = 0.95] but not hands-on experience [$F_{(1,34)}$ = 0.01, p > 0.05], and no significant interaction effect was found [$F_{(1,34)}$ = 0.13, p > 0.05]. It meant that the manipulation of variables was still valid.

Behavioral results

Well-structured problem-solving

A two-way ANOVA with hands-on experience and academic level as between-subjects factors, showed a significant main effect of the hands-on experience on the accuracy [$F_{(1,34)}$ = 4.34, p = 0.045, η^2_p = 0.11, 95% CI: 0.53∼0.73] and reaction time [$F_{(1,34)}$ = 4.39, p = 0.044, η^2_p = 0.11, 95% CI: 6847.76∼8273.06]. The main effect of academic level was also significant on the accuracy [$F_{(1,34)}$ = 6.04, p = 0.019, η^2_p = 0.15, 95% CI: 0.53∼0.74] and reaction time [$F_{(1,34)}$ = 9.85, p = 0.004, η^2_p = 0.23, 95% CI: 6687.32∼8411.80]. Interaction effect between hands-on experience and academic level was not found, suggesting overall a consistent trend in learning gains between the conditions.

Ill-structured problem-solving

We only recorded the accuracy of ill-structured problem-solving, and analysis of variance showed that both hands-on experience [$F_{(1,34)}$ = 22.31, p < 0.001, η^2_p = 0.39, 95% CI: 3.98∼6.40] and academic level [$F_{(1,34)}$ = 90.90, p < 0.001, η^2_p = 0.73, 95% CI: 3.22∼7.18] had significant main effects. There was a significant interaction effect between hands-on experience and academic level [$F_{(1,34)}$ = 5.37, p = 0.027, η^2_p = 0.14, 95% CI: 4.17∼7.73]. Further simple effect analysis found that, for those with low academic level, the accuracy of the operation group was higher

TABLE 1 Positions of the fNIRS channels.

Channels	MNI			AAL		BA	
	x	y	z	Region (L)	Coverage ratio	Region (L)	Coverage ratio
Ch1	−18	−44	78	SPL	0.46	5- SAC	0.77
Ch2	−27	−57	72	SPL	1.00	7- SAC	0.94
Ch3	−24	−66	68	SPL	0.97	7- SAC	1.00
Ch4	−24	−26	75	PCG	0.60	4- PMC	0.78
Ch5	−33	−45	72	SPL	0.69	5- SAC	0.59
Ch6	−34	−57	69	SPL	0.94	7- SAC	0.85
Ch7	−29	−68	64	SPL	1.00	7- SAC	1.00
Ch8	−40	−28	70	POCG	0.68	4- PMC	0.62
Ch9	−42	−44	66	IPL	0.42	40- Wernicke's area	0.71
Ch10	−40	−58	63	SPL	0.59	7- SAC	0.61
Ch11	−22	53	41	SFG	0.83	9- DLPFC	0.87
Ch12	−45	36	37	MFG	0.94	9- DLPFC	0.59
Ch13	−56	11	37	MFG	0.72	9- DLPFC	0.66
Ch14	−10	66	28	SFG	0.57	10- FPC	0.92
Ch15	−38	56	24	MFG	0.96	10- FPC	0.90
Ch16	−53	36	21	IFGtriang	0.83	46- DLPFC	0.92
Ch17	−63	6	28	MFG	0.47	44- Broca's area	0.86
Ch18	−26	68	10	SFG	0.86	10- FPC	1.00
Ch19	−46	53	5	MFG	0.53	10- FPC	0.65
Ch20	−57	31	9	IFGtriang	1.00	45- Broca's area	0.64

The location data listed in this table is the area with the greatest coverage probability. SPL, superior parietal lobe; IPL, inferior parietal lobe; POCG, postcentral gyrus; SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFGtriang, Inferior frontal gyrus, triangular part; DLPFC, dorsolateral prefrontal cortex; SAC, somatosensory association cortex; PMC, primary motor cortex; FPC, frontopolar cortex.

TABLE 2 Behavioral performance of problem-solving tasks.

Measures (n = 38)	Well-structured problem-solving		Ill-structured problem-solving
	Accuracy	Reaction time (ms)	Accuracy
Operation + H condition (n = 10)	0.72 (0.11)	6882.79 (844.43)	7.10 (0.74)
Operation + L condition (n = 10)	0.63 (0.15)	7615.17 (858.01)	4.80 (1.23)
Observation + H condition (n = 9)	0.65 (0.10)	7316.28 (957.40)	6.33 (0.71)
Observation + L condition (n = 9)	0.54 (0.11)	8383.99 (875.25)	2.56 (1.13)

H, high academic level; L, low academic level.

than that of the observation group [$F_{(1,34)}$ = 20.79, p < 0.000, η^2_p = 0.31], with no significant difference in high academic level students [$F_{(1,34)}$ = 2.89, p = 0.098]. The accuracy of high-achieving students was higher than low-achieving students, both in the operation group [$F_{(1,34)}$ = 27.48, p < 0.000, η^2_p = 0.45] and in the observation group [$F_{(1,34)}$ = 66.73, p < 0.000, η^2_p = 0.66]. Descriptive data are shown in Table 2.

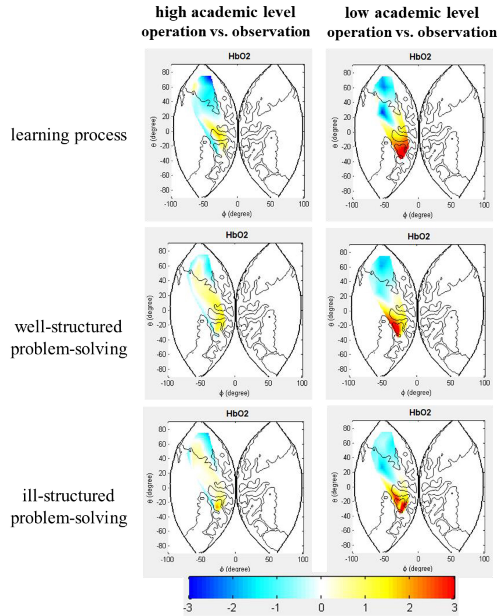


FIGURE 4
Brain activation images during the learning process and problem-solving process.

Functional near-infrared spectroscopy results

We analyzed the HbO_2 data in all 20 channels measured by fNIRS (see Figure 4). During the learning process, the result demonstrated that the main effects of hands-on experience were significant in SAC [Ch5: $F_{(1,34)}$ = 9.94, p = 0.030, η^2_p = 0.23, 95% CI:−0.002∼0.012] and PMC [Ch8: $F_{(1,34)}$ = 14.02, p = 0.020, η^2_p = 0.29, 95% CI:−0.002∼0.014]. A significantly greater activation was founded in the operation group than in the observation group. The main effect of academic level was significant in DLPFC [Ch13: $F_{(1,34)}$ = 12.75, p = 0.020, η^2_p = 0.27, 95% CI:0.004∼0.013]. HbO_2 variations in the low-achieving group were significantly higher than that in the high-achieving group. Interaction effect between hands-on experience and academic level was significant in SAC [Ch10: $F_{(1,34)}$ = 10.83, p = 0.040, η^2_p = 0.24, 95% CI:−0.002∼0.011]. Further simple effect analysis illustrated that for those with low academic level, a significant increase of HbO_2 concentration was observed in Ch10 [$F_{(1,34)}$ = 15.43, p < 0.000, η^2_p = 0.31] in the operation group compared to the observation group. No significant difference between the learning conditions was found in high-achieving students. Meanwhile, the HbO_2 variations in the high-achieving group were significantly higher than that in the low-achieving group in Ch10 [$F_{(1,34)}$ = 13.05, p = 0.001, η^2_p = 0.28] in the observation group, with no significant difference in the operation group.

During the problem-solving process, for the left SAC, the results demonstrated that the main effects of the hands-on experience were significant both in the well-structured problem-solving tasks [Ch5: $F_{(1,34)}$ = 9.80, p = 0.04, η^2_p = 0.22, 95% CI:0.002∼0.010; Ch6: $F_{(1,34)}$ = 11.57, p = 0.04, η^2_p = 0.25, 95% CI:0.001∼0.010] and in the ill-structured problem-solving tasks [Ch6: $F_{(1,34)}$ = 12.22, p = 0.020, η^2_p = 0.26, 95% CI:0.002∼0.010; Ch7: $F_{(1,34)}$ = 12.21, p = 0.010, η^2_p = 0.15, 95% CI: 0.002∼0.010]. A significantly higher activation in the operation group than in the observation group. No other significant effects were discovered (ps > 0.05). For the left DLPFC, the main effects of hands-on experience and academic level were significant in the well-structured problem-solving tasks (ps < 0.05). HbO_2 variations in the low-achieving group were significantly higher than that in the high-achieving group [Ch13: $F_{(1,34)}$ = 12.58, p = 0.02, η^2_p = 0.27, 95% CI:0.003∼0.008]. HbO_2 variations in the observation group were significantly higher than that in the operation group [Ch11: $F_{(1,34)}$ = 10.36, p = 0.02, η^2_p = 0.23, 95% CI:0.001∼0.011].

The relationship between brain activity and behavioral performance

Based on the fNIRS results, we concerned the brain regions of interest (ROI) correspond to the following Brodmann divisions: left SAC(Ch1,2,3,5,6,7,10), left DLPFC(Ch11,12,13,16), and HbO_2 concentration is averaged between channels in each ROI before statistical analysis.

Pearson's correlation analyses revealed a significantly correlation between the left SAC activity and problem-solving performance. The activation in left SAC was positively correlated with the accuracy (r = 0.448, p = 0.005) and negatively correlated with the reaction time (r = −0.441, p = 0.006) of well-structured problem-solving. Moreover, the activation of left SAC was positively correlated with the quiz score of ill-structured problem-solving (r = 0.470, p = 0.003). Obtaining this result proved that the extent of the sensorimotor brain system's involvement caused by hands-on experience is related to geometry reasoning.

Pearson's correlation analyses also revealed that the activation in left DLPFC was positively correlated with the reaction time of well-structured problem-solving (r = 0.432, p = 0.007), and negatively correlated with the quiz scores of ill-structured problem-solving (r = −0.471, p = 0.003).

Discussion

In the present study, we demonstrated positive effects of hands-on experience on geometry knowledge acquisition in middle-school. The knowledge of properties of the isosceles triangle was implemented in two different ways (hands-on operation vs. video observation). Students' geometry reasoning was examined through solving well-structured problem and ill-structured problem.

Our results showed that students who gained the knowledge of "properties of isosceles triangle" by hands-on operation outperformed students who observed the same phenomena, both in well-structured problem-solving tasks and in ill-structured problem-solving tasks, indicating the beneficial role of hands-on experience in geometry learning. More importantly, hands-on experience improved the students' performance across ability levels. Therefore, it could be reasonably against the viewpoints that both perceptual and interactive richness of instructional manipulatives would hinder symbolic inferences and compromise application of knowledge to new situations, thus resulting in lower learning outcomes (Brown et al., 2009; Kaminski et al., 2009; Pouw et al., 2014). These results supported the findings of Kontra et al. (2015) in physics learning, Stull et al. (2018) in chemistry learning, and Zacharia and Olympiou (2011) in science learning. Hands-on experience tied to the to-be-learned content could promote learning, even if learners have already had the ability to think abstractly.

However, our main interest was how hands-on experience enhance student's mathematics learning. Consistent with many previous studies, the operation group explored the characteristics of the isosceles triangle by folding and turning the paper, thus increasing the activation of the primary motor cortex (Bellebaum et al., 2013; Kontra et al., 2015). Compared with observational learning, hands-on operation required the integration of vision, haptics and other sensory information, and made the somatosensory association cortex activated. Later on, SAC reactivated during well- structured and ill-structured problems solving, contending that memory traces capture and reflect the perceptual components of past experience. Such evidences supported the reactivation hypothesis (Nyberg et al., 2001; Ianì, 2019), which amounted to the idea that memory trace consisted in a reactivation of the same sensorimotor regions initially activated in its perception. Similar results were obtained by Kontra et al. (2015). Activation in SAC was also positively correlated with performances in problem solving. Our findings offered a possible explanation for how hands-on experience enhanced understanding of geometry knowledge. Hands-on operation (relative to observation) increased activation of sensorimotor systems important for representing and processing geometry terms. This activation, in turn, enhanced understanding of the torque and isosceles triangle (as assessed *via* our quiz).

Math ability affects mathematical problem-solving (Krawec, 2014; Artemenko et al., 2019). This study further obtained consistent results that students with high academic level outperformed the students with low academic level in problem-solving tasks. In addition, recent studies have consistently suggested that activation in left DLPFC scales linearly with mental workload (Ayaz et al., 2012; Fishburn et al., 2014), and DLPFC activity also increased during mental effort anticipation of hard tasks as compared to easy tasks (Vassena et al., 2019). In this study, left DLPFC activated significantly higher in the low-achieving students than in the high-achieving students during the learning process and well-structed problem-solving process, no matter in which learning style (both hands-on operation and video observation), indicating that low-achieving students got more mental workload in geometry learning. Left DLPFC activated significantly higher in the observation group than in the operation group in the well-structed problem-solving tasks, which implied that hands-on operation served to off-load the demand of reasoning in the mind on to external objects (Stull et al., 2018), and thus supported the development of more complex understandings and improved geometry problem-solving. Recent psychophysical studies have provided evidence that each modality (visual, auditory, tactual) has its own working memory. If multiple modalities are employed compared to the use of a single modality for a same amount of information, the cognitive capacity increases and thus the cognitive load is reduced (Zacharia et al., 2012).

Hands-on experience can reduce the difference of problem-solving ability formed by academic level. When solving the ill structured problems that come from real life situations and require the participation of advanced cognitive processes, students with low academic level in the operation group outperformed those in the observation group. However, for those with high academic level, there was no significant difference. This result indicated that hands-on experience was more conducive to students with low academic level in geometry problem-solving.

In line with these findings, this would open the way for classroom practices. We can provide concrete manipulatives and meaningful operations in mathematics classroom, especially when students with lower academic level start to learn geometric knowledge. However, this is not to say that geometry reasoning only involves perception, but to recognize that sensorimotor information with the physical world is an integral part of learning, which has been neglected (Giardino, 2018).

Limitations and future directions

The present study is a preliminary attempt to investigate the effect and the neural basis underlying the value of hands-on experience in geometry learning in middle school. We tried to bring educational neuroscience methods to geometry learning, and bridged the gap between neuroscience and mathematics education (Ansari and Lyons, 2016). Focusing on our experiences with NIRS in the school setting, chances and limitations exist in the current study should be noted. First, fNIRS has the potential to be used in the natural environment (Soltanlou et al., 2018), thus future research can be extended to real classrooms. Second, the small sample size and recruitment from one middle school may limit external validity. Future studies with more participants should be conducted. Finally, this study discusses the influence of hands-on experience on geometry learning. Subsequent studies can extend learning contents to algebraic problem solving.

Conclusion

In summary, our findings indicated that hands-on experience improved geometry problems-solving performance for middle-school students. Hands-on operation caused the activation of the primary motor cortex and somatosensory association cortex during the learning process. Somatosensory association cortex was reactivated again during problem solving, which could effectively support geometry reasoning. Students with low academic level were more dependent on sensorimotor experience generated by perceptual and interactive richness of manipulatives.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Nantong University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

LS, LD, and WZ co-designed and co-performed the experiments. LS conducted statistical analyses and wrote the manuscript. LD and DT examined the data and revised the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding

This research was financially supported by grant 21YJC880062 from MOE (Ministry of Education in China) Project of Humanities and Social Sciences Youth Foundation, grant C-b/2020/01/18 from Educational Science Planning Project in Jiangsu Province, and Large Instruments Open Foundation of Nantong University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abrahamson, D., and Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *J. Learn. Sci.* 25, 203–239. doi: 10.1080/10508406.2016.1143370

Ansari, D., and Lyons, I. M. (2016). Cognitive neuroscience and mathematics learning: How far have we come? Where do we need to go? *ZDM* 48, 379–383. doi: 10.1007/s11858-016-0782-z

Artemenko, C., Soltanlou, M., Bieck, S. M., Ehlis, A. C., Dresler, T., and Nuerk, H. C. (2019). Individual differences in math ability determine neurocognitive processing of arithmetic complexity: A combined fNIRS-EEG study. *Front. Hum. Neurosci.* 13:227. doi: 10.3389/fnhum.2019.00227

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023

Bahnmueller, J., Dresler, T., Ehlis, A. C., Cress, U., and Nuerk, H. C. (2014). NIRS in motion-unraveling the neurocognitive underpinnings of embodied numerical cognition. *Front. Psychol.* 5:743. doi: 10.3389/fpsyg.2014.00743

Bellebaum, C., Tettamanti, M., Marchetta, E., Della Rosa, P., Rizzo, G., Daum, I., et al. (2013). Neural representations of unfamiliar objects are modulated by sensorimotor experience. *Cortex* 49, 1110–1125. doi: 10.1016/j.cortex.2012.03.023

Boakes, N. J. (2009). Origami Instruction in the middle school mathematics classroom: Its impact on spatial visualization and geometry knowledge of students. *RMLE Online* 32, 1–12. doi: 10.1080/19404476.2009.11462060

Brown, M. C., McNeil, N. M., and Glenberg, A. M. (2009). Using concreteness in education: Real problems, potential solutions. *Child Devel. Perspect.* 3, 160–164. doi: 10.1111/j.1750-8606.2009.00098.x

Burte, H., Gardony, A. L., Hutton, A., and Taylor, H. A. (2017). Think3d!: Improving mathematics learning through embodied spatial training. *Cogn. Res. Princ. Implicat.* 2, 1–18. doi: 10.1186/s41235-017-0052-9

Carbonneau, K. J., Marley, S. C., and Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *J. Educ. Psychol.* 105, 380–400. doi: 10.1037/a0031084

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *J. Exp. Psychol.*138, 351–367. doi: 10.1037/a0015854

Casasanto, D. (2016). Different bodies, different minds: The body specificity of language and thought. *Curr. Direct. Psychol. Sci.* 20, 378–383. doi: 10.1177/0963721411422058

Chang, T. L., and Hsin, H. T. (2021). The effect of the Self-explain–Discuss–Re-explain (SDR) learning strategy on high- and low-achieving fifth-grade students' achievement in science. *Res. Sci. Technol. Educ.* 39, 461–488. doi: 10.1080/02635143.2020.1772223

Chen, Y., Lian, R., Yang, L., Liu, J., and Meng, Y. (2017). Working memory load and reminder effect on event-based prospective memory of high- and low-achieving students in math. *J. Learn. Disabil.* 50, 602–608. doi: 10.1177/0022219416668322

Chu, M., and Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *J. Exp. Psychol.* 140, 102–116. doi: 10.1037/a0021790

Dupeyrat, C., Escribe, C., Huet, N., and Re, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *Int. J. Educ. Res.* 50, 241–250. doi: 10.1016/j.ijer.2011.08.005

Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8:76. doi: 10.3389/fnhum.2014.00076

Freina, L., Bottino, R., and Ferlino, L. (2018). Visuospatial abilities training with digital games in a primary school. *Int. J. Ser. Games* 5, 23–35. doi: 10.17083/ijsg.v5i3.240

Giardino, V. (2018). Manipulative imagination: How to move things around in mathematics. *Theoria* 33, 345–360. doi: 10.1387/theoria.17871

Hoge, R. D., Franceschini, M. A., Covolan, R. J. M., Huppert, T., Mandeville, J. B., and Boas, D. A. (2005). Simultaneous recording of task-induced changes in blood oxygenation, volume, and flow using diffuse optical imaging and arterial spin-labeling MRI. *NeuroImage* 25, 701–707. doi: 10.1016/j.neuroimage.2004.12.032

Hutmacher, F., and Kuhbandner, C. (2018). Long-term memory for haptically explored objects: Fidelity, durability, incidental encoding, and cross-modal transfer. *Psychol. Sci.* 29, 2031–2038. doi: 10.1177/0956797618803644

Ianì, F. (2019). Embodied memories: Reviewing the role of the body in memory processes. *Psychon. Bull. Rev.* 26, 1747–1766. doi: 10.3758/s13423-019-01674-x

Jin, S. H., Lee, S. H., Yang, S. T., and An, J. (2020). Hemispheric asymmetry in hand preference of right-handers for passive vibrotactile perception: An fNIRS study. *Scientif. Rep.* 10:13423. doi: 10.1038/s41598-020-70496-y

Kahney, D. (1993). *Problem solving: Current issues (2nd ed.)*. Milton Keynes: Open University Press.

Kaminski, J. A., Sloutsky, V. M., and Heckler, A. F. (2009). Supporting online materials for "The advantage of abstract examples in learning math." *Science* 320, 454–455.

Kontra, C., Lyons, D. J., Fischer, S. M., and Beilock, S. L. (2015). Physical Experience Enhances Science Learning. *Psychol. Sci.* 26, 737–749. doi: 10.1177/0956797615569355

Krawec, J. L. (2014). Problem representation and mathematical problem solving of students of varying math ability. *J. Learn. Disabil.* 47, 103–115. doi: 10.1177/0022219412436976

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instruct. Sci.* 26, 49–63. doi: 10.1007/978-94-017-2243-8_5

McNeil, N. M., and Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: Perspectives from development and education. *Child Devel. Perspect.* 3, 137–139. doi: 10.1111/j.1750-8606.2009.00093.x

Ministry of Education of the People's Republic of China (2012). *Compulsory education mathematics curriculum standards*, Vol. 47, 2011th Edition. Beijing: Beijing Normal University Press, 2–3.

Molina, D. R. J., Guevara, M. A., and Hernández, G. M. (2019). EEG correlation during the solving of simple and complex logical-mathematical problems. *Cogn. Affect. Behav. Neurosci.* 19, 1036–1046. doi: 10.3758/s13415-019-00703-5

Montague, M., Krawec, J., Enders, C., and Dietz, S. (2014). The effects of cognitive strategy instruction on math problem solving of middle-school students of varying ability. *J. Educat. Psychol.* 106, 469–481. doi: 10.1037/a0035176

Nathan, M. J., and Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cogn. Res.* 2:9. doi: 10.1186/s41235-016-0040-5

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., and Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychol. Sci.* 25, 903–910. doi: 10.1177/0956797613518351

Novak, M., and Schwan, S. (2021). Does touching real objects affect learning? *Educ. Psychol. Rev.* 33, 637–665. doi: 10.1007/s10648-020-09551-z

Nyberg, L., Petersson, K. M., Nilsson, L. G., Sandblom, J., Åberg, C., and Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage* 14, 521–528. doi: 10.1006/nimg.2001.0801

Okamoto, M., Dan, H., Shimizu, K., Takeo, K., Amita, T., Oda, I., et al. (2004). Multimodal assessment of cortical activation during apple peeling by NIRS and fMRI. *NeuroImage* 21, 1275–1288. doi: 10.1016/j.neuroimage.2003.12.003

Otten, M., van den Heuvel-Panhuizen, M., Veldhuis, M., Boom, J., and Heinze, A. (2020). Are physical experiences with the balance model beneficial for students' algebraic reasoning? An evaluation of two learning environments for linear equations. *Educ. Sci.* 10, 1–23. doi: 10.3390/educsci10060163

Palmiero, M., Piccardi, L., Giancola, M., Nori, R., D'Amico, S., and Olivetti Belardinelli, M. (2019). The format of mental imagery: From a critical review to an integrated embodied representation approach. *Cogn. Proc.* 20, 277–289. doi: 10.1007/s10339-019-00908-z

Pexman, P. M. (2017). The role of embodiment in conceptual development. *Lang. Cogn. Neurosci.* 34, 1274–1283. doi: 10.1080/23273798.2017.1303522

Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., et al. (2015). Using fiberless, wearable fnirs to monitor brain activity in real-world cognitive tasks. *J. Visualiz. Exp.* 106:53336. doi: 10.3791/53336

Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., et al. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *NeuroImage* 85, 64–71. doi: 10.1016/j.neuroimage.2013.06.062

Pouw, W. T. J. L., van Gog, T., and Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educ. Psychol. Rev.* 26, 51–72. doi: 10.1007/s10648-014-9255-5

Quaresima, V., and Ferrari, M. (2019). A mini-review on functional near-infrared spectroscopy (fNIRS): Where do we stand, and where should we go? *Photonics* 6:87. doi: 10.3390/photonics6030087

Rau, M. A., and Herder, T. (2021). Under which conditions are physical versus virtual representations effective? Contrasting conceptual and embodied mechanisms of learning. *J. Educ. Psychol.* 113, 1565–1586. doi: 10.1037/edu0000689

Reed, S. K. (2016). The structure of Ill-structured (and well-structured) problems revisited. *Educ. Psychol. Rev.* 28, 691–716. doi: 10.1007/s10648-015-9343-1

Rivella, C., Cornoldi, C., Caviola, S., and Giofrè, D. (2021). Learning a new geometric concept: The role of working memory and of domain-specific abilities. *Br. J. Educ. Psychol.* 91, 1537–1554. doi: 10.1111/bjep.12434

Scarapicchia, V., Brown, C., Mayo, C., and Gawryluk, J. R. (2017). Functional magnetic resonance imaging and functional near-infrared spectroscopy: Insights from combined recording studies. *Front. Hum. Neurosci.* 11:419. doi: 10.3389/fnhum.2017.00419

Sermier, D. R., and Bless, G. (2013). The impact of including children with intellectual disability in general education classrooms on the academic achievement of their low-, average-, and high-achieving peers. *J. Intellect. Devel. Disabil.* 38, 23–30. doi: 10.3109/13668250.2012.757589

Shadgan, B., Molavi, B., Reid, W. D., Dumont, G., and Macnab, A. J. (2010). Do radio frequencies of medical instruments common in the operating room interfere with near-infrared spectroscopy signals? *Adv. Biomed. Clin. Diag. Syst.* VIII: 755512. doi: 10.1117/12.842712

Singh, A. K., and Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *NeuroImage* 33, 542–549. doi: 10.1016/j.neuroimage.2006.06.047

Smith, C. P. (2018). Body-based activities in secondary geometry: An analysis of learning and viewpoint. *School Sci. Mathemat.* 118, 179–189. doi: 10.1111/ssm.12279

Soltanlou, M., Sitnikova, M. A., Nuerk, H. C., and Dresler, T. (2018). Applications of functional near-infrared spectroscopy (fNIRS) in studying cognitive development: The case of mathematics and language. *Front. Psychol.* 9:277. doi: 10.3389/fpsyg.2018.00277

Stull, A. T., Gainer, M. J., and Hegarty, M. (2018). Learning by enacting: The role of embodiment in chemistry education. *Learn. Instruct.* 55, 80–92. doi: 10.1016/j.learninstruc.2017.09.008

Swan, P., and Marshal, L. (2010). Revisiting mathematics manipulative materials. *Prim. Mathem. Classroom* 15, 13–19.

Tak, S., and Ye, J. C. (2014). Statistical analysis of fNIRS data A comprehensive review. *NeuroImage* 85, 72–91. doi: 10.1016/j.neuroimage.2013.06.016

Tian, F., Lin, Z.-J., and Liu, H. (2013). EasyTopo: A toolbox for rapid diffuse optical topography based on a standard template of brain atlas. *Optical Tomogr. Spectr. Tissue X* 8578:85782J. doi: 10.1117/12.2003907

Vallotton, C., Fusaro, M., Hayden, J., Decker, K., and Gutowski, E. (2015). Give me a hand: Differential effects of gesture type in guiding young children's problem-solving. *Instruct. Sci.* 43, 709–735. doi: 10.1007/s11251-015-9357-6

Vassena, E., Gerrits, R., Demanet, J., Verguts, T., and Siugzdaite, R. (2019). Anticipation of a mentally effortful task recruits dorsolateral prefrontal cortex: An fNIRS validation study. *Neuropsychologia* 123, 106–115. doi: 10.1016/j.neuropsychologia.2018.04.033

Weissgerber, S. C., Grünberg, C., and Neufeld, L. (2022). The interplay of math anxiety and math competence for later performance. *Soc. Psychol. Educ.* 25, 977–1002. doi: 10.1007/s11218-022-09700-y

Wilson, M. (2002). Six views of embodied cognition. *Psychometr. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322

Ye, J. C., Tak, S., Jang, K. E., Jung, J., and Jang, J. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *NeuroImage* 44, 428–447. doi: 10.1016/j.neuroimage.2008.08.036

Zacharia, Z. C., Loizou, E., and Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Res. Q.* 27, 447–457. doi: 10.1016/j.ecresq.2012.02.004

Zacharia, Z. C., and Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learn. Instruct.* 21, 317–331. doi: 10.1016/j.learninstruc.2010.03.001

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 14 March 2023
DOI 10.3389/fpsyg.2023.1110950

Check for updates

OPEN ACCESS

EDITED BY
Pedro Gil-Madrona,
University of Castilla-La Mancha, Spain

REVIEWED BY
Milagros Elena Rodriguez,
Universidad de Oriente, Venezuela
Na Li,
Central China Normal University, China

*CORRESPONDENCE
Hong Zhang
✉ zhanghongredg6@163.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 29 November 2022
ACCEPTED 27 February 2023
PUBLISHED 14 March 2023

CITATION
Zhang H, Ye S and Shi L (2023) Cognitive diagnosis and algorithmic cultural analysis of fourth-grade Yi students' mathematical skills in China: A case study of several primary schools in Puge county, Liangshan prefecture. *Front. Psychol.* 14:1110950.
doi: 10.3389/fpsyg.2023.1110950

Cognitive diagnosis and algorithmic cultural analysis of fourth-grade Yi students' mathematical skills in China: A case study of several primary schools in Puge county, Liangshan prefecture

Hong Zhang*, Shuang Ye and Lingke Shi

School of Mathematical Sciences, Sichuan Normal University, Chengdu, China

The Liangshan Yi Autonomous Prefecture is the largest area in China inhabited by the Yi people, and the original Yi characteristics and culture are well maintained. The Yi also have a high degree of ethnic and cultural intermingling with Tibetans, Han and other ethnic groups. The level of mathematical abilities directly determines the quality of mathematical learning of Yi students. Primary four is the stage of "concrete operations," and is a critical point in the development of mathematical symbolic awareness. In this study, the geographical location of the school and the financial income of the township in which the school is located were used as the basis for sampling, and the DINA model was used to diagnose the mathematical ability of fourth grade students in three rural Yi primary schools in Puge County. The study found that there was individual variability in the mathematical abilities of fourth grade Yi students, with 21 different types of cognitive error patterns identified, the main ones being five. In addition, the state of knowledge of fourth grade Yi students in arithmetic revealed that their overall level of mathematical ability was low, showing a lag, with none of the knowledge attributes of arithmetic being fully mastered. Cultural differences between the Chinese and Yi languages contribute to the difficulties that Yi students have in learning mathematical operations, including differences in understanding the place value system, zero, decimal expressions, and differences in the perception of multiplication and division. The above research can inform the implementation of targeted remediation for teaching and learning.

KEYWORDS

Chinese Yi ethnic group, mathematical ability, cognitive diagnosis, DINA model, algorithmic culture

Introduction

Problem statement

There are 55 ethnic minorities among China's 56 ethnic groups. Yi is one of the ethnic minorities who still use their own language and characters, distributed in Sichuan, Yunnan, Guizhou and other places. Liangshan Yi Autonomous Prefecture is the largest Yi settlement in China, located in the south of Sichuan Province, China, with more than 1.8 million people, which is a key area in the fight against educational poverty. Mathematical operation ability is the most basic ability in mathematics, in accordance with the needs of society and the requirements of educational development, the content and training requirements for the teaching of number and calculation in curriculum standards are constantly being improved in and outside China. The fourth grade of primary school is the transition period from the lower to the upper grades. This is the period when reversible thinking is formed, a sense of conservation of concepts can be developed, and concrete logical reasoning can be made. The fourth grade is also a critical point in the development of mathematical symbolic awareness, as students begin to move from empirical to theoretical thinking (Zhu and Ma, 2018). The level of mathematical operation ability directly determines the quality of Yi students' learning in mathematics. Studies have shown that the overall trend of Chinese students' mathematical ability is weakening (Zhou, 2011), while in Sichuan Province, the primary school students who have difficulties in arithmetic are mainly located in rural areas and Yi ethnic groups (Zhang et al., 2009).

Since ancient times, due to the existence of natural barriers such as mountains, rivers and gullies in Liangshan area, compared with Yunnan and Guizhou, the characteristics and cultural original ecology of Yi people in this area are well maintained. At the same time, Liangshan area is within the "Tibetan-Yi-Corridor," which is the necessary place for the ancient "Southern Silk Road." Yi people have more cultural blending with other ethnic groups, such as Tibetan and Han. Culture has a fundamental impact on mathematics education, and the level of mathematical ability contains the elements of algorithm culture. Therefore, it is necessary to analyze the Yi algorithmic culture based on the cognitive diagnosis of mathematical ability of the fourth grade Yi students in Liangshan Prefecture, Sichuan Province, China.

Literature review

Research on the mathematical ability of ethnic minority students is divided into three main categories. The first concerns using indicators such as correctness and frequency to portray students' arithmetic ability. For example, Zhou et al. (2010) portrayed the school readiness of children in Guangxi Zhuang Autonomous Region (GZAR) using the number frequency index, which differed from that of children of Han ethnicity. Zuo and Tao (1994) used the question correctness index to study the current state of the arithmetic ability of 7 to 8-year-old Han and Dai students. It was found that the language barrier of Dai children made arithmetic more difficult. The second regards cross-cultural comparison by means of difference analysis. For example, Zhou et al. (2005) compared the development of arithmetic ability in 9–12-year-old Wei and Han students and found that Wei children are lagging behind by 3 years. Lu (2004) pointed out that, compared to spatial imagination and reasoning skills, the highest average score for "Wei Wei," "Wei Han," and "Han Han" students was in arithmetic skills. Xia (2001) compared the arithmetic ability of Buyei and Han students and showed that Han students have better arithmetic ability than Buyei, and the more Han students there are, the stronger the students' arithmetic ability. The third concerns analysis of students' types of arithmetic errors. For example, He et al. (2020) found that primary school students in South Xinjiang mainly showed conceptual errors in arithmetic, and the main cause was that they did not understand the algorithm and arithmetic of complex operations, the meaning of the place value system, and the like.

The first two types of research pay more attention to the accuracy and scoring rate of the questions, which are general test scores of the subjects and do not provide more detailed diagnostic information. The third kind of research analyzes the causes of the errors in the algorithm and arithmetic of ethnic minority operations. It can also further analyze the influence of ethnic minority culture on the algorithm.

Research questions

In this study, we take fourth grade rural Yi students in Liangshan Prefecture, China as the object of study, conduct cognitive diagnosis and algorithmic cultural analysis on mathematical ability, provide reference for the implementation of targeted remedial measures in teaching, and advocate the pluralistic interaction of various ethnic cultures, leading to the conscious acceptance of cultural identity and citizenship by ethnic students (Da, 2019). Specifically, the following two research questions guided this study: (a) What is the knowledge state of arithmetic of the fourth grade Yi students in Liangshan Prefecture? (b) What are their mastery of arithmetic attributes and related cultural factors?

Materials and methods

Sample selection

Since the level of economic development has a significant positive correlation with educational input (Cai and Zheng, 2013), the economically underdeveloped Puge County was selected as the study area based on the monthly reports of the prefecture-wide and county and city gross domestic product of Liangshan Prefecture in the past 5 years. Puge County has a registered population of 219,000, of which 189,000 are Yi, accounting for 86.3%. It is a county inhabited by 24 ethnic minorities, mainly Yi, Han, and Hui. The 2017–2021 monthly reports of the statewide GDP of Liangshan Prefecture show that Puge County's GDP is ranked the 14th to 16th among the seventeen counties and cities in the prefecture[1], with

1 Statistics based on data published by the Liangshan Yi Autonomous Prefecture Bureau of Statistics for 2017–2021.

a low level of economic development, insufficient investment in education, and relatively weak teaching resources.

Using a stratified sampling method, taking the geographical location of the school and the economic development of the township where the school is located as the sampling basis. All fourth grade students from three schools in three townships were selected as the sample from 34 primary schools in Puge County[2], with a total of 7 classes and 312 fourth grade students, all of Yi ethnicity. The subjects had learned about addition, subtraction, multiplication, and division of integers and addition and subtraction of decimals before the test, which lasted 40 min. In all, 273 valid data were collected, including 82 from G school, 87 from Y school, and 104 from L school (**Table 1**).

Testing tools

Psychological and educational measures included the following instruments: classical test theory (CTT), item response theory (IRT), and cognitive diagnostic theory (CDT) (Mislevy and Robert, 1982). CTT only provides total test scores, merit rates, pass rates, or level classification specifics, and is concerned with student performance or where students' current level. The main purpose of IRT is to position subjects on a trait (e.g., academic achievement, spatial ability) scale (an instrument that portrays size, how much), that is, to give an ability value, and to select and place them accordingly. They both consider the psychological trait being measured as a "statistical structure" with no clear psychological meaning, and aim to give a holistic assessment of the individual at a macro level, assigning a value that indicates a position on a unidimensional, linear, continuous system of measures (Liu et al., 2006; Tu et al., 2012). CTT and IRT measurement theories focus on the results of test scores, and although they have the advantage of providing a macro-level assessment of students' overall ability levels, they do not provide the underlying knowledge structures behind the scores. Therefore, a cognitive diagnostic approach in a micro sense was used to test the mathematical ability of Yi fourth graders, emphasizing the combination of macro ability level and micro cognitive level assessment, providing the hidden knowledge structure status and cognitive attribute distribution behind the scores through the subjects' responses, and explaining the reasons for the subjects' errors (Nichols and Joldersma, 2008), thus providing targeted remediation instruction.

Development and testing of test instruments

Testing method

The test framework for cognitive diagnosis consists of two parts: the development of the test instrument and the test, and the use of the test instrument and the interpretation of the results (Gierl, 1997). It is necessary to first identify the cognitive attributes of mathematical operation skills and their hierarchical structure (Leighton et al., 2004) to guide the development of the test paper. Since the mathematical skills of fourth graders are an overall representation of the arithmetic knowledge learned in grades 1 to 4, the arithmetic content of primary school mathematics textbooks from grades 1 to 4 was first analyzed, which mainly included addition, subtraction, multiplication, and division of whole numbers and decimals. In-depth discussions with four current primary mathematics teachers resulted in the identification of 10 cognitive attributes of fourth-grade students' mathematical skills (**Table 2**).

After determining the cognitive properties, it was necessary to classify the 10 mathematical operation properties extracted above into their hierarchical structure. The history of mathematical development shows that different categories of operations are formed and developed gradually from simple to complex, concrete to abstract, and low to high level. Therefore, the knowledge and mastery of operations must also be progressively ordered and hierarchical (Wynn, 1992; Von Aster, 2000). Students were asked to elaborate their arithmetic processes in the form of oral reports, and the opinions of current mathematics teachers were sought to delineate the cognitive order of the above 10 arithmetic attributes, and then to verify the rationality of the attribute hierarchy using the hierarchy consistency index (HCI) values in cognitive diagnostic (Cui et al., 2007; Cui and Leighton, 2009). The final hierarchical relationship diagram of the attributes of mathematical operations knowledge for grades 1 to 4 was determined (**Figure 1**).

Development of the test instrument

First, the preparation of the test instrument requires the determination of the reachable matrix R describing the hierarchical structure of attributes of mathematical operation ability, which reflects all relations between cognitive attributes, including direct, indirect, and self-relations. It is a prerequisite for achieving an accurate diagnosis of each attribute. According to the reachable matrix R, with the help of the expansion algorithm, it is possible to obtain the ideal measurement model (Ding et al., 2012), that is, all test question types that satisfy the relationships of the attribute hierarchy.

The reachable matrix R cannot be obtained directly through the attribute hierarchy relationship graph, but it can be obtained with the help of graph theory knowledge. First, the adjacency matrix A corresponding to the mathematical operation knowledge attribute hierarchy relationship graph is given, based on the Boolean algorithm$R = (A + I)^n$, where I is the unit matrix, and then flexCDMS software is used to obtain the reachable matrix R. With the increasing number of n in$R = (A + I)^n$, when the value of R is stable and constant, then R is a reachable matrix (**Table 3**).

Once the reachable matrix R is obtained, 71 test question types that satisfy the structural relationship of the attribute hierarchy can be obtained using the expansion algorithm. When the test question types contain the reachable matrix R, it is possible to accurately diagnose each knowledge attribute to be examined, making the examination of the attributes structurally valid (Ding et al., 2011). Of the 71 question types, only 10 overlapped completely with the R matrix. The initial test prepared 43 questions based on the 10

2 According to the 'Directory of Schools in Puge County' published by the People's Government of Liangshan Yi Autonomous Prefecture in 2022.

TABLE 1 Sample selection.

School	Location	Annual revenue of the township where the school is located (million)	Effective sample size
G	Scenic area	380	82
Y	General township	136	87
L	General township	283.5	104

TABLE 2 Attributes of mathematical ability.

Properties	Elements
A1	In-table multiplication (1-digit by 1-digit)
A2	Multiplying multi-digit numbers by 1-digit integers
A3	Division of whole numbers with 1-digit divisors
A4	Adding whole numbers up to 10,000
A5	Subtraction of whole numbers up to 10,000
A6	Multiplying 2-digit numbers by 2-digit integers
A7	Multiplying 3-digit numbers by 2-digit integers
A8	Dividing whole numbers with 2-digit divisors
A9	Decimal addition
A10	Decimal subtraction

TABLE 3 R matrix under attribute hierarchy relations.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	1	1	1	0	0	1	1	1	0	0
A2	0	1	0	0	0	1	1	0	0	0
A3	0	0	1	0	0	0	0	1	0	0
A4	0	1	0	1	0	1	1	0	1	0
A5	0	0	1	0	1	0	0	1	0	1
A6	0	0	0	0	0	1	1	0	0	0
A7	0	0	0	0	0	0	1	0	0	0
A8	0	0	0	0	0	0	0	1	0	0
A9	0	0	0	0	0	0	0	0	1	0
A10	0	0	0	0	0	0	0	0	0	1

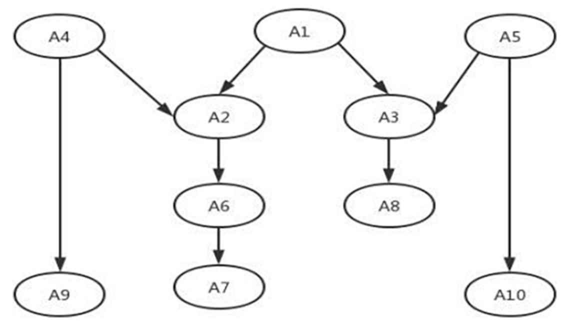


FIGURE 1
Hierarchical relationship of knowledge attributes of mathematical operations.

question types in the R matrix, and after eliminating 7 questions with low discrimination, 36 questions were finally formulated. Typically, test questions should be developed to enable multiple measures of each attribute, so that each attribute is measured at least three times (Tu et al., 2012).

Second, the Q-matrix describing the relationship between the questions and cognitive attributes of the whole paper needs to be obtained (Tatsuoka et al., 1995). The Q matrix corresponding to the 36 questions shows whether the attribute was examined for each question, where 1 is examined and 0 is not examined (Lawrence and DeCarlo, 2012; Feng et al., 2014; **Table 4**). Thus, the Q matrix was obtained for all test questions that satisfy the R matrix. The test questions were selected from the 2014 to 2020 Liangshan Prefecture 4th grade mathematics final exam question bank and prepared in consultation with teaching researchers and current teachers. Two types of questions are included: oral computation and columnar computation. There are 21 oral calculations and 15 vertical calculations. All questions are scored from 0 to 1, with 1 point per question, for a total of 36 points.

Testing of test instruments

Once the test paper is prepared, its quality needs to be checked against the measurement requirements of reliability, validity, completeness of attributes, and discrimination. According to flexCDMs calculations, the cognitive diagnostic reliability of each arithmetic attribute in the test paper is >0.75, showing good individual reliability. In addition, the average reliability is 0.9053, which indicates good overall reliability. For the validity of the test papers, since the development of cognitive diagnostic tests is done on the premise of attributes and attribute hierarchical relationships, the scientific soundness of the test directly depends on the soundness of the hierarchical relationships between attributes. Gierl et al. (2008) suggested that as long as the HCI is higher than 0.70, it indicates that a good cognitive hierarchical model has been constructed; that is, the assumed hierarchical structure of attributes is sound. The HCI of the test paper was calculated according to flexCDMs = $0.7426 > 0.70$, and thus the structural validity was good. In examining the completeness of the 10 selected attributes, a regression analysis was created with question difficulty as the dependent variable and arithmetic attributes as the independent variables, and all cognitive attributes were calculated to explain 82.2% of question difficulty with an effect value of 0.761, all greater than 60%. The knowledge attributes included in the test explained more than 60% of the difficulty, which means that the knowledge attributes selected for the test are complete (Junker and Sijtsma, 2001); thus, the 10 cognitive attributes finally identified in the test paper are more scientific and reasonable, and can reflect the complete knowledge of mathematical operations learned by the

TABLE 4 Q matrix corresponding to 36 questions.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
A1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	1	1	1
A2	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
A3	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	1	1	0
A4	0	1	1	1	0	0	1	1	0	0	1	1	0	0	0	0	0	1
A5	0	0	0	0	1	1	0	0	1	1	0	0	1	1	1	1	1	0
A6	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
A7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
A8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
A9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35	T36
A1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0
A2	1	0	1	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0
A3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
A4	1	1	1	0	1	1	0	0	1	1	1	1	1	1	0	0	0	0
A5	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	1	1	1
A6	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0
A7	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0
A8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
A9	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
A10	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1

fourth-grade students. The results of the analysis of the R-linguistic platform show that the average differentiation value of the 36 test questions is 0.57. In L. Ebel's opinion, according to the criteria that can be proposed to evaluate the merits of the questions according to the discrimination index (Ding, 2001), a discrimination of 0.4 or more indicates a very good question, a discrimination of 0.3 to 0.39 indicates a good question, and a discrimination of 0.2 to 0.29 indicates a fair question. Only one of the 36 test questions has a discrimination below 0.2, indicating that the discrimination of the test items is generally good.

Results

Existing cognitive diagnostic models can be divided into two categories: latent trait models (Fischer, 1973) and potential classification models (Tatsuoka, 1983). The latent trait model aims to analyze what latent traits the subject has, based on the scores obtained from the subject. Latent classification models are designed to classify the subject and help the subject find its place in the group. This test uses the DINA model (Henson, 2005; Wu et al., 2020), which is a latent classification model that allows the data collected from student responses to be processed to analyze the individual's knowledge status and mastery of cognitive attributes behind the scores (de la Torre, 2011). Compared to other potential classification models, the DINA model not only portrays the mental processes of students during arithmetic, but also involves only two parameters, the miss parameter and the guess parameter, which are easy to implement for estimation of these two parameters. Therefore, it is a concise and easy to explain model with the following advantages: both its miss and guess parameters are calculated based on the question level, and the complexity of the model is not affected by the number of attributes. Despite its simplicity, this model has been shown to have a good model fit (de la Torre and Douglas, 2004) and is an excellent model in cognitive diagnosis, which is widely used in practice.

Analysis of state of knowledge

The DINA cognitive diagnostic model for Yi fourth graders' mathematical ability allows us to obtain each student's arithmetic knowledge status (Tatsuoka, 2009). The DINA diagnostic model was used to diagnose each student's arithmetic knowledge, with a "1" indicating that the student mastered the attribute and a "0" indicating that the student did not master the attribute. The diagnosis of the knowledge status of mathematical ability of students in all three schools reflects the existence of individual differences among students (**Table 5**).

Based on the diagnostic results of the DINA model, the subjects' knowledge states were obtained by the great *a posteriori* estimation MAP calculation method, and then matched with the 72 ideal mastery patterns already obtained. It was found that 22 of the subjects' knowledge mastery states were in the ideal mastery pattern (**Table 6**).

In addition, the fourth grade Yi students showed a lag in their arithmetic skills. The subjects with the knowledge status of "0001000000," "0001100000," and "1001100000" were at the 2nd

TABLE 5 A total of 4th-grade Yi students' mathematical skills and cognitive diagnostic knowledge state.

School	Class	Name	Sex	State of knowledge (attribute mastery model)									
				A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Y	4.1	Ji***Xia	Female	1	1	1	1	1	1	1	1	1	1
Y	4.1	Jie**Xia	Male	1	1	1	1	1	1	1	0	1	0
Y	4.2	A***Zuo	Female	1	1	1	1	1	0	1	0	1	1
Y	4.2	E**Xia	Male	1	1	0	1	1	0	1	0	1	0
...													
G	4.1	Ji***Wai	Female	1	1	0	1	1	1	1	0	1	0
G	4.1	Ri** Dao	Male	1	1	1	1	1	1	1	0	1	0
G	4.2	Nai**Xia	Male	1	0	0	1	1	0	0	0	1	0
G	4.2	La***Zha	Female	1	1	0	1	0	1	1	0	1	0
...													
L	4.1	A**Qu	Male	1	1	0	1	1	1	1	0	1	1
L	4.1	La**Cong	Male	1	1	0	1	1	0	1	0	1	0
L	4.2	Bao***Niui	Female	1	1	0	1	0	0	0	0	0	0
L	4.2	Ji***Za	Female	1	1	0	1	1	1	1	0	1	0
L	4.3	A*** Li	Female	1	1	0	1	1	1	1	0	1	1
L	4.3	Shi*** Niu	Female	1	1	1	1	1	1	1	1	1	1
...													

For example, the arithmetic knowledge state of Xie**Xia is "11111111010," while the arithmetic knowledge state of Nai**Xia is "1001100010." Based on the cognitive diagnostic knowledge state table, teachers can grasp each student's strengths and weaknesses in arithmetic and teach them according to their needs. For example, A**Qu's knowledge state is "1101111011," which means that the student has not yet mastered the division of integers by one digit (A3) and the division of integers by two digits (A8); therefore, the teacher needs to strengthen the remedial learning of attributes A3 and A8 for targeted homework. ***Indicates the first name in the student's name.

TABLE 6 Classification of subjects' ideal mastery patterns.

State of knowledge	Frequency	Percentages	State of knowledge	Frequency	Percentages
1101000000	7	2.56%	1001000010	3	1.10%
0001000000	7	2.56%	0001100000	1	0.37%
1101011000	3	1.10%	1001100000	4	1.47%
1101110000	2	0.73%	1101111000	11	4.03%
0001100010	1	0.37%	1001100010	8	2.93%
1101010010	1	0.37%	1111111100	3	1.10%
1101011010	8	2.93%	1111111010	7	2.56%
1111111001	1	0.37%	1101111010	6	2.20%
1111110101	3	1.10%	1111111101	3	1.10%
1111110101	2	0.73%	1101111011	14	5.13%
1111111011	25	9.16%	1111111111	120	43.96%

According to the above classification table, 240 out of 273 subjects whose mastery patterns conformed to the corresponding hierarchical structure relationship in Figure 1 were successfully classified as having ideal mastery patterns, with a high rate of 87.91%. As can be seen from Table 5, except for the full mastery pattern (11111111111), all the other 21 patterns indicated that Yi fourth graders made different types of cognitive errors, mainly "111111011," "1101111011," "1101111000," "1001100010," and "1101011010," which are the five types, constituting 24.18% of the total number and 43.15% of the total cognitive errors.

grade level, while those with the knowledge status of "1101000000" and "1101110000" were at the 3rd grade level. The subjects with the knowledge status of "1101000000" and "1101110000" were at the level of grade 2, and those with the knowledge status of "1101000000" and "1101110000" were at the level of grade 3. A total of 60.08% of the fourth grade Yi students mastered all the arithmetic knowledge (A1∼A6) learned in grades 1–3, and only 43.96% mastered all the arithmetic knowledge (A1–A10) learned in grades 1–4.

Analysis of cognitive attribute mastery

According to the mastery statistics of each arithmetic attribute (Table 7), the overall level of mathematical ability of the fourth-grade students in Liangshan Yi Prefecture is low, and none of the ten arithmetic knowledge attributes has been fully mastered, among which integer addition within 10,000 (A4) is the best mastered, and integer division with 2-digit divisors (A8) is the worst mastered.

Among the operations A1 to A6 studied in grades 1 to 3, in-table multiplication (A1), multiplication of multi-digit numbers

TABLE 7 Cognitive attribute mastery statistics.

Property code	Number of attribute holders	Percentage of attribute mastery
A1	259	94.87%
A2	249	91.21%
A3	185	67.77%
A4	264	96.70%
A5	239	87.55%
A6	209	76.56%
A7	225	82.42%
A8	132	48.35%
A9	214	78.39%
A10	181	66.30%

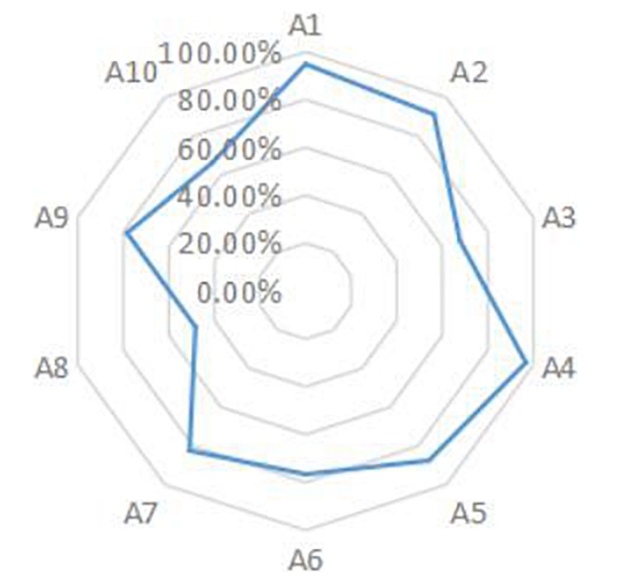


FIGURE 2
Radar chart of student knowledge attribute mastery.

by 1-digit numbers (A2), addition of integers within 10,000 (A4), and subtraction of integers within 10,000 (A5) were relatively well mastered, but these six attributes were not fully mastered by students, indicating that Yi fourth graders had "underachieved" when it comes to the arithmetic knowledge learned in grades 1 to 3. Figure 2 is a visual representation of the radar chart.

The less than satisfactory mastery of the property of arithmetic (A7 to A10) by Yi fourth graders could be attributed to the failure to master the arithmetic knowledge learned in grades 1 to 3 (A1 to A6). Division of integers with 2-digit divisors (A8) studied in grade 4 had the lowest mastery rate of 48.35%, that is, more than half of the students did not master this property. The reason for this is that its prerequisites are mastery of the knowledge attribute in-table multiplication (A1), division by whole numbers with 1-digit divisors (A3), and subtraction of whole numbers up to 10,000 (A5), and as many as 49.81% did not master the prerequisites, resulting in subjects who did not master A1, A3, and A5 also failing to master division by whole numbers with 2-digit divisors (A8). Similarly, students at this grade level had difficulty in decimal operations, with mastery of decimal addition (A9) (78.39%) and decimal subtraction (A10) (66.30%) ranking in the bottom five and bottom two, respectively.

Summary of main results

Cognitive diagnosis through the DINA model leads to the following main outcomes.

First, the state of knowledge of fourth grade Yi students in arithmetic is captured. Students' mastery of ten cognitive attributes is accurately described, and individual differences in the mathematical abilities of students scoring the same are then identified as a means of targeting training and remediation.

Second, all cognitive error patterns and major cognitive error patterns in arithmetic were obtained for fourth grade Yi students. The knowledge states obtained from the diagnosis were matched with 72 ideal mastery patterns to obtain 21 different types of cognitive error patterns, which in turn led to the identification of the main five error patterns, accounting for 43.15% of the total cognitive errors.

Third, the overall mastery of the properties of arithmetic knowledge by fourth grade Yi students was understood. The overall level of mathematical ability of fourth grade Yi students in Liangshan Prefecture was low, and none of the knowledge attributes of arithmetic were fully mastered. A total of 60.08% of the fourth grade Yi students mastered all of the arithmetic knowledge learned in grades 1 to 3, while only 43.96% mastered all of the arithmetic knowledge learned in grades 1 to 4, showing an overall lag in arithmetic skills.

Discussion

Differences in traditional Yi cultural understanding of the place value system, zero, and decimal expressions

From a semiotic point of view, all real mathematical activities are related to the representation of mathematical objects rather than to the mathematical objects themselves (Otte, 2006). The students' relationship to reality is mediated by symbols or symbolic processes. The errors made by the fourth grade Yi students in performing the four operations were related to the representation and understanding of the place value system. The following were the errors in operations: rounding errors in addition, debit errors in subtraction, spatial arrangement errors and rounding errors in multiplication, and failure to consider the divisor as a whole and misplacement of the quotient in division (Figure 3). All reflect barriers to Yi fourth graders' representation and understanding of the place value.

The Yi language differs from Han Chinese notation in that the Yi language does not have a place value system. The "place value system" refers to the fact that the same symbol is assigned a different value depending on its relative position in the numerical representation. The Yi word for "eleven" is written as "++," not a direct spelling of "ten" and "one," which implies the idea of "ten

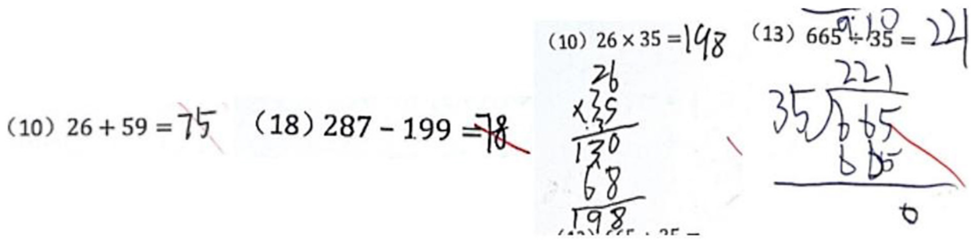


FIGURE 3
Reminding error, debit error, spatial arrangement error, failure to consider divisor as a whole.

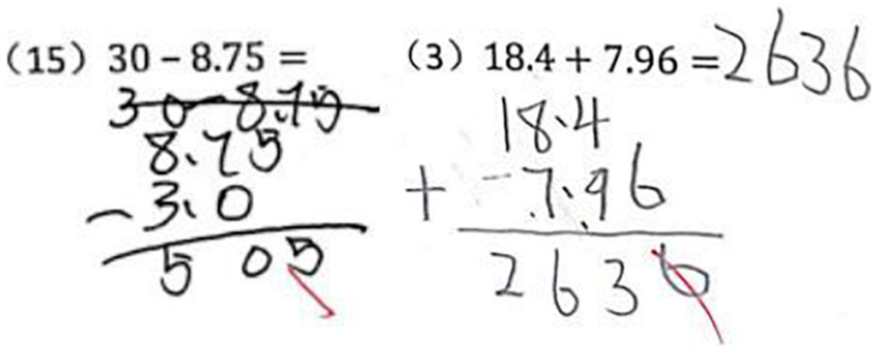


FIGURE 4
Improper understanding of decimal concepts and omission of decimal point.

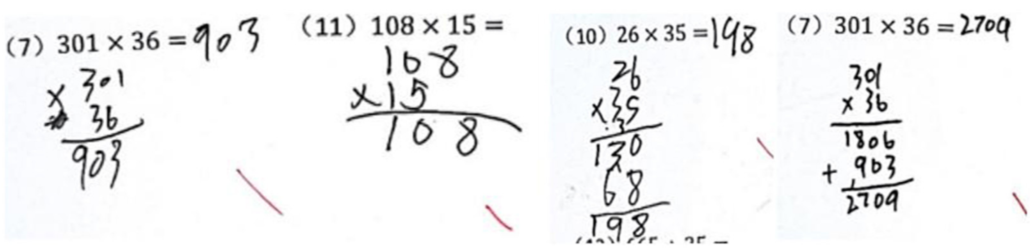


FIGURE 5
High alignment and start-from-high-bit arithmetic error, recede-to-the-right error.

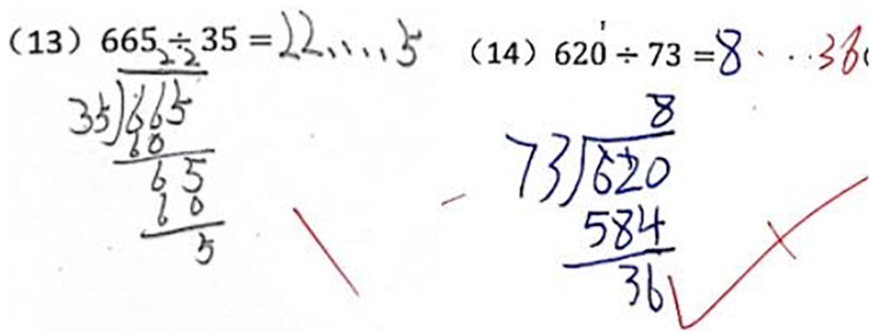


FIGURE 6
Confusing division operations for addition, subtraction and multiplication operations, not estimating quotients correctly, missing remainder errors.

plus a starting number." Similarly, when "twenty" is expressed in Yi, it is not "ㄷ十," but "ㄷ卄," which means "two plus another ten." In addition, the number "11" is written as "十十" when expressed in Yi, but the number "1" is expressed as "ㄴ." The number "20" is written as "ㄷ卄" in Yi, but the number "2" is represented as "ㄷ" (Wang, 2018). The above notation indicates that the Yi algorithm does not have a place value system, which leads to barriers to understanding the concept of place value system for Yi students.

The absence of "zero" and "decimal expressions" in Yi algorithms is also a major obstacle for Yi students to learn decimal operations. Usually, even when people of one ethnicity learn the language of another people, they must translate themselves in their own heads when they learn the culture of that ethnicity (Yi and Ye, 1996). Fourth graders of the Yi ethnic group also translate from Chinese to Yi when they learn mathematical operations and think in Yi. However, the Yi do not have a word for "zero" or an expression for decimals. Interviews with Yi students revealed that they use different units to represent numbers in different digits; for example, 10.09, which is pronounced as "ten point zero nine" in Chinese, is expressed in Yi as "10 yuan and 9 cents," omitting the "zero" and "decimal point." According to the results of the cognitive diagnostic assessment, the probability of mastering decimal addition and subtraction for Yi fourth graders was only 78.39 and 66.30%, respectively, and the students' misunderstanding of decimal concepts and omission of decimal points (**Figure 4**) were related to the lack of "zero" and "decimal expression" in Yi algorithms.

The place value system, zero sign, and decimal expressions in the number system all involve symbolic meaning, and constructing meaning for these symbols requires a complex process of negotiation through discourse (Presmeg, 2006). In teaching arithmetic to Yi students, teachers can create an algorithmic symbol system based on ethnic culture and use a semiotic chain model of instruction to create a series of abstract concepts, while preserving important relationships in students' everyday practice, to help Yi students build a complete cognition of the number system.

Differences in Yi-Han algorithmic culture lead to different perceptions of multiplication and division

Yi fourth graders all had problems with spatial misalignment when multiplying integers: that is, three-digit by two-digit integer multiplication where the multiplier is aligned with the first digit of the multiplied number and only the higher digit of the multiplier is counted in vertical calculations, or two-digit by two-digit integer multiplication (three-digit by two-digit integer multiplication) where the multiplication tens digit calculation is shifted one place to the right (**Figure 5**). In Han multiplication, the multiplier is aligned with the last digit of the multiplied number, and each digit of the multiplied number is multiplied by the last digit of the multiplied number. The product of the second and subsequent multiplications is shifted one digit to the left. However, in Yi multiplication, the multiplier is aligned with the highest digit of the multiplied number, and the digit in the highest digit of the multiplied number is multiplied by each digit of the multiplied number. The product of the second and subsequent multiplications is shifted one digit to the right. The second and subsequent products are shifted one place to the right, which is different but the result is the same (Wu et al., 1996).

Mathematical activities require different symbolic representation systems. This highlights a key issue in mathematical understanding: how learners can identify the same representational objects through symbolic representations produced by different representational systems (Duval, 2006). Yi fourth graders either used the Yi algorithm but incomplete calculations when performing multi-digit multiplication operations or confused the use of the Yi multi-digit multiplication algorithm with the Chinese multi-digit multiplication algorithm. Since the two operations are exactly opposite, this confused the thinking of Yi fourth graders in multi-digit multiplication operations and resulted in different arithmetic errors. Therefore, Yi teachers should focus on explaining to students the differences and connections between the Yi and Chinese algorithmic cultures when teaching multi-digit multiplication operations.

Yi fourth graders had the poorest mastery of division, especially division by whole numbers with two-digit divisors. Typical errors of Yi fourth graders in division operations included confusing division operations with addition, subtraction, and multiplication operations; not estimating quotients correctly; and missing remainders (**Figure 6**). Yi minorities commonly try to multiply and subtract to roughly solve some simple division, and often ignore the remainder when they cannot divide whole (Wei and Qin, 2014). Fourth-grade Yi students are more unfamiliar with division operations than with other operations; therefore, teachers should focus on teaching students the differences between division operations and addition, subtraction, and multiplication operations, and on developing students' estimation skills to lay the foundation for them to be able to estimate quotients correctly. For division with remainders, teachers should explain the origin of the remainder to students to avoid the error of ignoring the remainder.

The history of the development of mathematics shows that the development of symbolic representations is necessary for the development of mathematical thought. Mathematical objects beginning with numbers are not objects that can be directly perceived or observed, and the learning of numbers must be based on a system of representations. The problems and difficulties in minority mathematics pedagogy can be understood to some extent in terms of cultural conflicts (Bishop, 1994). Therefore, it is necessary to study both the acquisition and transmission of culture in the teaching of mathematics, to continuously improve the process of instructional design and implementation, and to effectively improve the quality of teachers' knowledge construction systems.

Conclusion

Native languages have a significant impact on people's ability to process mathematical problems, and languages representing different cultural backgrounds can affect the way people's brains process mathematical information (Tang et al., 2006). There are many differences between Yi and Han in terms of sentence structure. For example, Han Chinese uses subject-verb-object expressions, while Yi uses subject-verb-predicate expressions. Interviews with fourth-grade Yi students revealed that Yi students would read questions in a different order than in Han Chinese, which resulted in their inability to understand the meaning of the question. For example, "3 minus 2" is expressed in Yi as "32 minus," "6 plus 7" is expressed in Yi as "67 plus," and "4 times 5" is expressed in Yi as "45 times." Thus, the cultural differences between Chinese and Yi lead to difficulties for students in learning mathematical operations.

As Yi language mathematics teachers suffer from a lack of ontological knowledge base of mathematics subjects, insufficient multicultural integration skills, weak professional development skills, and ambiguous professional development beliefs, students mostly rely on their personal linguistic translations to understand abstract mathematical knowledge. The generally low Chinese language level of Yi students makes it inevitable that some incorrect translations will occur during the conversion from Chinese-mathematical language-Yi language, resulting in a misunderstanding of mathematical knowledge, theorems, formulas, logic, etc., which reduces the effectiveness of learning (Li, 2012). From the aspects of mathematical subject content compensation, mathematical problem identification in multiple contexts, and ethnicized material mining for teaching design, it is essential to facilitate the professional knowledge development of minority mathematics teachers and to improve the arithmetic ability of minority students (Sun and Zhang, 2020).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Sichuan Normal University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

HZ and SY: conceptualization, methodology, formal analysis, writing—original draft preparation, and several rounds of revision. SY and LS: data collection and editing. HZ and LS: writing—subsequent drafts, reviewing, revising, proofreading, and final draft, and preparation and editing for submission. HZ: supervision and project administration. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the National Natural Science Foundation of China: Historical Evolution of Tibetan Weights and Measures and the Development of Tibetan Algorithmic Culture (12126504), and the Evolution and Transmission of Mathematical Symbols (11471232).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bishop, A. J. (1994). Cultural conflicts in mathematics education: developing a research agenda. *For. Learn. Math.* 14, 15–18.

Cai, L., and Zheng, Y. (2013). Empirical research on the relation between education investment and economic development level in ethnic areas: taking three ethnic autonomous prefectures in Sichuan for example. *Guizhou Ethn. Stud.* 34, 114–118.

Cui, Y., and Leighton, J. P. (2009). The hierarchy consistency index: evaluating person fit for cognitive diagnostic assessment. *J. Educ. Meas.* 46, 429–449. doi: 10.1111/j.1745-3984.2009.00091.x

Cui, Y., Leighton, J. P., Gierl, M. J., and Hunka, S. M. (2007). The hierarchy consistency index: a person-fit statistic for the attribute hierarchy method. *Diss. Abst. Int. Sec. A Human. Soc. Sci.* 68:4193.

Da, W. (2019). Theoretical conception of ethnic education practice under the community consciousness of the Chinese nation. *Ethn. Educ. Res.* 30, 38–42.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika* 76, 179–199. doi: 10.1007/s11336-011-9207-7

de la Torre, J., and Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika* 69, 333–339. doi: 10.1007/BF02295640

Ding, S., Luo, F., and Wang, W. (2012). Extension to tatsuoka's q matrix theory. *Psychol. Exploration* 32, 417–422.

Ding, S., Wang, W., and Yang, S. (2011). The design of cognitive diagnostic test blueprints. *Psychol. Sci.* 34, 258–265.

Ding, X. (2001). *Psychometrics*. Kaifeng: Henan University Press.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educ. Stud. Math.* 6, 103–131. doi: 10.1007/s10649-006-0400-z

Feng, Y., Habing, B. T., and Huebner, A. (2014). Parameter estimation of the reduced RUM using the EM algorithm. *Appl. Psychol. Meas.* 38, 137–150. doi: 10.1177/0146621613502704

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychol.* 37, 359–374. doi: 10.1016/0001-6918(73)90003-6

Gierl, M. J. (1997). Comparing cognitive representations of test developers and students on a mathematics test with Bloom's taxonomy. *J. Educ. Res.* 91, 26–32. doi: 10.1080/00220679709597517

Gierl, M. J., Wang, C., and Zhou, J. (2008). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in algebra on the SAT. *J. Technol. Learn. Assessment* 6, 1–53.

He, W., Dong, L., Fa, X., Shao, W., and Lang, J. (2020). A study of calculation mistakes made by primary students in South Xinjiang–based on Xinjiang large-scale assessment. *J. Math. Educ.* 29:80.

Henson, R. (2005). Test construction for cognitive diagnosis. *Appl. Psychol. Meas.* 29, 262–277. doi: 10.1177/0146621604272623

Junker, B. W., and Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Appl. Psychol. Meas.* 25, 258–272. doi: 10.1177/01466210122032064

Lawrence, T., and DeCarlo. (2012). Recognizing uncertainty in the Q-Matrix via a Bayesian extension of the DINA model. *Appl. Psychol. Meas.* 36, 447–468. doi: 10.1177/0146621612449069

Leighton, J. P., Gierl, M. J., and Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka's rule-space approach. *J. Educ. Meas.* 41, 205–237. doi: 10.1111/j.1745-3984.2004.tb01163.x

Li, J. (2012). *On' Calculation Thinking Differences Between Upper Grades Yi Rural Primary Students:a Case Study of Puge Country, Sichuan Province*. Texas: Southwest University.

Liu, S., Dai, H., and Zhou, J. (2006). Test theory for a new generation of tests–the origin and features of cognitive diagnostic theory. *Psychol. Exploration* 4, 73–77.

Lu, X. (2004). Cross-cultural research of mathematical ability of Uygur and Han middle school students in different regions. *Ethn. Educ. Res.* 5, 46–52.

Mislevy, and Robert, J. (1982). Foundations of a new test theory. *Ets Res. Rep.* 2, i–32. doi: 10.1002/j.2333-8504.1982.tb01336.x

Nichols, P. D., and Joldersma, K. (2008). Cognitive diagnostic assessment for education: theory and applications edited by Leighton, J. P. & Gierl, M. J. J. *Educ. Meas.* 45, 407–411. doi: 10.1111/j.1745-3984.2008.00072.x

Otte, M. (2006). Mathematical epistemology from a Peircean semiotic point of view. *Educ. Stud. Math.* 61, 11–38. doi: 10.1007/s10649-006-0082-6

Presmeg, N. (2006). Semiotics and the "connections" standard: significance of semiotics for teachers of mathematics. *Educ. Stud. Math.* 61, 163–182. doi: 10.1007/s10649-006-3365-z

Sun, L., and Zhang, H. (2020). Empirical research on the development of Yi minority mathematics teachers' professional knowledge and skills from a multicultural perspective: a case of high school teachers who are using the "type- I model" in Liangshan Yi ethnic area as an example. *J. Math. Educ.* 29, 91–95.

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., et al. (2006). Arithmetic processing in the brain shaped by cultures. *Proc. Natl. Acad. Sci. U S A.* 103, 10775–10780. doi: 10.1073/pnas.0604416103

Tatsuoka, K. K. (1983). Rule space: an approach for dealing with misconceptions based on item response theory. *J. Educ. Meas.* 20, 345–354. doi: 10.1111/j.1745-3984.1983.tb00212.x

Tatsuoka, K. K. (2009). *Cognitive Assessment: an Introduction to the Rule Space Method*. New York: Routledge. doi: 10.4324/9780203883372

Tatsuoka, K. K., Nichols, P. D., Chipman, S. F., and Brennan, R. L. (1995). "Architecture of knowledge structures and cognitive diagnosis: a statistical pattern recognition and classification approach," in *Cognitively Diagnostic Assessment*, eds P. D. Nichols, S. F. Chipman and R. L. Brennan (New York, NY: Psychology Press)

Tu, D., Cai, Y., and Ding, S. (2012). *Cognitive Diagnosis: Theory, Methods and Applications*. Beijing: Beijing Normal University Press.

Von Aster, M. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *Eur. Child Adolesc. Psychiatry* 9, S41–S57. doi: 10.1007/s007870070008

Wang, Y. (2018). *The Weights and Measures in the Traditional Culture of the Yi Nationality*. China: Sichuan Normal University.

Wei, D., and Qin, Z. (2014). *General History of Ethnic Cultures in Western China - Science and Technology Volume*. Kunming: Yunnan People's Publishing House.

Wu, H., Liang, X., Yürekli, H., Becker, B. J., Paek, I., and Binici, S. (2020). Exploring the impact of q-matrix specifications through a DINA Model in a large-scale mathematics assessment. *J. Psychoeduc. Assess.* 38, 581–598. doi: 10.1177/0734282919867535

Wu, S., Zhou, G., and Zhou, K. (1996). A preliminary study of Yi mathematics. *J. Southwest Nationalities Coll.* 17, 124–131.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cogn. Psychol.* 24, 220–251. doi: 10.1016/0010-0285(92)90008-P

Xia, X. (2001). Cross-cultural study on mathematical operational ability of junior middle school students' in the areas of Bouyei and the Han. *J. Math. Educ.* 3, 84–87.

Yi, J. K., and Ye, X. (1996). Traditional mathematics of the Liangshan Yi and bilingual education of the Liangshan Yi. *J. Southwest Nationalities Coll.* 17, 132–137.

Zhang, X., He, L., and Zheng, D. (2009). Investigation report on the basic mathematical calculating skills of the sixth-grade students in primary school. *J. Math. Educ.* 18, 42–45.

Zhou, X., Song, B., Chen, X., Jiang, J., Yang, Y., and Xiao, X. (2010). On school readiness of rural Zhuang minority children in Guangxi. *Presch. Educ. Res.* 12, 7–14.

Zhou, Y. (2011). Investigation of middle school students' level of mathematical calculation skill and ability in the new curriculum objectives. *J. Math. Educ.* 20, 24–27.

Zhou, Y., Pan, F., and Tang, J. (2005). Comparison and analysis about children's mathematical ability between Weis and Hans. *J. Guangxi Normal University* 2, 60–63.

Zhu, L., and Ma, Y. (2018). An empirical study on the development level of students' mathematical symbol consciousness in compulsory education. *Curriculum Teaching Material Method* 38, 87–94.

Zuo, M., and Tao, Y. (1994). An experimental study on the development of mathematical arithmetic ability of 7-8 year old children. *Psychol. Sci.* 1, 55–57.

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 28 March 2023
DOI 10.3389/fpsyg.2023.1091126

OPEN ACCESS

EDITED BY
Xinya Liang,
University of Arkansas, United States

REVIEWED BY
Lang Chen,
Santa Clara University, United States
Peijie Jiang,
Hunan Normal University, China

*CORRESPONDENCE
Kan Guo
✉ guokan@bnu.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 06 November 2022
ACCEPTED 06 March 2023
PUBLISHED 28 March 2023

CITATION
Ji Z and Guo K (2023) The association between working memory and mathematical problem solving: A three-level meta-analysis. *Front. Psychol.* 14:1091126.
doi: 10.3389/fpsyg.2023.1091126

The association between working memory and mathematical problem solving: A three-level meta-analysis

Zhongtian Ji and Kan Guo*

School of Mathematical Sciences, Beijing Normal University, Beijing, China

Although working memory (WM) is an important factor in mathematical problem solving (MPS), it remains unclear how well WM relates to MPS. Thus, we aimed to determine this relationship by using a meta-analysis. We searched electronic databases for studies published between 2000 and 2020 and established operational criteria. We conducted Egger's regression tests and created funnel plots to test for publication bias. Finally, a three-level meta-analytic model analysis of data from 130 studies involving 43,938 participants and 1,355 effect sizes revealed a moderate relationship between WM and MPS (r = 0.280, 95% CI = [0.263, 0.314]). Moreover, moderator analyses showed that: (1) dressed-up word problems were more strongly tied to WM than to intra-mathematical problems; (2) the central executive function showed the strongest relation with MPS, whereas the phonological loop had the weakest; (3) gender ratio had significant moderating effects; and (4) some of the above-mentioned significant moderating effects were unique after controlling for other factors. Implications for research and practice were also discussed.

KEYWORDS

working memory, mathematical problem solving, executive functions, word problem, three-level meta-analysis

1. Introduction

Problem solving has, for some time, occupied a prominent position in education research (e.g., Lawson, 2003; Felmer et al., 2016; Priemer et al., 2020). In recent years, incorporating problem solving in school mathematics instruction has become a major area of interest within the field of mathematics education (e.g., Jitendra et al., 2005; Popham et al., 2020). From the perspective of teaching, research on the cognitive level has not been properly transferred to pedagogical issues, and remains separate from practice. Focusing on cognitive training is likely to aid in far-transferring students' performance. However, previous studies have failed to show any stable evidence or provide impetus for teachers' practice. Research on cognitive abilities and mathematical performance thus far has provided ideas for further exploration, and it is possible to identify the predictors for problem solving.

Working memory (WM) is frequently mentioned with regards to cognitive abilities in mathematics. Extensive studies have established that WM is related to students' mathematical performance (e.g., Meyer et al., 2010; Ching, 2017; Wu et al., 2017; Korhonen et al., 2018; Fuchs et al., 2020). In previous studies from Baddeley and Hitch (1974) and Baddeley (1992, 2003, 2010), WM referred to a system to provide temporary storage and manipulation of the information necessary while performing complex tasks. As for mathematical problem solving (MPS), memory systems also appear to be decisive factors (Ambrus and Barczi-Veres, 2016). However, studies have yielded mixed results. For instance, Peng et al. (2016) estimated the average correlation between WM and mathematical word problem solving skills to be 0.37 while Song et al. (2011) found that students with lower WM capacity performed better on medium difficult problems than students with higher WM capacity. Solving mathematical problems appears to be complex as the process involves considerable phases (Pongsakdi et al., 2020). For mathematics, more attention should be given to "knowledge." However, the same "knowledge" might involve different problem solving strategies and cognitive processes. For example, to calculate 8+5, a child who uses a retrieval strategy might solve it by recalling from memory, but a child who uses a decomposition strategy might break it down to 8 + (2 + 3). These are obviously different and the individual differences in memory abilities are related to individual differences in MPS. The relationship between WM and mathematics has been examined by meta-analysis (e.g., Peng et al., 2016). Although the classification divides mathematical skills (e.g., basic number knowledge, whole-number calculations, fractions, Peng et al., 2016), focusing on cognitive processes and MPS, the results are complex. For example, for geometry problem solving tasks from a standardized geometrical achievement test (Mammarella et al., 2012), children were required to calculate the area of complex figures and solve complex geometrical problems. Both calculation skills and geometry knowledge are necessary. Overall, MPS emphasizes students' cognitive processes, providing better future direction.

In MPS, the process comprises several phases that are not necessarily performed sequentially: (1) understanding the problem information and situation; (2) translating problems into a mathematical model; (3) solving the mathematical model with mathematical skills; (4) interpreting and examining results with respect to the problem situation; and (5) communicating the results of the original problems (Pongsakdi et al., 2020). The integrity and overlapping of these processes make them difficult to break down. However, different kinds of problems might cause students varying degrees of cognitive load making it difficult to interpret students' cognitive processes (e.g., Ayres, 2001; Jäder et al., 2017; Voica et al., 2020). In the studies of Blum et al. (2007), mathematical problems were categorized into three main types with varying focuses on cognitive processes: intra-mathematical problems, dressed-up word problems, and modeling problems. Schukajlow et al. (2012) described an intra-mathematical problem as a problem without any connection to the real world. The beginning of the cognitive process follows the mathematical model directly. Problems that appeared the most in class were dressed-up word problems (e.g., Blum, 2015). The cognitive activities involved are more complex than when solving intra-mathematical problems since the mathematical models have been dressed- up according to real-life situations. In solving modeling problems, there always exists a modeling loop and students are required to go back and forth between reality and mathematics (e.g., Herbst, 2019). Taken together, according to students' performance on different problems, it helps to speculate and understand how cognitive factors, including WM, relate to the performance of MPS.

Furthermore, several studies have highlighted that WM relates to mathematics with the strength of these relations differing across components of WM (e.g., Costa et al., 2011; Rennie et al., 2014; Bullen et al., 2020). With their focus on knowledge and skills, previous studies do not offer an adequate explanation for the meaning between those components and the cognitive process during MPS. This article provides valuable insight for understanding the association. For instance, the central executive function might be more vital for dressed-up problems and modeling problems because students must identify what information is useful for solving the problem, plan how to apply what they know comprehensively in real life, and make decisions on how to manage the information. In solving intra-mathematical problems, the phonological loop may play a more important role in phonological awareness and coding in counting (e.g., equations, problems about a sequence of numbers). Therefore understanding the relationship between the three components and MPS is of practical significance, and will provide a fresh angle for educators to re-examine the cognitive processes in MPS. Moreover, researchers have operationalized WM and these components in a variety of ways (e.g., operation span, block span, sentence span). Previous studies have investigated the difference between MPS and WM measures based on reading and counting (Perlow and Jattuso, 2018). Clarifying these problems will also be beneficial for our understanding of the nature of MPS.

To address educators' concerns, it is hoped that this research will contribute to a deeper understanding of students' characteristics (e.g., grade level, gender ratio). For example, because of the changing focus of math instruction (i.e., a heavier focus on counting/calculations in primary school and at young ages, and on complex problems when the student reaches middle or high school), the role of WM may also evolve. In terms of the difficulty of the problems mentioned earlier, even when distinguished in one study, they are not comparable on a wide scale across studies. This research sought to remedy this issue by analyzing MPS by focusing more on cognitive processes rather than on the difficulty of knowledge. It is not certain whether intra-mathematical problems are easier or more difficult than dressed-up problems in different school periods. By exploring problems or cognitive processes, the results will be more general without the limitations of age or other sample types.

In summary, although studies have focused on mathematics, research has yet to systematically investigate MPS. Most studies are limited in that they may be generalizable only to mathematics knowledge, but MPS differs from mathematics in a number of important ways (e.g., understanding, comprehension, and monitoring). The aim of this study was to develop a better understanding of MPS. This paper is structured as follows: We first focus on problems that stress cognitive processes in MPS. We then discuss the components and measures of WM. This is followed by a discussion on the relationship between WM and MPS. We then explore the influence of sample type which is independent of knowledge leading to purer results. Besides, a challenge for this research is that one study may involve two types of mathematical problems or several subsystems of WM and is therefore necessary to extract more than one effect size for the same study. Common methods such as selecting only one effect size per study used in the traditional meta-analysis are unlikely to be appropriate for this study. However, by the three-level meta-analysis, all the useful effect sizes could be extracted and the heterogeneity of within-study variance was calculated to ensure independency (Ran et al., 2022). The three-level meta-analysis model has been proven to be as effective to estimate the parameters in meta-analysis as other traditional random effects approaches, with the additional advantage that multilevel models are more flexible (Van den Noortgate and Onghena, 2003; Ran et al., 2022). For example, multiple predictors can be incorporated into this model (Fernández-Castilla et al., 2020). No previous study has used the three-level meta-analysis model for analyzing the association between WM and MPS. To conclude, we answer the following two questions:

(1) What is the size of the relationship between WM and MPS?
(2) Does the relationship between WM and MPS vary as a function of (a) task type or (b) participant characteristics?

2. Methods

2.1. Data collection

Figure 1 outlines the inclusion, search, and coding procedures. To identify studies for the three-level meta-analysis, we first searched electronic databases (i.e., ERIC, PubMed, Medline, PsycINFO, ProQuest Educational, Scopus, and the China National Knowledge Infrastructure) for studies published between 2000 and 2020. We used problem solving, math,* and working memory in our search, as well as the AND command. We removed the duplicates at first and contacted authors who published studies that we could not find and asked for their papers or unpublished data.

2.2. Operational criteria for inclusion and the elimination of studies

For the target topics included in this study, we first established operational criteria to determine the indicators. We included studies that had considered WM as a whole model or measured one of the WM components (e.g., the central executive function).

Regarding MPS, we considered two kinds of outcomes (intra-mathematical problems and dressed-up word problems). As the included research focused more on young students, there was no research using modeling problems to investigate the relationship. To be considered an intra-mathematical problem, the task had to include arithmetic problems or natural operations with no relationship to reality (e.g., addition and subtraction). To be considered a dressed-up word problem, the task had to include mathematical reasoning, applied math problems, and some necessary problem situations.

Furthermore, if a dissertation was published as an article, we only considered the article itself. After applying these criteria, we identified 130 studies with sample sizes ranging from 20 to 5,234.

2.3. Coding procedures

We created an online coding form. Two researchers coded the following content of the selected studies separately: (a) participant characteristics (e.g., sample size, grade level, country/region, typically developing students vs. students with difficulties); (b) type of MPS outcome (intra-mathematical problem/dressed up word problem); (c) type of WM (unspecified WM/central executive/visuospatial sketchpad/phonological loop); (d) tool for measuring WM (inventory/operation/block/sentence/digit/spot/others).

All effect sizes were then coded. Several studies have reported more than one measure to examine the relationship between WM and MPS. According to the basic rules of the three-level meta-analysis, all relevant effect sizes from each selected study were coded without reducing the number of effect sizes in any way. To ensure the robust reliability of this study, two independent recorders double-coded all the primary studies and checked the data to ensure coding accuracy. The consensus rate (Cohen's kappa) varied between 95 and 100%. Most differences in coding were because of the lack of effective and comprehensive information provided in several studies that described the samples and measures. After revisiting the studies, and discussing the differences, two independent recorders reached an agreement (see **Supplementary Appendix A**).

2.4. Moderator variables

In each study, both groups of important moderators were coded that might explain the significance of the residual within- and between-study variance.

2.4.1. Task type

We classified the WM tasks into the central executive, the visuospatial sketchpad, the phonological loop, and unspecified WM, which were unspecified in primary studies. We coded the measurement of WM to determine whether the overall effect size varied across tools. As highlighted earlier, the classification of WM tests was determined on the basis of the surface-level elements of test content. (e.g., inventory, operation, block, sentence, digit, and spot). We categorized MPS tasks into intra-mathematical and dressed-up word problems.

2.4.2. Participant characteristics

We included four participant characteristics in this study. First, we coded gender based on the ratio of males included in the samples. Second, we categorized the grade levels as follows: elementary, middle, and high school. Third, we coded the cultural environment as a category variable (1 = Eastern, 0 = Western) according to the students' country or region reported in the study (Ran et al., 2022). Finally, we coded whether the sample included students with difficulties (e.g., math difficulties and dyslexia).

2.5. Statistical analysis

We used the metafor package for the R statistical program (Viechtbauer, 2010) and the three-level meta-analytic model

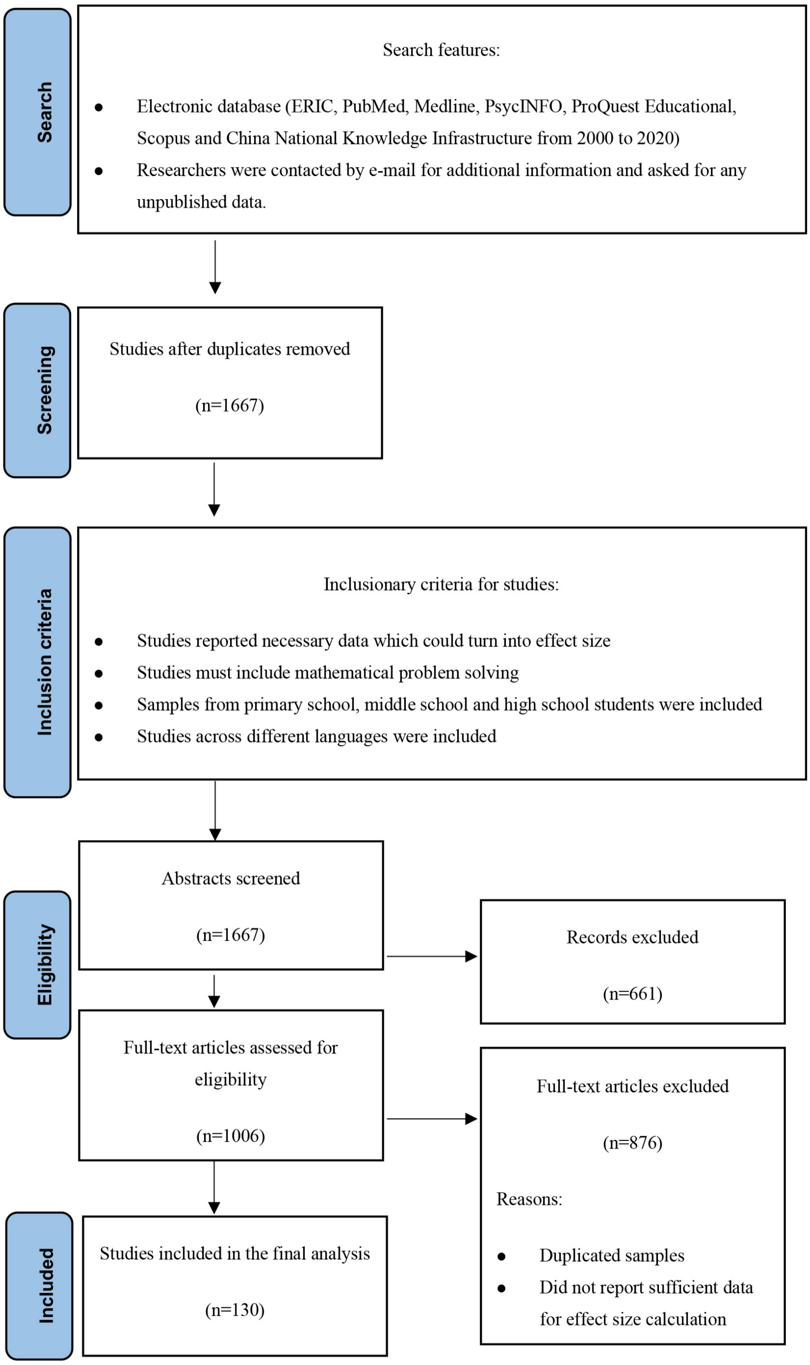


FIGURE 1
Flow diagram for search and inclusion on studies.

tutorial (Assink and Wibbelink, 2016) for the analysis. Since three-level models assume a normal distribution of effect sizes (Van den Noortgate et al., 2013), it is necessary to transform all data into Fisher's Z-values. We applied the Fisher Z-transformation first to conduct the meta-analysis, and then Fisher's z-values were converted back, respectively into correlation coefficients for interpretability (Hedges and Olkin, 2014; Card, 2015). Pearson's r value was not provided in almost 20 of the 130 studies. However, using alternative formulas from, among other sources, Lenhard and Lenhard (2016), we were able to compute the r correlation and then transformed it into Fisher's Z-score. For instance, we used a procedure for converting standardized β to r and then r could be used directly as an effect size. We were also able to compute Cohen's d value using the statistic about means, standard deviations and sample sizes from the treatment and control group, and the t-statistic from the group test and then transformed d into the r-value (Borenstein et al., 2009). As noted in the introduction, we used a three-level random effects model and included all effect sizes in the same study. The three-level meta-analytic model considers three variance components distributed across the model's three

10.3389/fpsyg.2023.1091126

levels: sampling variance of the extracted effect sizes at Level 1, variance between effect sizes from the same study at Level 2, and variance between studies at Level 3 (Assink and Wibbelink, 2016).

To establish whether the variation in the *r*-value between the studies was significant, we used the *Q* test of homogeneity (Hedges and Olkin, 2014). We also computed the 95% CI for each overall effect size to provide more information regarding the correlation. We calculated the variance at Level 1 according to Cheung's (2014) formula and applied the log-likelihood-ratio test to examine heterogeneity at levels 2 and 3. Furthermore, we tested for significance and calculated the distribution of the overall variance.

We also explored moderator variables as potential sources of additional variance in the effect size. We used linear models to predict the study's outcome from the moderator variables, both for the continuous (i.e., gender ratio) and categorical (i.e., school level, task type, and sample characteristics) moderators. Universal classifications were chosen as the reference category to clarify the findings between different task types (e.g., unspecified working memory in WM tasks, both in MPS tasks and others in WM tests). Furthermore, we used a multiple moderator model to scrutinize the unique effect of significant moderators in the univariate analyses and added all significant moderators to the model. We tested the degree of difference between the subsets of studies using a *Q* test and by comparing the correlation magnitude with CIs between the study subsets. Similarly, we investigated the variances at levels 2 and 3.

2.6. Publication bias

To test for publication bias, we first conducted Egger's regression tests (Egger et al., 1997) to test the relationship between the size of the effects from each study and the associated standard error (Georgiou et al., 2020; Ran et al., 2022). If the results of the linear regression showed no significant difference, there was no publication bias. Furthermore, we created funnel plots to test for publication bias. In the funnel plot, the standard error was plotted on the y-axis and the effect size on the x-axis, and if publication bias exists, the funnel would not be symmetric (Ludwig et al., 2019; Borenstein et al., 2021).

3. Results

3.1. Study features

Of the 130 publications included in our final analysis, 24 reported results on all three components of WM, and 55 reported results on both MPS outcomes. There were 43,938 students represented, with sample sizes ranging from 20 to 5,234. Moreover, one study exclusively focused on female students. The number of effect sizes in each study ranged from 1 to 36.

3.2. Meta-analytic results

The three-level meta-analytic model demonstrated that the overall mean correlations between WM and MPS were significant (r = 0.280, $p < 0.001$, 95% CI = [0.263, 0.314]). Additionally, the log-likelihood-ratio test showed significant heterogeneity ($p < 0.001$) at the within-study variance (Level 2) and the between-study variance. Exactly 16.34% of the total variance could be attributed to variance at Level 1, 26.05% of the total variance could be attributed to the differences between the effect sizes within studies at Level 2, and 57.61% could be attributed to the between-study variance (Level 3).

3.3. Results of the moderator analyses

First, we delved into the role of the three components of WM, two types of MPS tasks, and the measured elements (WM) in the relationship of interest. As outlined in **Table 1**, they were all significant moderators. Studies testing the central executive produced significantly larger correlations than those that tested the other two components (0.303 > 0.265 > 0.248, $p < 0.001$). Compared with other tasks, studies using intra-mathematical problems only generated a significantly smaller correlation than those using dressed-up word problems (0.309 > 0.259, $p < 0.001$). Besides, measuring WM by operation showed a larger relation than any other WM tests ($p < 0.001$).

Second, we further analyzed the moderating effects of participant characteristics, including gender, school level, culture, and sample characteristics. As seen in **Table 2**, the gender ratio was a significant moderator. The regression coefficient was positive (β = 0.016), implying that this association was stronger in boys. Besides, the correlation between WM and MPS was stable across school level, cultural background, and unfolding situation.

Previous studies have demonstrated that moderators might be interrelated (Hox et al., 2017). Therefore, we added all the significant moderators to the multiple moderator model to examine what effects were really relevant. As mentioned earlier, we chose universal classifications as the reference category. The omnibus test showed significant results, $F(12,973) = 7.676$, $p < 0.001$, suggesting that at least one of the regression coefficients of the moderators significantly deviated from zero. Based on the findings in **Table 3**, we were able to assert that the components of WM, MPS tasks and measured elements for WM were not confounded by the gender ratio. These results indicated these three moderators had a uniquely moderating effect on the association.

3.4. Publication bias

The results of Egger's test suggested that publication bias should be ignored in the meta-analysis because the *p*-value of this test exceeded 0.05. The symmetric distribution of the funnel plot was depicted in **Figure 2**, which indicated that the results of our meta-analysis were stable and reliable.

4. Discussion

The purpose of this meta-analysis was to estimate the size of the relationship between WM and MPS and to determine if different factors (task type/participant characteristics) moderate

TABLE 1 Relation between working memory (WM) and task types.

Moderator variable	*k*	Intercept/mean z (95% CI)	β (95% CI)	Mean *r*	*F*(df1, df2)	*p*-value	Level 2 variance	Level 3 variance
a. Variable								
Unspecified working memory (RC)	465	0.301*** (0.271, 0.332)		0.292	*F*(3,1353) = 7.384	<0.001***	0.008***	0.018***
Central executive	341	0.313*** (0.278, 0.347)	0.011 (−0.024, 0.047)	0.303				
Phonological loop	303	0.253*** (0.222, 0.285)	−0.048** (−0.079, −0.016)	0.248				
Visuo-spatial sketchpad	248	0.271*** (0.238, 0.303)	−0.031 (−0.063, 0.002)	0.265				
b. Mathematical problem solving tasks								
Both (RC)	70	0.343*** (0.290, 0.396)		0.330	*F*(2,277) = 17.360	<0.001***	0.008***	0.020***
Intra-mathematical problem	912	0.265*** (0.237, 0.293)	−0.078** (−0.127, −0.029)	0.259				
Dressed-up word problem	297	0.319*** (0.289, 0.350)	−0.024 (−0.075, 0.027)	0.309				
c. Measured element (WM)								
Others (RC)	608	0.285*** (0.256, 0.313)		0.278	*F*(6,274) = 4.335	<0.001***	0.009***	0.019***
Operation	22	0.410*** (0.329, 0.491)	0.125** (0.045, 0.205)	0.388				
Block	175	0.280*** (0.245, 0.315)	−0.005 (−0.034, 0.024)	0.273				
Sentence	97	0.298*** (0.259, 0.337)	0.014 (−0.020, 0.047)	0.289				
Digit	277	0.274*** (0.243, 0.305)	−0.010 (−0.035, 0.014)	0.267				
Spot	129	0.342*** (0.304, 0.380)	0.058*** (0.025, 0.090)	0.329				
Inventory	3	0.250 (−0.057, 0.557)	−0.035 (−0.343, 0.274)	0.245				

k = numbers of correlations; mean *z* = mean effect size (Fisher's *z*); β = estimated regression coefficient; *r* = correlation size (Pearson's *r*) for studies belonging to different categories of the moderator variable; Level 2 variance = variance between effect sizes extracted from the same study; Level 3 variance = variance between studies; RC = reference category. $^{**}p < 0.01$; $^{***}p < 0.001$.

TABLE 2 Relation between working memory (WM) and participant characteristics.

Moderator variable	*k*	Intercept/mean z (95% CI)	β (95% CI)	Mean *r*	*F*(df1, df2)	*p*-value	Level 2 variance	Level 3 variance
a. Gender	1063	0.260*** (0.232, 0.288)	0.016*** (0.009, 0.024)	0.254	*F*(1,1061) = 16.639	<0.001***	0.007***	0.015***
b. School level								
Primary school (RC)	1188	0.280*** (0.255, 306)		0.273	*F*(2,1236) = 0.533	0.587	0.009***	0.015***
Middle school	36	0.281*** (0.144, 0.418)	0.001 (−0.138, 0.140)	0.274				
High school	25	0.364*** (0.207, 0.521)	0.084 (−0.075, 0.243)	0.349				
c. Culture								
Western (RC)	1066	0.288*** (0.258, 0.317)		0.280	*F*(1,1234) = 0.023	0.881	0.009***	0.017***
Eastern	170	0.283*** (0.225, 0.341)	−0.005 (−0.070, 0.060)	0.276				
d. Sample characteristics								
Typically-developing students (RC)	1050	0.286*** (0.258, 0.313)		0.278	*F*(1,1355) = 0.348	0.556	0.008***	0.018***
Partly children with difficulties	87	0.300*** (0.253, 0.348)	0.015 (−0.034, 0.063)	0.291				

k = numbers of correlations; mean *z* = mean effect size (Fisher's z); β = estimated regression coefficient; *r* = correlation size (Pearson's *r*) for studies belonging to different categories of the moderator variable; Level 2 variance = variance between effect sizes extracted from the same study; Level 3 variance = variance between studies; RC = reference category. $^{***}p < 0.001$.

their relationship. When we considered any correlation between any components of WM, we found significant links between WM and MPS (the average correlation was 0.280), which are similar to those reported in previous meta-analyses (Peng et al., 2016). Additionally, this relationship was significantly influenced by publication characteristics and task type, but not by participant characteristics. Concretely speaking, in terms of WM, all components of WM demonstrated significant ties with MPS, and the central executive showed the strongest relationship ($r = 0.303$). Regarding the WM tests, operation span had a strong relationship

TABLE 3 Multiple moderator model on the relation between working memory (WM) and mathematical problem solving (MPS).

Moderator variable			*k*	β (95% CI)	
	Intercept			0.428(0.332, 0.524)***	
a. Gender ratio				−0.057(−0.125, 0.010)	
b. Components of working memory	Central executive		243	−0.003(−0.045, 0.038)	
	Phonological loop		203	−0.020(−0.062, 0.022)	
	Visuo-spatial sketchpad		214	−0.061(−0.119, −0.003)*	
c. Mathematical problem solving tasks	Intra-mathematical problem		691	−0.050(0.015, 0.084)**	
	Dressed up word problem		242	−0.026(−0.125, 0.073)	
c. Measured element (WM)	Operation		13	0.102(0.012, 0.191)*	
	Block		147	−0.033(−0.072, 0.005)	
	Sentence		91	0.006(−0.027, 0.039)	
	Digit		200	−0.029(−0.057, −0.002)*	
	Spot		97	0.025(−0.011, 0.061)	
	Inventory		3	−0.064(−0.331, 0.203)	
Multiple moderator model	$k = 986$	$F(12,973) = 7.676$	$p < 0.001$	Level 2 0.007***	Level 3 0.013***

k = numbers of correlations; β = estimated regression coefficient; Level 2 = variance between effect sizes extracted from the same study; Level 3 = variance between studies. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

with MPS. In the field of outcomes for MPS, the connection between WM and dressed-up word problems was stronger. Furthermore, although gender ratio had significance, the results of the multiple moderator model indicated that it was not as steady as expected.

In MPS, diverse strategies and cognitive processes were identified as different types of math problems (Star and Rittle-Johnson, 2008; Hu et al., 2017; Rott et al., 2021). Since different types of problems require different cognitive abilities, we investigated this by examining the relationship between WM and intra-mathematical or dressed-up word problems (Rellensmann and Schukajlow, 2017; Krawitz and Schukajlow, 2018). In this meta-analysis, intra-mathematical problems appeared to be purely mathematical tasks and could not lead to any reality-related mental activities. In contrast, dressed-up word problems are more common in daily life and may require more cognitive resources. Students may have to identify missing or useful information, "undress" the problems, and experience the process of mathematization.

In this meta-analysis, we found that two kinds of mathematical problems were both positively related to WM and the types of MPS indeed moderated the relationship. Furthermore, dressed-up word problems showed stronger links than intra-mathematical problems (r = 0.309 for dressed-up word problems). Additionally, the results of the multiple moderator model indicated similar results. Taken together, these findings imply that cognitive processes drive the relationship between WM and MPS, and highlight their important roles. Understanding the problem situations and translating them into a mathematical model might draw upon significant WM resources. Currently, problem solving is no longer thought of as solving pure mathematical problems (Holmes et al., 2017; Priemer et al., 2020). Recent developments in science, technology, engineering, and mathematics (STEM) and project-based learning place a strong emphasis on 21st-century skills, such as solving problems in reality (Markham et al., 2003; Chen and Yang, 2019; Priemer et al., 2020). As mathematical educators, we also expect students to apply math to real-life scenarios. Students are required to practice solving more reality-related problems, such as ill-structured problems (e.g., Jäder et al., 2017), rather than intra-mathematical problems, thus they also need more regulation of cognition and experience more complex processes. Based on such a trend, WM may play a larger role in identifying valuable information, organizing, and monitoring total performance, and rethinking outcomes in real life.

Regarding WM, we found that the relationship between MPS and WM is indeed affected by components of WM, and the central executive function indicated the strongest relationship with MPS, whereas the phonological loop had the weakest relationship (r = 0.303 for the central executive, r = 0.248 for the phonological loop, and r = 0.265 for the visuospatial sketchpad). Given the variety of WM tests, we also found that each WM test-MPS relationship was well documented and the tests measuring operations showed stronger links than other types. Previous studies have demonstrated that WM tests measure both cognitive abilities and other skills (e.g., counting numbers) according to how the construct was conducted and assessors always use the tool relating best to their criterion of interest (Perlow and Jattuso, 2018). However, the results show that practitioners should pay more attention to cognitive characteristics or relationship between WM tests and other constructs. Besides, several cognitive processes of the executive system might influence WM in MPS, such as controlling, encoding, and retrieval strategies, and suppressing unnecessary information (Miyake et al., 2000; Oberauer et al., 2003). Multiple moderator model analyses also proved that the central executive function showed a stronger relationship than the other two components. Taken together, students need to better master how to organize the entire MPS process. To promote students' performance in MPS, training the central executive function would be a powerful strategy. As mentioned earlier, because problem

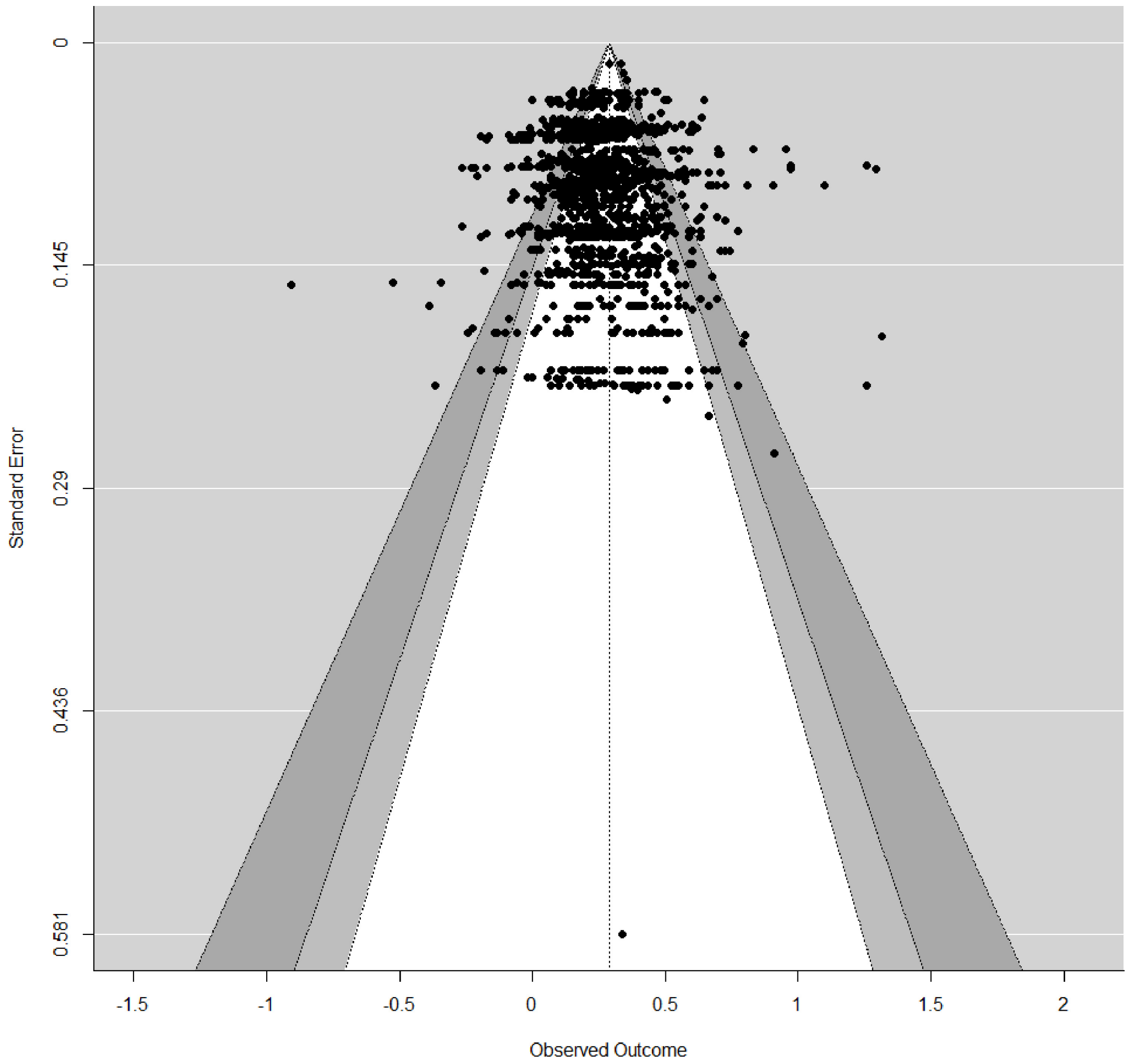


FIGURE 2
Funnel plot of the overall mean r analysis.

solving integrates information from different branches of math, students may need to have a solid command of switching strategies flexibly, and the relationship between WM and MPS should be invariant.

Although gender ratio showed significance, its moderating effects were not robust after controlling for other factors. School level and other participant characteristics did not moderate the relationship between WM and MPS either. Rooted in the idea that cognitive universals exist in this relationship, we claimed that the contributions made here have wide applicability. Additionally, although the school level had a non-significant moderation effect, students at higher school levels showed a larger relationship (0.349 > 0.274 > 0.273). Young students always rely on some basic problem solving strategies such as finger counting to solve mathematical problems at the beginning of formal schooling (Ramirez et al., 2016; Palmer and van Bommel, 2020). By practicing these strategies, they develop strong problem–answer associations. Thus, when they grow up and use more advanced problem solving strategies, they rely heavily on memory-based processes (Siegler and Shrager, 1984; Laski et al., 2013). Another possible reason is due to the different developing patterns among the three components of WM. Studies focusing on working memory have found that the central executive matures later (Palmer, 2000; Muñoz-Pradas et al., 2021), which provides a plausible explanation for our finding. However, the current research still pays insufficient attention to the relationship between WM and MPS in high school students. We recommend additional research to focus on the relationship in senior grades, and apply lab-based findings to actual situations (Cui and Guo, 2022). Clearly, our meta-analysis synthesized research from multiple sources and obtained relatively more reliable conclusions than a single study, thus to some extent making up for the current deficiencies. A greater focus on that could produce interesting findings that develop a deeper understanding of the relationships between WM and MPS.

Our study has some limitations. Although we searched for unpublished papers, this might be a problem for the current study, with a bias toward significant effects. Thus, although funnel plot analyses and Egger's test confirmed that publication bias was probably not a major problem in this study, we may have missed studies that reported non-significant results. Besides, we acknowledge that some of the categories in the moderator analyses did not include many studies. For instance, when investigating the role of school level/grade in the relationship, we had only three students in high school and four in middle school. This may have influenced the chances of finding significant differences. Third, we did not examine the relationship between problem solving and WM. The reason for this circumstance is that primary studies rarely report on the relationship between WM and each process, such as identifying information. Hence, we strongly recommend that future empirical research about this relationship pay more attention to the specific process of MPS. This would allow researchers to further explain how WM is related to MPS. Finally, we did not control for the role of instruction in the relationship between WM and MPS. Previous studies indicate that students' MPS competency can be improved through training and practice (Witt, 2011). Different forms of instruction can alter the cognitive processes involved in specific problems.

In summary, the present meta-analysis applied a three-level, meta-analytic model to quantitatively synthesize the overall association between WM and MPS. The manuscript, therefore, adds to a growing body of research on the role of WM (e.g., Reber and Kotovsky, 1997; Justicia-Galiano et al., 2017). Therefore, all evidence supports the significantly positive correlation between WM and MPS, suggesting that there are benefits if we develop students' WM abilities, which are linked to mathematical performance. Subsequent moderator analyses demonstrated some significant moderators that could explain differences in the strength of the relationship, namely publication characteristics and the task type of WM, as well as MPS. These results have direct implications for instruction and interventions in programming. However, this meta-analysis also underscores areas for future research, including processes of MPS and specific populations [e.g., students with math-related disabilities, which may have significant benefits in terms of mathematical cognition (Geary et al., 2005)].

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

KG collected and organized the data and critically reviewed and revised the manuscript. ZJ performed the statistical analysis. KG and ZJ wrote the original draft. Both authors contributed to the manuscript revision and approved the submitted version.

Funding

This study was supported by a grant from the National Social Science Fund of China (CHA180266).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1091126/full#supplementary-material

References

Ambrus, A., and Barczi-Veres, K. (2016). "Teaching mathematical problem solving in Hungary for students who have average ability in mathematics," in *Posing and Solving Mathematical Problems. Research in Mathematics Education*, eds P. Felmer, E. Pehkonen, and J. Kilpatrick (Cham: Springer), 137–156. doi: 10.1007/978-3-319-28023-3_9

*Anderisson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Appl. Cognit. Psychol.* 21, 1201–1216. doi: 10.1002/acp.1317

*Andersson, U. (2008a). Mathematical competencies in children with different types of learning difficulties. *J. Educ. Psychol.* 100:48. doi: 10.1037/0022-0663.100.1.48

*Andersson, U. (2008b). Working memory as a predictor of written arithmetical skills in children: the importance of central executive functions. *Br. J. Educ. Psychol.* 78, 181–203. doi: 10.1348/000709907X209854

*Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: findings from a 3-year longitudinal study of children with different types of learning difficulties. *J. Educ. Psychol.* 102, 115–134. doi: 10.1037/a0016838

*Andersson, U., and Lyxell, B. (2007). Working memory deficit in children with mathematical difficulties: a general or specific deficit? *J. Exp. Child Psychol.* 96, 197–228. doi: 10.1016/j.jecp.2006.10.001

*Aran Filippetti, V., and Cristina Richaud, M. (2017). A structural equation modeling of executive functions, IQ and mathematical skills in primary students: differential effects on number production, mental calculus and arithmetical problems. *Child Neuropsychol.* 23, 864–888. doi: 10.1080/09297049.2016.1199665

*Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S., Swigart, A. G., and Menon, V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia* 51, 2305–2317. doi: 10.1016/j.neuropsychologia.2013.06.031

Assink, M., and Wibbelink, C. J. (2016). Fitting three-level meta-analytic models in R: a step-by-step tutorial. *Quant. Methods Psychol.* 12, 154–174. doi: 10.20982/tqmp.12.3.p154

*Attout, L., and Majerus, S. (2015). Working memory deficits in developmental dyscalculia: the importance of serial order. *Child Neuropsychol.* 21, 432–450. doi: 10.1080/09297049.2014.922170

*Attout, L., Noel, M.-P., and Majerus, S. (2014). The relationship between working memory for serial order and numerical development: a longitudinal study. *Dev. Psychol.* 50, 1667–1679. doi: 10.1037/a0036496

Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemp. Educ. Psychol.* 26, 227–248. doi: 10.1006/ceps.2000.1051

Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

Baddeley, A. (2010). Working memory. *Curr. Biol.* 20, 136–140. doi: 10.1016/j.cub.2009.12.014

Baddeley, A. D., and Hitch, G. (1974). "Working memory," in *Psychology of Learning and Motivation*, Vol. 8, ed. K. Federmeier (Cambridge, MA: Academic press), 47–89. doi: 10.1016/S0079-7421(08)60452-1

*Bahar, A., and Maker, C. J. (2015). Cognitive backgrounds of problem solving: a comparison of open-ended vs. closed mathematics problems. *Eurasia J. Math. Sci. Technol. Educ.* 11, 1531–1546. doi: 10.12973/eurasia.2015.1410a

*Bailey, D. H., Littlefield, A., and Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: individual and sex differences from first to sixth grades. *J. Exp. Child Psychol.* 113, 78–92. doi: 10.1016/j.jecp.2012.04.014

*Barnes, M. A., Raghubar, K. P., English, L., Williams, J. M., Taylor, H., and Landry, S. (2014). Longitudinal mediators of achievement in mathematics and reading in typical and atypical development. *J. Exp. Child Psychol.* 119, 1–16. doi: 10.1016/j.jecp.2013.09.006

*Berg, D. H. (2008). Working memory and arithmetic calculation in children: the contributory roles of processing speed, short-term memory, and reading. *J. Exp. Child Psychol.* 99, 288–308. doi: 10.1016/j.jecp.2007.12.002

*Bizzaro, M., Giofre, D., Girelli, L., and Cornoldi, C. (2018). Arithmetic, working memory, and visuospatial imagery abilities in children with poor geometric learning. *Learn. Individ. Differ.* 62, 79–88. doi: 10.1016/j.lindif.2018.01.013

Blum, W. (2015). "Quality teaching of mathematical modelling: what do we know, what can we do?," in *The Proceedings of the 12th international congress on mathematical education*, (Cham: Springer), 73–96. doi: 10.1007/978-3-319-12688-3_9

Blum, W., Galbraith, P. L., Henn, H. W., and Niss, M. (2007). *Modelling and Applications in Mathematics Education: The 14th ICMI Study*. Boston, MA: Springer. doi: 10.1007/978-0-387-29822-1

Borenstein, M., Cooper, H., Hedges, L., and Valentine, J. (2009). Effect sizes for continuous data. *Handb. Res. Synth. Meta Anal.* 2, 221–235.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2021). *Introduction to Meta-Analysis*. New York, NY: John Wiley & Sons. doi: 10.1002/9781119558378

*Bresgi, L., Alexander, D. L. M., and Seabi, J. (2017). The predictive relationships between working memory skills within the spatial and verbal domains and mathematical performance of Grade 2 South African learners. *Int. J. Educ. Res.* 81, 1–10. doi: 10.1016/j.ijer.2016.10.004

*Bullen, J. C., Swain Lerro, L., Zajic, M., McIntyre, N., and Mundy, P. (2020). A developmental study of mathematics in children with autism spectrum disorder, symptoms of attention deficit hyperactivity disorder, or typical development. *J. Autism Dev. Disord.* 50, 4463–4476. doi: 10.1007/s10803-020-04500-9

*Cai, D., Georgiou, G. K., Wen, M., and Das, J. P. (2016). The role of planning in different mathematical skills. *J. Cognit. Psychol.* 28, 234–241. doi: 10.1080/20445911.2015.1103742

Card, N. A. (2015). *Applied Meta-Analysis for Social Science Research*. New York, NY: Guilford Publications.

*Caviola, S., Mammarella, I. C., Lucangeli, D., and Cornoldi, C. (2014). Working memory and domain-specific precursors predicting success in learning written subtraction problems. *Learn. Individ. Differ.* 36, 92–100. doi: 10.1016/j.lindif.2014.10.010

*Chan, B. M.-Y., and Ho, C. S. (2010). The cognitive profile of Chinese children with mathematics difficulties. *J. Exp. Child Psychol.* 107, 260–279. doi: 10.1016/j.jecp.2010.04.016

*Chan, W. W. L., and Wong, T. T.-Y. (2019). Visuospatial pathways to mathematical achievement. *Learn. Instruct.* 62, 11–19. doi: 10.1016/j.learninstruc.2019.03.001

Chen, C. H., and Yang, Y. C. (2019). Revisiting the effects of project-based learning on students' academic achievement: a meta-analysis investigating moderators. *Educ. Res. Rev.* 26, 71–81. doi: 10.1016/j.edurev.2018.11.001

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol. Methods* 19:211. doi: 10.1037/a0032968

*Ching, B. H. H. (2017). Mathematics anxiety and working memory: longitudinal associations with mathematical performance in Chinese children. *Contemp. Educ. Psychol.* 51, 99–113. doi: 10.1016/j.cedpsych.2017.06.006

*Cormier, D. C., Bulut, O., McGrew, K. S., and Singh, D. (2017). Exploring the relations between Cattell–Horn–Carroll (CHC) cognitive abilities and mathematics achievement. *Appl. Cognit. Psychol.* 31, 530–538. doi: 10.1002/acp.3350

*Cornoldi, C., Carretti, B., Drusi, S., and Tencati, C. (2015). Improving problem solving in primary school students: the effect of a training programme focusing on metacognition and working memory. *Br. J. Educ. Psychol.* 85, 424–439. doi: 10.1111/bjep.12083

*Costa, A. J., Lopes Silva, J. B., Chagas, P. P., Krinzinger, H., Lonneman, J., Willmes, K., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning? *Front. Psychol.* 2:368. doi: 10.3389/fpsyg.2011.00368

Cui, X., and Guo, K. (2022). Supporting mathematics learning: a review of spatial abilities from research to practice. *Curr. Opin. Behav. Sci.* 46:101176. doi: 10.1016/j.cobeha.2022.101176

*De Smedt, B., Swillen, A., Devriendt, K., Fryns, J.-P., Verschaffel, L., Boets, B., et al. (2008). Cognitive correlates of mathematical disabilities in children with velo-cardio-facial syndrome. *Genetic Counsel.* 19, 71–94.

*De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x

*Dong, Y. (2003). *Gongzuo jiyi Wenti Biaozheng Fangshi dui Xiaoxuesheng Shuxue Wenti Jiejue Yingxiang de yanjiu [Research on the influence of working memory and problem representation on pupils' mathematical problem solving]*. Changchun: Dongbei shifan daxue.

*Dube, A. K., and Robinson, K. M. (2010). Accounting for individual variability in inversion shortcut use. *Learn. Individ. Differ.* 20, 687–693. doi: 10.1016/j.lindif.2010.09.009

Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj* 315, 629–634. doi: 10.1136/bmj.315.7109.629

*Fanari, R., Meloni, C., and Massidda, D. (2019). Visual and spatial working memory abilities predict early math skills: a longitudinal study. *Front. Psychol.* 10:2460. doi: 10.3389/fpsyg.2019.02460

Felmer, P. L., Pehkonen, E., and Kilpatrick, J. (2016). *Posing and Solving Mathematical Problems*. Suiza: Springer International Publishing. doi: 10.1007/978-3-319-28023-3

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., and Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: a systematic review. *Behav. Res. Methods* 52, 2031–2052. doi: 10.3758/s13428-020-01373-9

*Foley, A. E., Vasilyeva, M., and Laski, E. V. (2017). Children's use of decomposition strategies mediates the visuospatial memory and arithmetic accuracy relation. *Br. J. Dev. Psychol.* 35, 303–309. doi: 10.1111/bjdp.12166

*Friedman, L. M., Rapport, M. D., Orban, S. A., Eckrich, S. J., and Calub, C. A. (2018). Applied problem solving in children with ADHD: the mediating roles of working memory and mathematical calculation. *J. Abnormal Child Psychol.* 46, 491–504. doi: 10.1007/s10802-017-0312-7

*Friedman, S. L., Scholnick, E. K., Bender, R. H., Vandergrift, N., Spieker, S., Pasek, K. H., et al. (2014). Planning in middle childhood: early predictors and later outcomes. *Child Dev.* 85, 1446–1460. doi: 10.1111/cdev.12221

*Fuchs, L., Fuchs, D., Compton, D., Powell, S., Seethaler, P., Capizzi, A., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29

*Fuchs, L. S., Compton, D. L., Fuchs, D., Powell, S. R., Schumacher, R. F., Hamlett, C. L., et al. (2012). Contributions of domain-general cognitive resources and different forms of arithmetic development to pre-algebraic knowledge. *Dev. Psychol.* 48, 1315–1326. doi: 10.1037/a0027475

*Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., and Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Sci. Stud. Read.* 19, 204–223. doi: 10.1080/10888438.2015.1005745

*Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., and Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *J. Educ. Psychol.* 100, 30–47. doi: 10.1037/0022-0663.100.1.30

*Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., et al. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Dev. Psychol.* 46, 1731–1746. doi: 10.1037/a0020662

*Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., and Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem solving: insights into word-problem solving as a form of text comprehension. *Sci. Stud. Read.* 22, 152–166. doi: 10.1080/10888438.2017.1398259

*Fuchs, L. S., Gilbert, J. K., Powell, S. R., Cirino, P. T., Fuchs, D., Hamlett, C. L., et al. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus prealgebraic knowledge. *Dev. Psychol.* 52, 2085–2098. doi: 10.1037/dev0000227

*Fuchs, L. S., Powell, S. R., Fall, A. M., Roberts, G., Cirino, P., Fuchs, D., et al. (2020). Do the processes engaged during mathematical word-problem solving differ along the distribution of word-problem competence? *Contemp. Educ. Psychol.* 60:101811. doi: 10.1016/j.cedpsych.2019.101811

*Gathercole, S., Alloway, T., Willis, C., and Adams, A. (2006). Working memory in children with reading disabilities. *J. Exp. Child Psychol.* 93, 265–281. doi: 10.1016/j.jecp.2005.08.003

*Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510

Geary, D. C., Hoard, M. K., and Campbell, J. I. D. (2005). *The Handbook of Mathematical Cognition*. New York, NY: Psychology Press.

Georgiou, G. K., Guo, K., Naveenkumar, N., Vieira, A. P. A., and Das, J. P. (2020). PASS theory of intelligence and academic achievement: a meta-analytic review. *Intelligence* 79:101431. doi: 10.1016/j.intell.2020.101431

*Giofre, D., Mammarella, I. C., and Cornoldi, C. (2014). The relationship among geometry, working memory, and intelligence in children. *J. Exp. Child Psychol.* 123, 112–128. doi: 10.1016/j.jecp.2014.01.002

*Greiner de Magalhaes, C., Mervis, C. B., and Cardoso-Martins, C. (2021). Cognitive predictors of arithmetic, reading, and spelling in Brazilian Portuguese-speaking children. *Read. Writ.* 34, 171–198. doi: 10.1007/s11145-020-10062-0

*Hajibaba, M., Farzad, R., and Hassan, A. (2013). A psychological model for mathematical problem solving based on revised bloom taxonomy for high school girl students. *Res. Math. Educ.* 17, 199–220. doi: 10.7468/jksmed.2013.17.3.199

*Hale, J., Fiorello, C., Bertin, M., and Sherman, R. (2003). Predicting math achievement through neuropsychological interpretation of WISC-III variance components. *J. Psychoeduc. Assess.* 21, 358–380. doi: 10.1177/073428290302100404

*Hecht, S., Torgesen, J., Wagner, R., and Rashotte, C. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: a longitudinal study from second to fifth grades. *J. Exp. Child Psychol.* 79, 192–227. doi: 10.1006/jecp.2000.2586

*Hecht, S. A., Close, L., and Santisi, M. (2003). Sources of individual differences in fraction skills. *J. Exp. Child Psychol.* 86, 277–302. doi: 10.1016/j.jecp.2003.08.003

Hedges, L. V., and Olkin, I. (2014). *Statistical Methods for Meta-Analysis*. Cambridge, MA: Academic press.

Herbst, P. (2019). "Geometric modeling tasks and opportunity to learn geometry: the ranking triangles task revisited," in *Problem Solving in Mathematics Instruction and Teacher Professional Development*, eds P. Felmer, P. Liljedahl, and B. Koichu (Cham: Springer), 123–143. doi: 10.1007/978-3-030-29215-7_7

Holmes, S. D., He, Q., and Meadows, M. (2017). An investigation of construct relevant and irrelevant features of mathematics problem-solving questions using comparative judgement and Kelly's repertory grid. *Res. Math. Educ.* 19, 112–129. doi: 10.1080/14794802.2017.1334576

Hox, J. J., Moerbeek, M., and Van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications*. London: Routledge. doi: 10.4324/9781315650982

Hu, Y., Wu, B., and Gu, X. (2017). An eye tracking study of high-and low-performing students in solving interactive and analytical problems. *J. Educ. Technol. Soc.* 20, 300–311.

*Huijsmans, M. D. E., Kleemans, T., van der Ven, S. H. G., and Kroesbergen, E. H. (2020). The relevance of subtyping children with mathematical learning disabilities. *Res. Dev. Disabil.* 104:103704. doi: 10.1016/j.ridd.2020.103704

*Iglesias-Sarmiento, V., Carriedo Lopez, N., and Rodriguez Rodriguez, J. L. (2015). Updating executive function and performance in reading comprehension and problem solving. *Anal. Psicol.* 31, 298–309. doi: 10.6018/analesps.31.1.158111

*Imbo, I., and Vandierendonck, A. (2007). The development of strategy use in elementary school children: working memory and individual differences. *J. Exp. Child Psychol.* 96, 284–309. doi: 10.1016/j.jecp.2006.09.001

Jäder, J., Sidenvall, J., and Sumpter, L. (2017). Students' mathematical reasoning and beliefs in non-routine task solving. *Int. J. Sci. Math. Educ.* 15, 759–776. doi: 10.1007/s10763-016-9712-3

*Jansen, B. R. J., De lange, E., and Van der Molen, M. J. (2013). Math practice and its influence on math skills and executive functions in adolescents with mild to borderline intellectual disability. *Res. Dev. Disabil.* 34, 1815–1824. doi: 10.1016/j.ridd.2013.02.022

*Jenks, K. M., van Lieshout, E. C., and de Moor, J. M. (2012). Cognitive correlates of mathematical achievement in children with cerebral palsy and typically developing children. *Br. J. Educ. Psychol.* 82, 120–135. doi: 10.1111/j.2044-8279.2011.02034.x

Jitendra, A. K., Sczesniak, E., and Deatline-Buchman, A. (2005). An exploratory validation of curriculum-based mathematical word problem-solving tasks as indicators of mathematics proficiency for third graders. *Sch. Psychol. Rev.* 34, 358–371. doi: 10.1080/02796015.2005.12086291

*Jogi, A.-L., and Kikas, E. (2016). Calculation and word problem-solving skills in primary grades—Impact of cognitive abilities and longitudinal interrelations with task-persistent behaviour. *Br. J. Educ. Psychol.* 86, 165–181. doi: 10.1111/bjep.12096

*Jordan, N. C., Glutting, J., and Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learn. Individ. Differ.* 20, 82–88. doi: 10.1016/j.lindif.2009.07.004

*Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., and Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *J. Exp. Child Psychol.* 116, 45–58. doi: 10.1016/j.jecp.2013.02.001

*Justicia-Galiano, M. J., Martín-Puga, M. E., Linares, R., and Pelegrina, S. (2017). Math anxiety and math performance in children: the mediating roles of working memory and math self-concept. *Br. J. Educ. Psychol.* 87, 573–589. doi: 10.1111/bjep.12165

*Keeler, M., and Swanson, H. (2001). Does strategy knowledge influence working memory in children with mathematical disabilities? *J. Learn. Disabil.* 34, 418–434. doi: 10.1177/002221940103400504

*Khng, K. H., and Lee, K. (2009). Inhibiting interference from prior knowledge: arithmetic intrusions in algebra word problem solving. *Learn. Individ. Differ.* 19, 262–268. doi: 10.1016/j.lindif.2009.01.004

*Kleemans, T., Segers, E., and Verhoeven, L. (2018). Role of linguistic skills in fifth-grade mathematics. *J. Exp. Child Psychol.* 167, 404–413. doi: 10.1016/j.jecp.2017.11.012

*Korhonen, J., Nyroos, M., Jonsson, B., and Eklöf, H. (2018). Additive and multiplicative effects of working memory and test anxiety on mathematics performance in grade 3 students. *Educ. Psychol.* 38, 572–595. doi: 10.1080/01443410.2017.1356449

Krawitz, J., and Schukajlow, S. (2018). Do students value modelling problems, and are they confident they can solve such problems? Value and self-efficacy for modelling, word, and intra-mathematical problems. *Zdm* 50, 143–157. doi: 10.1007/s11858-017-0893-1

Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., and Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learn. Individ. Differ.* 23, 123–130. doi: 10.1016/j.lindif.2012.08.001

Lawson, M. J. (2003). *Problem solving*. Netherlands: Springer, 511–524. doi: 10.1007/978-94-017-3368-7_35

*Lee, K., Ng, E. L., and Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *J. Educ. Psychol.* 101, 373–387. doi: 10.1037/a0013843

*Lee, K., Ng, S., Ng, E., and Lim, Z. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *J. Exp. Child Psychol.* 89, 140–158. doi: 10.1016/j.jecp.2004.07.001

*Lee, K., Ng, S. F., Bull, R., Pe, M. L., and Ho, R. H. M. (2011). Are patterns important? An investigation of the relationships between proficiencies in patterns, computation, executive functioning, and algebraic word problems. *J. Educ. Psychol.* 103, 269–281. doi: 10.1037/a0023068

*Lee, K., and Peh, Y. X. (2008). Differences in working memory profiles amongst children with low versus average academic performances. *Int. J. Creat. Prob. Solv.* 18, 21–33.

*LeFevre, J.-A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L., et al. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *J. Exp. Child Psychol.* 114, 243–261. doi: 10.1016/j.jecp.2012.10.005

Lenhard, W., and Lenhard, A. (2016). *Computation of Effect Sizes*. Available online at: https://www.psychometrica.de/effect_size.html (accessed October 27, 2019).

*Li, X. D., Nie, Y. Y., Pang, A. L., and Lin, C. D. (2003). Gongzuo jiyi dui xiaoxue sannianji xuesheng jiejue bijiao wenti de yingxiang[The effects of working memory on performance of third grade solving compare problems]. *Xinli Fazhan yu Jiaoyu* 3, 41–45.

*Li, Y., and Geary, D. C. (2013). Developmental gains in visuospatial memory predict gains in mathematics achievement. *PLoS One* 8:e70160. doi: 10.1371/journal.pone.0070160

*Liu, Y. Q. (2019). *Sannianji Xiaoxuesheng Zhixing Gongneng Yu Shuxue Wenti Jiejue de Guanxi Jiqi Cujin Yanjiu [The Relation between the Executive Function of the Third Grade Pupils and Mathematical Problems Solving and Promoting Research]*. Baoding: Hebei University.

Ludwig, C., Guo, K., and Georgiou, G. K. (2019). Are reading interventions for English language learners effective? A meta-analysis. *J. Learn. Disabil.* 52, 220–231. doi: 10.1177/0022219419825855

*Lukowski, S. L., Soden, B., Hart, S. A., Thompson, L. A., Kovas, Y., and Petrill, S. A. (2014). Etiological distinction of working memory components in relation to mathematics. *Intelligence* 47, 54–62. doi: 10.1016/j.intell.2014.09.001

*Mabbott, D. J., and Bisanz, J. (2008). Computational skills, working memory, and conceptual knowledge in older children with mathematics learning disabilities. *J. Learn. Disabil.* 41, 15–28. doi: 10.1177/0022219407311003

*Maennamaa, M., Kikas, E., Peets, K., and Palu, A. (2012). Cognitive correlates of math skills in third-grade students. *Educ. Psychol.* 32, 21–44. doi: 10.1080/01443410.2011.621713

Mammarella, I. C., Todeschini, M., Englaro, G., Lucangeli, and Cornoldi, C. (2012). *Geometria test [Test of geometry]*. Trento: Erickson.

Markham, T., Larmer, J., and Ravitz, J. L. (2003). *Project Based Learning Handbook: A Guide to Standards-Focused Project Based Learning for Middle and High School Teachers*. Novato, CA: Buck Institute for Education.

*Matejko, A. A., and Ansari, D. (2017). How do individual differences in children's domain specific and domain general abilities relate to brain activity within the intraparietal sulcus during arithmetic? An fMRI study. *Hum. Brain Mapp.* 38, 3941–3956. doi: 10.1002/hbm.23640

*Metcalfe, A. W., Ashkenazi, S., Rosenberg-Lee, M., and Menon, V. (2013). Fractionating the neural correlates of individual working memory components underlying arithmetic problem solving skills in children. *Dev. Cognit. Neurosci.* 6, 162–175. doi: 10.1016/j.dcn.2013.10.001

*Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., and Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learn. Individ. Differ.* 20, 101–109. doi: 10.1016/j.lindif.2009.08.004

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognit. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Muñoz-Pradas, R., Díaz-Palacios, M., Rodriguez-Martínez, E. I., and Gómez, C. M. (2021). Order of maturation of the components of the working memory from childhood to emerging adulthood. *Curr. Res. Behav. Sci.* 2:100062. doi: 10.1016/j.crbeha.2021.100062

*Navarro, J. I., Aguilar, M., Alcalde, C., Ruiz, G., Marchena, E., and Menacho, I. (2011). Inhibitory processes, working memory, phonological awareness, naming speed, and early arithmetic achievement. *Span. J. Psychol.* 14, 580–588. doi: 10.5209/rev_SJOP.2011.v14.n2.6

*Ng, J., Lee, K., and Khng, K. H. (2017). Irrelevant information in math problems need not be inhibited: students might just need to spot them. *Learn. Individ. Differ.* 60, 46–55. doi: 10.1016/j.lindif.2017.09.008

*Nunes, T., Bryant, P., Barros, R., and Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *Br. J. Educ. Psychol.* 82, 136–156. doi: 10.1111/j.2044-8279.2011.02033.x

Oberauer, K., Süß, H. M., Wilhelm, O., and Wittman, W. W. (2003). The multiple faces of working memory: storage, processing, supervision, and coordination. *Intelligence* 31, 167–193. doi: 10.1016/S0160-2896(02)00115-0

Palmer, H., and van Bommel, J. (2020). Young students posing problem-solving tasks: what does posing a similar task imply to students? *Zdm Math. Educ.* 52, 743–752. doi: 10.1007/s11858-020-01129-x

Palmer, S. (2000). Working memory: a developmental study of phonological recoding. *Memory* 8, 179–193. doi: 10.1080/096582100387597

*Passolunghi, M. C., Cargnelutti, E., and Pastore, M. (2014). The contribution of general cognitive abilities and approximate number system to early mathematics. *Br. J. Educ. Psychol.* 84, 631–649. doi: 10.1111/bjep.12054

*Passolunghi, M. C., Cargnelutti, E., and Pellizzoni, S. (2019). The relation between cognitive and emotional factors and arithmetic problem-solving. *Educ. Stud. Math.* 100, 271–290. doi: 10.1007/s10649-018-9863-y

*Passolunghi, M. C., and Mammarella, I. C. (2010). Spatial and visual working memory ability in children with difficulties in arithmetic word problem solving. *Eur. J. Cognit. Psychol.* 22, 944–963. doi: 10.1080/09541440903091127

*Passolunghi, M. C., Mammarella, I. C., and Altoe, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Dev. Neuropsychol.* 33, 229–250. doi: 10.1080/87565640801982320

*Paul, J. M., Gray, S. A., Butterworth, B. L., and Reeve, R. A. (2019). Reading and math tests differentially predict number transcoding and number fact speed longitudinally: a random intercept cross-lagged panel approach. *J. Educ. Psychol.* 111, 299–313. doi: 10.1037/edu0000287

Peng, P., Namkung, J., Barnes, M., and Sun, C. (2016). A meta-analysis of mathematics and working memory: moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *J. Educ. Psychol.* 108, 455–473. doi: 10.1037/edu0000079

Perlow, R., and Jattuso, M. (2018). A comparison of computation span and reading span working memory measures' relations with problem-solving criteria. *Psychol. Rep.* 121, 430–444. doi: 10.1177/0033294117729183

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., and Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM* 52, 33–44. doi: 10.1007/s11858-019-01118-9

Popham, M., Adams, S., and Hodge, J. (2020). Self-regulated strategy development to teach mathematics problem solving. *Intervent. Sch. Clin.* 55, 154–161. doi: 10.1177/1053451219842197

Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rösken-Winter, B., Tiemann, R., et al. (2020). A framework to foster problem-solving in STEM and computing education. *Res. Sci. Technol. Educ.* 38, 105–130. doi: 10.1080/02635143.2019.1600490

*Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., and Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: the role of problem solving strategies. *J. Exp. Child Psychol.* 141, 83–100. doi: 10.1016/j.jecp.2015.07.014

Ran, G., Li, J., Zhang, Q., and Niu, X. (2022). The association between social anxiety and mobile phone addiction: a three-level meta-analysis. *Comput. Hum. Behav.* 130:107198. doi: 10.1016/j.chb.2022.107198

Reber, P. J., and Kotovsky, K. (1997). Implicit learning in problem solving: the role of working memory capacity. *J. Exp. Psychol.* 126:178. doi: 10.1037/0096-3445.126.2.178

*Reeve, R. A., Gray, S. A., Butterworth, B. L., and Paul, J. M. (2018a). Variability in single digit addition problem-solving speed over time identifies typical, delay and deficit math pathways. *Front. Psychol.* 9:1498. doi: 10.3389/fpsyg.2018.01498

*Reeve, R. A., Reynolds, F., Paul, J., and Butterworth, B. L. (2018b). Culture-independent prerequisites for early arithmetic. *Psychol. Sci.* 29, 1383–1392. doi: 10.1177/0956797618769893

Rellensmann, J., and Schukajlow, S. (2017). Does students' interest in a mathematical problem depend on the problem's connection to reality? An analysis of students' interest and pre-service teachers' judgments of students' interest in problems with and without a connection to reality. *ZDM* 49, 367–378. doi: 10.1007/s11858-016-0819-3

*Rennie, B., Beebe-Frankenberger, M., and Swanson, H. L. (2014). A longitudinal study of neuropsychological functioning and academic achievement in children with and without signs of attention-deficit/hyperactivity disorder. *J. Clin. Exp. Neuropsychol.* 36, 621–635. doi: 10.1080/13803395.2014.921284

*Resing, W. C. M., Bakker, M., Pronk, C. M. E., and Elliott, J. G. (2017). Progression paths in children's problem solving: the influence of dynamic testing, initial variability, and working memory. *J. Exp. Child Psychol.* 153, 83–109. doi: 10.1016/j.jecp.2016.09.004

Rott, B., Specht, B., and Knipping, C. (2021). A descriptive phase model of problem-solving processes. *ZDM Math. Educ.* 53, 737–752. doi: 10.1007/s11858-021-01244-3

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., and Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educ. Stud. Math.* 79, 215–237. doi: 10.1007/s10649-011-9341-2

*Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., et al. (2012). Early predictors of high school mathematics achievement. *Psychol. Sci.* 23, 691–697. doi: 10.1177/0956797612440101

Siegler, R. S., and Shrager, J. (1984). "Strategy choices in addition and subtraction: how do children know what to do," in *Origins of Cognitive Skills*, ed. C. Sophian (Hillsdale, NJ: Lawrence Erlbaum), 229–293.

*Song, G. W., He, W. G., and Kong, W. (2011). Wenti biaozheng gongzuo jiyi dui xiaoxuesheng shuxue wenti jiejue de yingxiang [Influence of problem representation and working memory span on pupils' mathematical problem solving]. *Xinli Xuebao* 11, 1283–1292.

*Spencer, M., Fuchs, L. S., and Fuchs, D. (2020). Language-related longitudinal predictors of arithmetic word problem solving: a structural equation modeling approach. *Contemp. Educ. Psychol.* 60:101825. doi: 10.1016/j.cedpsych.2019.101825

*St Clair-Thompson, H., and Gathercole, S. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *Q. J. Exp. Psychol.* 59, 745–759. doi: 10.1080/17470210500162854

Star, J. R., and Rittle-Johnson, B. (2008). Flexibility in problem solving: the case of equation solving. *Learn. Instruct.* 18, 565–579. doi: 10.1016/j.learninstruc.2007.09.018

*Swanson, H. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Mem. Cognit.* 32, 648–661. doi: 10.3758/BF03195856

*Swanson, H., and Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *J. Educ. Psychol.* 96, 471–491. doi: 10.1037/0022-0663.96.3.471

*Swanson, H., and Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: both executive and phonological processes are important. *J. Exp. Child Psychol.* 79, 294–321. doi: 10.1006/jecp.2000.2587

*Swanson, H. L (2006a). Cognitive processes that underlie mathematical precociousness in young children. *J. Exp. Child Psychol.* 93, 239–264. doi: 10.1016/j.jecp.2005.09.006

*Swanson, H. L. (2006b). Cross-sectional and incremental changes in working memory and mathematical problem solving. *J. Educ. Psychol.* 98, 265–281. doi: 10.1037/0022-0663.98.2.265

*Swanson, H. L., Jerman, O., and Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *J. Educ. Psychol.* 100:343. doi: 10.1037/0022-0663.100.2.343

*Swanson, H. L., Kong, J. E., and Petcu, S. D. (2019). Individual differences in math problem solving and executive processing among emerging bilingual children. *J. Exp. Child Psychol.* 187:104653. doi: 10.1016/j.jecp.2019.06.006

*Swanson, H. L., Moran, A. S., Bocian, K., Lussier, C., and Zheng, X. (2013). Generative strategies, working memory, and word problem solving accuracy in children at risk for math disabilities. *Learn. Disabil. Q.* 36, 203–214. doi: 10.1177/0731948712464034

*Swanson, H. L., Orosco, M. J., and Lussier, C. M. (2014). The effects of mathematics strategy instruction for children with serious problem-solving difficulties. *Except. Child.* 80, 149–168. doi: 10.1177/001440291408000202

*Szucs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Dev. Sci.* 17, 506–524. doi: 10.1111/desc.12144

*Tam, Y. P., Wong, T. T.-Y., and Chan, W. W. L. (2019). The relation between spatial skills and mathematical abilities: the mediating role of mental number line representation. *Contemp. Educ. Psychol.* 56, 14–24. doi: 10.1016/j.cedpsych.2018.10.007

*Tavakoli, H. M. (2016). The relationship between accuracy of numerical magnitude comparisons and children's arithmetic ability: a study in iranian primary school children. *Eur. J. Psychol.* 12, 567–583. doi: 10.5964/ejop.v12i4.1175

*Traff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *J. Exp. Child Psychol.* 116, 139–156. doi: 10.1016/j.jecp.2013.04.007

*Träff, U., Olsson, L., Skagerlund, K., and Östergren, R. (2018). Cognitive mechanisms underlying third graders' arithmetic skills: expanding the pathways to mathematics model. *J. Exp. Child Psychol.* 167, 369–387. doi: 10.1016/j.jecp.2017.11.010

*Träff, U., and Passolunghi, M. C. (2015). Mathematical skills in children with dyslexia. *Learn. Individ. Differ.* 40, 108–114. doi: 10.1016/j.lindif.2015.03.024

*Trakulphadetkrai, N. V., Courtney, L., Clenton, J., Treffers-Daller, J., and Tsakalaki, A. (2020). The contribution of general language ability, reading comprehension and working memory to mathematics achievement among children with English as additional language (EAL): an exploratory study. *Int. J. Biling. Educ. Biling.* 23, 473–487. doi: 10.1080/13670050.2017.1373742

*Trezise, K., and Reeve, R. A. (2014a). Working memory, worry, and algebraic ability. *J. Exp. Child Psychol.* 121, 120–136. doi: 10.1016/j.jecp.2013.12.001

*Trezise, K., and Reeve, R. A. (2014b). Cognition-emotion interactions: patterns of change and implications for math problem solving. *Front. Psychol.* 5:840. doi: 10.3389/fpsyg.2014.00840

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., and Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behav. Res. Methods* 45, 576–594. doi: 10.3758/s13428-012-0261-6

Van den Noortgate, W., and Onghena, P. (2003). Multilevel meta-analysis: a comparison with traditional meta-analytical procedures. *Educ. Psychol. Meas.* 63, 765–790. doi: 10.1177/0013164403251027

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03

*Villeneuve, E. F., Hajovsky, D. B., Mason, B. A., and Lewno, B. M. (2019). Cognitive ability and math computation developmental relations with math problem solving: an integrated, multigroup approach. *Sch. Psychol.* 34, 96–108. doi: 10.1037/spq0000267

*Viterbori, P., Traverso, L., and Usai, M. C. (2017). The role of executive function in arithmetic problem-solving processes: a study of third graders. *J. Cognit. Dev.* 18, 595–616. doi: 10.1080/15248372.2017.1392307

*Viterbori, P., Usai, M. C., Traverso, L., and De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: a longitudinal study. *J. Exp. Child Psychol.* 140, 38–55. doi: 10.1016/j.jecp.2015.06.014

Voica, C., Singer, F. M., and Stan, E. (2020). How are motivation and self-efficacy interacting in problem-solving and problem-posing? *Educ. Stud. Math.* 105, 487–517. doi: 10.1007/s10649-020-10005-0

*Wang, A. Y., Fuchs, L. S., and Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. *Learn. Individ. Differ.* 52, 79–87. doi: 10.1016/j.lindif.2016.10.015

*Wang, F., Kuang, H., Wang, J., and Zhang, Z. H. (2010). Xiaoxue dinianji ertong shuxue yingyongti jiejue nengli yu gongzuo jiyi de guanxi yanjiu [The study on the relation between word problem solving and working memory of Lower Grade Children in Primary School]. *Z. Xuexiao Weisheng* 4, 471–473.

Witt, M. (2011). School based working memory training: preliminary finding of improvement in children's mathematical performance. *Adv. Cognit. Psychol.* 7:7. doi: 10.2478/v10053-008-0083-3

*Wong, T. T. Y. (2018). Is conditional reasoning related to mathematical problem solving? *Dev. Sci.* 21:e12644. doi: 10.1111/desc.12644

*Wong, T. T.-Y., and Ho, C. S.-H. (2017). Component processes in arithmetic word-problem solving and their correlates. *J. Educ. Psychol.* 109, 520–531. doi: 10.1037/edu0000149

*Wu, S. S., Chen, L., Battista, C., Watts, A. K. S., Willcutt, E. G., and Menon, V. (2017). Distinct influences of affective and cognitive factors on children's non-verbal and verbal mathematical abilities. *Cognition* 166, 118–129. doi: 10.1016/j.cognition.2017.05.016

*Yip, E. S. K., Wong, T. T. Y., Cheung, S. H., and Chan, K. K. W. (2020). Do children with mathematics learning disability in Hong Kong perceive word problems differently? *Learn. Instruct.* 68:101352. doi: 10.1016/j.learninstruc.2020.101352

*Zebec, M. S., Demetriou, A., and Kotrla-Topic, M. (2015). Changing expressions of general intelligence in development: a 2-wave longitudinal study from 7 to 18 years of age. *Intelligence* 49, 94–109. doi: 10.1016/j.intell.2015.01.004

*Zhang, J., Cheung, S. K., Wu, C., and Meng, Y. (2018). Cognitive and affective correlates of Chinese children's mathematical word problem solving. *Front. Psychol.* 9:2357. doi: 10.3389/fpsyg.2018.02357

*Zhang, X., Rasanen, P., Koponen, T., Aunola, K., Lerkkanen, M.-K., and Nurmi, J.-E. (2017). Knowing, applying, and reasoning about arithmetic: roles of domain-general and numerical skills in multiple domains of arithmetic learning. *Dev. Psychol.* 53, 2304–2318. doi: 10.1037/dev0000432

*Zhao, Y. (2012). *Shuxue Lingyu Zhishi Xuexi Zhong Gongzuo Jiyi yu Xueye Shuiping Dui Wenti Jiejue de Yingxiang [The Influence of Working Memory and Achievement Level on Problem-Solving in Mathematic Disciplinary Domain Knowledge]*. Guangzhou: Guangzhou University.

*Zheng, X., Swanson, H. L., and Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *J. Exp. Child Psychol.* 110, 481–498. doi: 10.1016/j.jecp.2011.06.001

*Zhong, N. N. (2009). Xiaoxue gaonianji ertong yingyongti biaozheng shuiping gongzuo jiyi dui wenti jiejue de yingxiang [The influence of elementary school students' representational level in math word problem and working memory on problem-solving]. *Zhongguo Teshu Jiaoyu* 4, 77–81.

*Zhu, M., Cai, D., and Leung, A. W. S. (2017). Number line estimation predicts mathematical skills: difference in grades 2 and 4. *Front. Psychol.* 8:1576. doi: 10.3389/fpsyg.2017.01576

*Zhu, X. B., Wang, J. L., and Li, X. F. (2011). Shikongjian gongzuo jiyi he feiyanyu liuti zhili zai xiaoxuesheng shuxue wenti jiejue zhong de zuoyong [A study on the effect of visuospatial working memory and non-verbal fluid intelligence on pupils' mathematical performance]. *Xinli Kexue* 34, 845–851.

*Zhu, Z. J. (2013). *Gongzuo Jiyi ji Tuili Nengli zai Zhongxuesheng Shuxue Wenti Jiejue Zhong de Zuoyong [The Role of Working Memory and Reasoning Ability in Middle School Students' Mathematical Problem Solving]*. Jining: Qufu Normal University.

Studies preceded by an asterisk were included in the meta-analysis.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 March 2023
DOI 10.3389/fpsyg.2023.1148754

OPEN ACCESS

EDITED BY
Belen Garcia-Manrubia,
University of Murcia, Spain

REVIEWED BY
Grace Vesga,
Universidad Antonio Nariño, Colombia
Gemma Sala Sebastià,
University of Barcelona, Spain

*CORRESPONDENCE
Lianchun Dong
✉ lianchun.dong@muc.edu.cn

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 20 January 2023
ACCEPTED 27 February 2023
PUBLISHED 28 March 2023

CITATION
Dong L, Jia X and Fei Y (2023) How growth mindset influences mathematics achievements: A study of Chinese middle school students. *Front. Psychol.* 14:1148754.
doi: 10.3389/fpsyg.2023.1148754

How growth mindset influences mathematics achievements: A study of Chinese middle school students

Lianchun Dong[1]*, Xiaoying Jia[2] and Yaxin Fei[1]

[1]College of Science, Minzu University of China, Beijing, China, [2]Shenzhen Zhenneng School, Shenzhen, Guangdong, China

Introduction: It has been suggested that students with growth mindsets are more likely to achieve better mathematics learning results than their counterparts with fixed mindsets. However, inconsistent and some even contradictory results have been reported in recent studies which examined the associations between growth mindset and mathematics achievements, suggesting the complexity regarding the effects of growth mindset on academic achievements.

Methods: This study aims to examine students' growth mindsets, failure attributions, intrinsic motivation, mathematics self-efficacy, mathematics anxiety and mathematics achievements in one model to capture the sophisticated functioning processes of growth mindset. A total number of 266 middle school students in China participated in this study. Students' mindset and related variables (i.e.,motivations to learn mathematics, attributions of failure in mathematics, mathematics anxiety, mathematics self-efficacy) were measured at year 7, the first year of junior middle school in China. These students' mathematics learning outcomes were tracked from year 7 to year 9, the end of junior middle school. Structural equation modeling (SEM) was used to investigate the relations among students' growth mindsets, failure attributions, intrinsic motivation, mathematics self-efficacy, mathematics anxiety and mathematics achievements.

Results: The results show that: (1) growth mindset doesn't directly predict mathematics achievements; (2) growth mindset indirectly influences mathematics achievements through intrinsic motivation; (3) failure attributions and mathematics self-efficacy sequentially mediate the association between growth mindset mathematics achievements; (4) failure attributions and mathematics anxiety sequentially mediate the relationship between growth mindset mathematics achievements.

Discussion: The results of this study contribute a better understanding about how growth mindsets make impacts on middle school students' mathematics achievements. These findings have important implications for mathematics education in that we could not simply cultivate students' growth mindset in schools with expectations of higher mathematics learning outcomes. Instead, along with the growth mindset intervention, it is fundamental to make interventions on students' intrinsic motivation, failure attribution, mathematics self-efficacy, and mathematics anxiety in mathematics teaching and learning.

KEYWORDS

growth mindset, fixed mindset, mathematics, achievements, attributions, intrinsic motivation, mathematics self-efficacy, mathematics anxiety

1. Introduction

The theory of growth mindset has attracted researcher's interest in the past decade. A growth mindset refers to the belief that one person's ability can be developed through efforts, whereas a fixed mindset means viewing ability as fixed and unchangeable (Yeager and Dweck, 2012; Xu et al., 2022). During the last decades, there has been a controversy regarding whether a growth mindset could predict mathematics learning outcomes (Burnette et al., 2013; Yeager and Dweck, 2020). Some researchers reported a positive association between a growth mindset and mathematics achievement (Blackwell et al., 2007), whereas other studies showed no or negative correlations between these two variables (Bahník and Vranka, 2017; Li and Bates, 2019; Yeager and Dweck, 2020).

These results prompted researchers to reconsider the growth mindsets' predictive role in students' mathematics learning. To obtain further insights into the above discrepancies, researchers proposed to conduct further investigations to understand the sophisticated mechanisms in which growth mindsets and related variables (e.g., motivation and attributions) function together to influence academic outcomes (Burgoyne et al., 2020; Yeager and Dweck, 2020). Thus, it is necessary to examine multiple related variables and their mutual connections together in one study. This study aims to investigate the effects of a growth mindset on students' mathematics achievement by considering failure attributions, intrinsic motivation, mathematics self-efficacy, and mathematics anxiety together.

2. Literature review

This section presents previous studies involving the connections between a growth mindset and mathematics achievements and the impacts of failure attributions, intrinsic motivation, mathematics self-efficacy, and mathematics anxiety.

2.1. Failure attribution in mathematics learning and growth mindset

Failure attributions refer to students' inclination to attribute their academic failure or setbacks to possible impacting factors. Failure attributions orient students toward different patterns of responding to failure and setbacks in learning (e.g., whether students take remedy strategies or give up) and thereby have significant impacts on academic outcomes (Weiner, 2010; Organisation for Economic Co-operation and Development and Programme for International Student Assessment, 2019).

It is reported that students with growth mindsets tend to interpret academic failure and setbacks differently from those with fixed mindsets (De Castella and Byrne, 2015; Martin et al., 2017). When reflecting on the cause of failure, students with growth mindsets usually focus on controllable characteristics (e.g., insufficient efforts). For example, Hong et al. (1999) reported that students with growth mindsets are more likely to make effort attribution, believing the cause of failure is their lack of effort. By contrast, students with fixed mindsets tend to attribute failure to uncontrollable aspects (e.g., the lack of ability). Previous studies reported that fixed mindsets are significantly associated with ability attribution in the failure contexts, indicating that students with fixed mindsets perceived low ability as the reason for failure (Robins and Pals, 2002; Tempelaar et al., 2014; Smiley et al., 2016). The ability attribution is also called helpless attribution by Yeager and Dweck (2020), who claimed that students with a fixed mindset are inclined to attribute their undesirable performances to a stable flaw and thereby show helpless behavior when facing academic setbacks.

2.2. Mathematics self-efficacy and growth mindset

Self-efficacy refers to individuals' beliefs about their abilities to achieve a certain level of performance in specific activities or tasks (Bandura, 1977). Self-efficacy has been reported to make an impact on students' mathematics learning processes and outcomes (Hackett and Betz, 1989; Pajares and Kranzler, 1995; Zimmerman, 2000; Huang et al., 2019; Organisation for Economic Co-operation and Development and Programme for International Student Assessment, 2019). Researchers pointed out that growth and fixed mindset are significantly correlated with students' mathematics self-efficacy (Young and Urdan, 1993; Abdullah, 2008; Todor, 2014; van Aalderen-Smeets et al., 2018). It was reported that higher levels of mathematics self-efficacy are more likely to be observed when students adopt a growth mindset, whereas students with fixed mindsets usually have lower levels of mathematics self-efficacy (Todor, 2014). van Aalderen-Smeets et al. (2018) claimed that the adoption of a growth mindset could help students to maintain a relatively stable level of self-efficacy when facing failure and difficulties in learning (van Aalderen-Smeets et al., 2018). In contrast, students with a fixed mindset tend to believe failure is the result of a lack of capability rather than effort, which usually leads to a decline in self-efficacy.

Davis et al. reported that a fixed mindset could result in the feeling of helplessness when facing setbacks and challenges, which in turn decreases mathematics self-efficacy (Davis et al., 2010). Recent intervention studies reported that growth mindset intervention could improve students' self-efficacy (Samuel and Warner, 2019; Zhang et al., 2022). For example, a two-semester intervention based on a growth mindset was executed in a college statistics course. At the end of the intervention, students in the intervention group reported a higher level of mathematics self-efficacy than those in the control group (Samuel and Warner, 2019).

The roles of self-efficacy are also underlined when investigating the processes of how growth and fixed mindset impact on academic learning (Burgoyne et al., 2020). Some researchers conducted a regression analysis to confirm that a growth mindset positively predicts self-efficacy and a fixed mindset negatively predicts self-efficacy (BrÅten et al., 2005). Some researchers claimed that students with fixed mindsets responded differently to failure, depending on the level of students' self-efficacy (Dweck and Leggett, 1988; Gonida et al., 2006). In general, low persistence and challenge avoidance are more likely to be observed when students are more inclined to believe intelligence is unchallengeable. However, if these students have a higher level of self-efficacy, they are less likely to hesitate when facing an academic challenge. Instead, these students with a higher level of self-efficacy, despite their tendency toward a fixed mindset, could embrace challenges in academic learning and maintain persistence when experiencing failure.

Researchers also reported that self-efficacy plays a mediating role between a growth mindset and academic achievements. Leondari and Gialamas (2002) found that a growth mindset promotes students' orientation toward learning goals, leading to a higher level of self-efficacy, which in turn enhances academic achievements (Leondari and Gialamas, 2002). Other studies reported that a growth mindset leads to higher academic achievements through the chain mediating effects of mathematics self-efficacy and beliefs of failure (Su et al., 2021).

Failure attributional style is closely linked with self-efficacy (Schunk, 1981; Wang et al., 2008; Siegle et al., 2009; Cheng and Chiou, 2010). Students' self-efficacy declines as a result of attributing failure to stable and uncontrollable factors (e.g., the ability of students with fixed mindsets; Silver et al., 1995). By contrast, when making adaptive attribution for academic failure (e.g., attributing failure to the lack of effort), students can develop their self-efficacy and believe their academic performance can be improved by continuous efforts (Baird et al., 2009).

Further studies show that there is a chain-mediating role of effort beliefs, failure attribution, and positive remedy strategies in the connections between a growth mindset and academic achievements (Blackwell et al., 2007). However, Cheng and Chiou (2010) reported that the level of self-efficacy can be weakened when students attribute failure to personal factors, regardless of whether it is personal ability or personal efforts. By contrast, students' self-efficacy can be improved when attributing failure to situational factors, such as the difficulty level of a test and luck.

2.3. Mathematics anxiety and growth mindset

Mathematics anxiety refers to students' feelings of anxiety and tension when interfering with mathematics knowledge and tasks (Richardson and Suinn, 1972; Essau et al., 2008; Radišić et al., 2014; Dirzyte et al., 2021; Li et al., 2021). Mathematics anxiety usually leads to the avoidance of mathematics tasks and activities (Hembree, 1990; Dowker et al., 2016), disrupts students' attention and working memory (Cohen and Rubinsten, 2017), and results in poor mathematics performances (Wang et al., 2015; Byrnes and Miller-Cotto, 2016; Ramirez et al., 2018; Kaskens et al., 2020; Geary et al., 2021).

Mathematics anxiety is also related to failure attributions. Researchers reported that students attributing failure in mathematics to ability tend to have a high level of mathematics anxiety (Arkin et al., 1983; Hunsley, 1987). Dweck and Licht (1980) also suggested that failure attributions can explain the observed gender differences regarding learning anxiety. Further studies investigated the effects of failure attributions on mathematics anxiety, presenting that effort attributions predicted a low level of mathematics anxiety in exam contexts, whereas ability attributions led to a high level of mathematics anxiety (Bandalos et al., 1995).

In addition, researchers also claimed that students' anxious experiences in mathematics learning are always related to their perceptions of personal abilities (Clark, 2021; Young and Dyess, 2021). Intervention studies also reported that developing students' growth mindset in mathematics classrooms can successfully relieve mathematics anxiety and increase mathematics achievements (Boaler, 2013; Smith and Capuzzi, 2019; Clark, 2021; Young and Dyess, 2021). These findings support the assumption that a growth mindset might influence mathematics achievements through attributions and mathematics anxiety.

2.4. Intrinsic motivation to learn mathematics and growth mindset

Learning motivation refers to the psychological processes that drive students to engage in learning activities and specific tasks. In general, there are two types of the motivation behind mathematics learning: learning mathematics because of individual enjoyment and interest (i.e., intrinsic motivation) and learning mathematics due to external rewards (i.e., extrinsic or instrumental motivation; Ryan and Deci, 2009; Wigfield et al., 2009; Organisation for Economic Co-operation and Development and Programme for International Student Assessment, 2019). Compared with extrinsic motivation, intrinsic motivation is more likely to contribute to high-quality learning behavior (e.g., high-level engagement, more persistence, and efforts) and outcomes (e.g., deep understanding and high academic achievements; Murayama et al., 2012; Taylor et al., 2014; Gottfried, 2019; Karlen et al., 2019).

Previous studies show that students' mindsets can make an impact on their learning motivation (Molden and Dweck, 2000; Dweck, 2002). Students with growth mindsets tend to have a stronger intrinsic motivation to learn, whereas learners with fixed mindsets are more likely to be motivated by external rewards (Blackwell et al., 2007; Komarraju and Nadler, 2013; Tempelaar et al., 2014). Blackwell et al. (2007) conducted an intervention to promote students' growth mindset, reporting that students in the experimental group showed larger improvements in intrinsic motivation than those in the control group. This finding was supported by other intervention studies, which highlighted that effective mindset interventions contributed to maintaining a high level of students' motivation to learn mathematics (Priess-Groben and Hyde, 2016).

Researchers also reported that the link between a growth mindset and better academic results could be explained by considering the effects of a growth mindset on learning motivations. Compared with those with a fixed mindset, students with a growth mindset usually focus more on skill improvements by learning, and thereby are inclined to maintain a high level of intrinsic motivation, which results in better learning behavior and outcomes, especially when facing challenging learning tasks (Burnette et al., 2013; Karlen et al., 2019). Degol et al. (2017) found that students holding a malleable perspective of mathematics intelligence tend to have more appreciation for the value of mathematics and for personal connections with mathematics and develop more intrinsic motivation to engage in mathematics learning, which ultimately leads to higher mathematics achievements.

3. The present study and research hypotheses

This study aims to include growth mindset, failure attribution, intrinsic motivation, mathematics self-efficacy, and mathematics anxiety together to examine the processes in which a growth mindset influences mathematics achievements. Based on the above discussion, a conceptual research model is established to demonstrate the hypothesized relationships (see Figure 1). The research hypotheses in this study are outlined below.

H1: A growth mindset would predict mathematics achievements. Students with growth mindsets would have higher mathematics achievements than students with fixed mindsets.

H2: Intrinsic motivation to learn mathematics would mediate the relationship between a growth mindset and mathematics achievement. Students with growth mindsets are more likely to have stronger learning motivation than their peers with fixed mindsets and, thereby, have better learning results in mathematics.

H3: There would be a significant sequential mediation path from growth mindset to perceived self-responsibility for failing in mathematics, to mathematics self-efficacy, and then to mathematics achievement. Compared with students with fixed mindsets, students with growth mindsets are more inclined to attribute undesirable mathematics learning results to themselves (e.g., efforts) rather than external factors (e.g., luck). Therefore, students with growth mindsets have a better feeling of control over their mathematics learning, which leads to higher mathematics self-efficacy, believing that they would improve their mathematics learning by making more effort. This could, in turn, contribute to better mathematics learning results.

H4: There would be a significant sequential mediation path from a growth mindset to perceived self-responsibility for failing in mathematics, to mathematics anxiety, and then to mathematics achievement. As stated in H3, a growth mindset leads to the tendency to attribute undesirable mathematics learning results to oneself (e.g., efforts) rather than external factors (e.g., luck). This attribution style can reduce students' mathematics anxiety because the efforts to improve learning mathematics can be controlled by students. With a lower level of mathematics anxiety, students are more likely to make higher mathematics achievements.

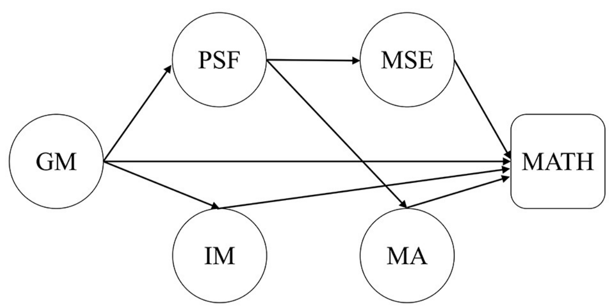


FIGURE 1
Conceptual research model. GM, growth mindset; MSE, mathematics self-efficacy; IM, intrinsic motivation to learn mathematics; MA, mathematics anxiety; PSF, perceived self-responsibility for failing in mathematics; MATH, mathematics achievement.

4. Materials and methods

4.1. Participants

This study selected junior middle school students from one middle school in Hengshui City, Hebei Province, China. We advertised our research projects to a couple of middle schools in Hebei Province. But because this study requires the tracking of students' achievements over the period through years 7–9, most of the middle schools could not make sure whether they could complete the collections of students' achievements in such a long period of time. Therefore, only one middle school agreed to participate in this study and was interested in examining the effects of students' growth mindset on mathematics achievements over students' whole middle school life.

This junior middle school is located in the urban region of the city, consisting of 3-year levels, respectively years 7, 8, and 9. In China education system, students have 6 years of primary education, followed by 3 years of junior middle school study, i.e., years 7, 8, and 9. At the end of year 9, students sit for a high-stake test that determines whether students can be admitted by senior middle schools.

At the end of year 7, participants filled in the scale of growth mindset, mathematics self-efficacy, motivation to learn mathematics, mathematics anxiety, and perceived self-responsibility for failing in mathematics. Then we tracked participants' mathematics learning achievements from years 8 to 9. Participants' mathematics achievements at three points were collected: the end of year 8, the end of the first semester of year 9, and the end of the second semester of year 9. We did not collect data about students' ages, but according to the Chinese education system, school education from years 1 to 9 is compulsory for all children, and thereby all participants in the study were 12 or 13 years old in year 7.

A total number of 266 students participated in all rounds of data collection. Only students participating in all rounds of data collection were included in the data analysis.

4.2. Instruments

4.2.1. Mathematics achievements

When exampling the effects of a growth mindset on mathematics achievements, one of the issues raised by researchers is the limitations of the cross-sectional design adopted in most of the previous studies. The measurement of student mathematics achievements at one single point in time might constrain the observation of the long-term impacts that growth mindsets might exert on students' mathematics learning. Thus, this study attempted to collect students' achievement data at multiple points.

Participants' mathematics achievements were measured by mathematics assessments designed and administered by the local education department. The items of mathematics assessments involved the mathematics facts, skills, and problem-solving strategies required in the National Mathematics Curriculum Standards issued by the Ministry of Education in China. Three mathematics tests' results were collected separately at the end of year 8, the middle, and the end of year 9. Each test lasted 2 h and had a total score of 120. An average of the three tests' results was used to measure students' overall mathematics achievements in junior middle school study.

4.2.2. Growth mindset

The growth mindset measure was adopted from Dweck et al.'s (1995) and Dweck (1999) work. Five items are included in the growth mindset scale. Two items are growth mindset statements (e.g., "You can always greatly change how intelligent you are") and the other three are fixed mindset statements (e.g., "Your intelligence is something about you that you can't change very much"). The 6-point Likert responses were used, and participants needed to select one choice from 1 (strongly disagree) to 6 (strongly agree). The fixed mindsets items were scored reversely. Scores on the six items were added altogether as an overall growth mindset score, with higher scores indicating stronger beliefs in a growth mindset. The reliability of this measure was $\alpha = 0.79$ ($N = 266$).

4.2.3. Mathematics self-efficacy

Mathematics self-efficacy measure was adopted from the context questionnaire in PISA 2009 survey (Organisation for Economic Co-operation and Development, 2010; PISA 2009 Shanghai Committee, 2016). This measure included seven mathematics tasks with different levels of cognitive load (e.g., "Solving an equation like $3x + 5 = 17$," "Using a train schedule to figure out how long it would take to get from one place to another"). Participants were asked to express their confidence in solving these tasks by selecting a choice from 1 (not at all confident) to 4 (very confident). The sum of scores on all seven tasks was used as the overall measure of mathematics self-efficacy, with higher scores representing a higher level of mathematics self-efficacy. The reliability of this measure was $\alpha = 0.87$ ($N = 266$).

4.2.4. Intrinsic motivation to learn mathematics

Intrinsic motivation to learn mathematics measure was adopted from the context questionnaire in PISA 2009 survey (Organisation for Economic Co-operation and Development, 2010; PISA 2009 Shanghai Committee, 2016). This measure included four items describing the motivation in mathematics learning (e.g., "I do mathematics because I enjoy it"). Participants need to select one choice from 1 (strongly disagree) to 4 (strongly agree). The sum of scores on all four items was used as the overall measure of intrinsic motivation to learn mathematics, with higher scores representing stronger intrinsic motivation. The reliability of this measure was $\alpha = 0.89$ ($N = 266$).

4.2.5. Mathematics anxiety

Mathematics anxiety measure was adopted from the context questionnaire in PISA 2009 survey (Organisation for Economic Co-operation and Development, 2010; PISA 2009 Shanghai Committee, 2016). This measure included three items about mathematics learning experiences (e.g., "I often worry that it will be difficult for me in mathematics classes"). Participants need to select one choice from 1 (strongly disagree) to 4 (strongly agree). The sum of scores on all three items was used as the overall measure of mathematics anxiety, with higher scores indicating a higher level of anxiety in mathematics learning. The reliability of this measure was $\alpha = 0.86$ ($N = 266$).

4.2.6. Perceived self-responsibility for failing in mathematics

In order to measure students' failure attributions, perceived self-responsibility for failing in mathematics measure was adopted from the context questionnaire in PISA 2009 survey (Organisation for Economic Co-operation and Development, 2010; PISA 2009 Shanghai Committee, 2016). This measure provides a scenario about participants' undesirable results in mathematics quizzes and asks participants whether they agree with six given statements (e.g., "I'm not very good at solving mathematics problems"). Participants need to select one choice from 1 (very likely) to 4 (not at all likely). The sum of scores on all six items was used as the overall measure of perceived self-responsibility for failing in mathematics, with higher scores indicating a higher tendency to attribute undesirable mathematics learning results to oneself. The reliability of this measure was $\alpha = 0.79$ ($N = 266$).

4.3. Data analysis

SPSS 22.0 and Mplus 7.0 were used for data analysis. Correlation analysis was conducted to examine the relationship between different measures. Structure equation modeling with latent variables was used to test the mediating effects of the motivational variables (i.e., mathematics self-efficacy, intrinsic motivation to learn mathematics, mathematics anxiety, and perceived self-responsibility for failing in mathematics) in the relationship between growth mindset and mathematics achievement.

5. Results

5.1. Descriptive statistics

Table 1 presents descriptive statistics, and Table 2 shows the results of correlation analyses. It can be seen that all variables are intercorrelated significantly except that between growth mindset and mathematics achievement. Both growth mindset and mathematics achievement are positively correlated with mathematics self-efficacy, intrinsic motivation to learn mathematics, and perceived self-responsibility for failing in mathematics but negatively correlated with mathematics anxiety.

TABLE 1 Summary of descriptive statistics.

Measures	*N*	*Min*	*Max*	*Mean*	*SD*
Growth mindset	266	5	30	22.81	5.09
Math self-efficacy	266	7	28	23.86	3.85
Intrinsic motivation to learn math	266	4	16	12.02	2.63
Math anxiety	266	3	12	7.27	2.30
Perceived self-responsibility for failing in math	266	6	24	17.60	3.27
Math achievement	266	9	110	51.57	25.0

TABLE 2 Summary of correlation analyses.

	1	2	3	4	5
1. Growth mindset					
2. Math self-efficacy	0.21**				
3. Intrinsic motivation to learn math	0.36**	0.49**			
4. Math anxiety	−0.17**	−0.42**	−0.38**		
5. Perceived self-responsibility for failing in math	0.33**	0.52**	0.69**	−0.51**	
6. Math achievement	0.10	0.49**	0.42**	−0.37**	0.45**

**$p < 0.01$.

5.2. SEM analyses

5.2.1. Common method biases

Because self-reported items were used in the measurement, Harman's single-factor test was used to assess common method bias. Exploratory factor analysis (EFA) was conducted on all the items. Unrotated EFA shows that KMO (Kaiser–Meyer–Olkin's measure) is 0.90 (>0.8), and χ^2 for Bartlett's sphericity test is significant ($p < 0.001$). There are five factors with an eigenvalue higher than one, and the first common factor explained 33.1% of the total variance in the variables. This proportion is less than the threshold of 40%, suggesting that the problem of common method bias is not present in the data (see Table 3).

5.2.2. Structural equation modeling results

Structural equation modeling with latent variables was used to examine the fit of our full model (see Figure 2). Growth mindset was indexed by five items, mathematics self-efficacy was indexed by seven items, intrinsic motivation to learn mathematics was indexed by four items, mathematics anxiety was indexed by three items, and perceived self-responsibility for failing in mathematics was indexed by six items (see the 4.2 section, for detailed explanations about all the items). Mathematics achievement was used as an outcome

TABLE 3 Harman's single-factor test.

	KMO	χ^2	*P*	Number of factors	Explained variance by the first factor
All items	0.90	3326.93	<0.001	25	33.1%
Recommended threshold	>0.8	-	<0.05	>1	<40%

variable. The full model had an acceptable fit to the data (see Table 4).

We did not a direct link between a growth mindset and mathematics achievement, which is inconsistent with the hypothesized model (see Figure 1 in Section 3). Actuality, it can be seen in Figure 2 that a growth mindset did not directly predict mathematics achievement, but rather made impacts on mathematics achievement *via* other mediating variables.

A growth mindset positively predicted both perceived self-responsibility for failing in mathematics ($\beta = 0.379$, $p < 0.001$) and intrinsic motivation to learn mathematics ($\beta = 0.356$, $p < 0.001$). Perceived self-responsibility for failing in mathematics positively predicted mathematics self-efficacy ($\beta = 0.637$, $p < 0.001$) and negatively predicted mathematics anxiety ($\beta = -0.637$, $p < 0.001$). Mathematics achievement was positively predicted by mathematics self-efficacy ($\beta = 0.348$, $p < 0.001$) and intrinsic motivation to learn mathematics ($\beta = 0.199$, $p < 0.05$) but negatively predicted by mathematics anxiety ($\beta = -0.152$, $p < 0.05$).

5.2.3. Mediation analysis

The full model suggests indirect effects of a growth mindset on mathematics achievements through three mediational pathways: (1) the sequential mediation effect from growth mindset to perceived self-responsibility for failing in mathematics to mathematics self-efficacy to mathematics achievements; (2) the sequential mediation effect from growth mindset to perceived self-responsibility for failing in mathematics to mathematics anxiety to mathematics achievements; (3) intrinsic motivation to learn mathematics mediates the relationship between growth mindset and mathematics achievements.

The mediation effects were evaluated by the bootstrapping method with 1,000 bootstrap data samples, and the results are shown in Table 5. It can be seen that zero is not included in the 95% confidence intervals, suggesting that all the indirect effects are statistically significant.

6. Discussion

6.1. A growth mindset does not directly predict mathematics achievements

This study shows that a growth mindset does not directly predict mathematics achievements, indicating that H1 is not supported. This is not consistent neither with the mindset theory

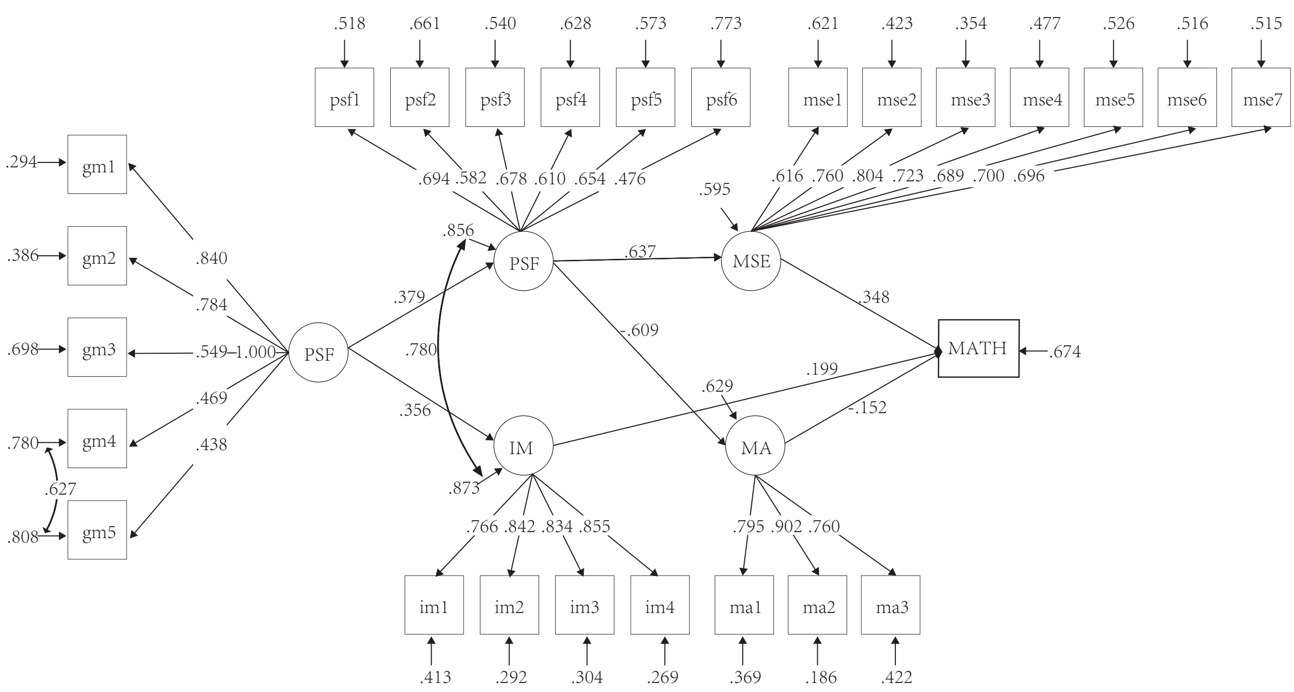


FIGURE 2
The structural equations model of the relationships growth mindset. GM, growth mindset; MSE, mathematics self-efficacy; IM, intrinsic motivation to learn mathematics; MA, mathematics anxiety; PSF, perceived self-responsibility for failing in mathematics; MATH, mathematics achievement. All paths $p < 0.05$.

TABLE 4 Fit indices of the full model.

	χ	df	χ^2/df	RMSEA	CFI	TLI	SRMR
Full model	510.22	291	1.753	0.053	0.933	0.925	0.058
Suggested threshold	-	-	<5	<0.08	>0.9	>0.9	<0.08

TABLE 5 Summary of mediation analysis.

Path	Effect size	95% confidence interval
GM→PSF→MSE→MATH	0.084	[0.047, 0.133]
GM→PSF→MA→MATH	0.035	[0.006, 0.082]
GM→IM→MATH	0.071	[0.020, 0.141]
Total	0.190	[0.118, 0.262]

GM, growth mindset; MSE, mathematics self-efficacy; IM, intrinsic motivation to learn mathematics; MA, mathematics anxiety; PSF, perceived self-responsibility for failing in mathematics; MATH, mathematics achievement.

proposed by Yeager and Dweck (2020) nor with the evidence of the direct link between a growth mindset and academic achievements (Costa and Faria, 2018). However, this finding is in line with the results reported by some other researchers (e.g., Li and Bates, 2019; Burgoyne et al., 2020). This study supports the claim that the connections between a growth mindset and mathematics achievements are sophisticated, depending on such factors as cultural contexts (Yeager and Dweck, 2020; Dong and Kang, 2022).

By citing the PISA data (Schleicher, 2019), Yeager and Dweck (2020) claimed that the links between a growth mindset and achievement might be the weakest in Chinese contexts compared with other cultures, such as Western contexts. The weak link was attributed to the Chinese culture of valuing diligence and efforts, and thereby there might be little space to get improvements in study hours or exam scores resulting from the adoption of a growth mindset. In order to further explain the heterogeneity, Yeager and Dweck (2020) suggested conducting cross-cultural studies to compare the effects of a growth mindset in different cultures.

6.2. The mediating role of intrinsic motivation

This study finds that a growth mindset indirectly influences mathematics achievements through intrinsic motivation, supporting H2. This is consistent with the previous findings that students with growth mindset motivation are more likely to maintain strong learning motivation and consequently achieve better results than those with fixed mindsets (Burnette et al., 2013; Degol et al., 2017; Karlen et al., 2019). Previous studies selected high school students in Western contexts and showed that students viewing intelligence as malleable are more inclined to engage in challenging mathematics activities and tasks than students with fixed mindsets (Jones et al., 2011; Degol et al., 2017; Karlen et al., 2019). Therefore, students with growth mindsets are more likely to develop an interest and enjoyment in mathematics learning, which leads to fewer chances to withdraw from mathematics tasks and a higher possibility of obtaining desirable mathematics achievements. By collecting data on middle school students in China, this study provides further evidence that the mediating roles of intrinsic motivation can be observed in the stage of middle school learning in Chinese contexts.

6.3. The sequential mediation effect of failure attribution and mathematics self-efficacy

This study shows that a growth mindset indirectly influences mathematics achievements through other factors. It presents that perceived self-responsibility for failing in mathematics and mathematics self-efficacy sequentially mediate the association between growth mindset mathematics achievements, supporting H3. This finding highlights the mediating roles of failure attribution between growth mindset and academic achievements, which is consistent with previous studies (Cheng and Chiou, 2010; Yeager and Dweck, 2020).

This study also reports the mediating role of mathematics self-efficacy in the relationship between growth mindset and mathematics achievements, supporting the previous findings that high mathematics self-efficacy is a significant variable when explaining the process in which growth mindset could influence mathematics achievements (Su et al., 2021). It is noteworthy that Su et al. focused on primary school students in their study, highlighting the roles of mathematics self-efficacy in primary mathematics learning. By observing similar results in the middle school students' samples, this study provides further evidence that the roles of mathematics self-efficacy in mindset theory are generalizable across grade levels.

Another contribution of this study is to report the sequential mediation effects of failure attributions and mathematics self-efficacy, highlighting how these two variables might work together to explain the relationship between growth mindset and mathematics achievements. It is claimed that mindset theory constructs a meaning system to explain discrepancies regarding students' responses to challenging situations and failures (Dweck and Yeager, 2019; Yeager and Dweck, 2020). Failure attributions have been reported to be the core of mindset theory in that different mindsets guide students toward different attributions in the face of failure, which in turn exerts great impacts on students' learning behavior and outcomes (Yeager and Dweck, 2020). However, very few studies consider the roles of attribution and self-efficacy together in one model to examine the functioning mechanisms of a growth mindset (Burgoyne et al., 2020). By reporting the sequential mediation effects of failure attributions and mathematics self-efficacy, this study contributes to a better understanding of the processes of how a growth mindset impact on academic achievements.

6.4. The sequential mediation effect of failure attribution and mathematics anxiety

This study also presents that perceived self-responsibility for failing in mathematics and mathematics anxiety sequentially mediate the relationship between growth mindset mathematics achievements, supporting H4. This finding suggests the necessity of including mathematics anxiety in order to better understand the processes by which a growth mindset can influence academic achievement.

The inclusion of mathematics anxiety in this study is based on previous findings that mathematics anxiety is observed to decrease during the interventions aiming to foster students' development of a growth mindset (Smith and Capuzzi, 2019; Clark, 2021; Young and Dyess, 2021). However, when examining the processes in which a growth mindset might influence academic achievements, previous studies (e.g., Yeager and Dweck, 2020) rarely include learning anxiety as a variable in the model, making it unclear whether learning anxiety can help to explain the relationship between growth mindset, attributions, and academic achievements. Thus, this study contributes to show evidence that a growth mindset can shape students' failure attributions, and thereby reduce their mathematics anxiety, which consequently results in better mathematics achievements.

7. Conclusion and limitations

This study examined Chinese middle school students' growth mindsets, failure attributions, intrinsic motivation, mathematics self-efficacy, mathematics anxiety, and mathematics achievements in one model, aiming to better understand how growth mindset impact on students' mathematics achievements over the period from years 7 to 9. The findings show that a growth mindset does not directly predict mathematics achievements in middle school study, but rather indirectly influences mathematics achievements through other variables, i.e., intrinsic motivation, failure attribution, mathematics self-efficacy, and mathematics anxiety. These findings have important implications for mathematics education in that we could not simply cultivate students' growth mindset in schools with expectations of higher mathematics learning outcomes. Instead, along with the growth mindset intervention, it is fundamental to make interventions on students' intrinsic motivation, failure attribution, mathematics self-efficacy, and mathematics anxiety in mathematics teaching and learning.

This study has some limitations that need to be considered when interpreting the results. First, the sample included only middle school students in China, and it is unclear whether the same results can be generalized to other cultural contexts. Thus, more investigations in different settings are necessary to test the generalizability of the findings in this study. Second, this study did not consider all the impacting factors related to a growth mindset, such as goal orientations and effort beliefs. Future research is needed to construct a more sophisticated model to consider these factors altogether to investigate their interrelationships. Third, although students' mathematics achievements were measured several times from years 7 to 9, we measured other variables (e.g., mathematics anxiety) only in year 7. Because other variables (e.g., mathematics anxiety) might change over the period, this study could not take into account the changes in these variables and the corresponding impacts on students' mathematics achievements. Further studies can measure these variables several times to track the possible changes in these variables to better understand how mathematics achievements might be influenced by these variables through middle school study.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee in College of Science, Minzu University of China. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

LD designed the study, collected the data, and revised the draft of the manuscript. XJ and YF analyzed the data and drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Social Science Foundation of China (Grant number: CHA200260).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdullah, M. N. L. Y. (2008). Children's implicit theories of intelligence: Its relationships with self-efficacy, goal orientations, and self-regulated learning. *Int. J. Learn.* 15, 47–56. doi: 10.18848/1447-9494/CGP/v15i02/45623

Arkin, R. M., Kolditz, T. A., and Kolditz, K. K. (1983). Attributions of the test-anxious student. *Pers. Soc. Psychol. B* 9, 271–280. doi: 10.1177/0146167283092012

Bahník, Š., and Vranka, M. A. (2017). Growth mindset is not associated with scholastic aptitude in a large sample of university applicants. *Pers. Indiv. Differ.* 117, 139–143. doi: 10.1016/j.paid.2017.05.046

Baird, G. L., Scott, W. D., Dearing, E., and Hamill, S. K. (2009). Cognitive self-regulation in youth with and without learning disabilities: Academic self-efficacy, theories of intelligence, learning vs. performance goal preferences, and effort attributions. *J. Soc. Clin. Psychol.* 28, 881–908. doi: 10.1521/jscp.2009.28.7.881

Bandalos, D. L., Yates, K., and Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *J. Educ. Psychol.* 87, 611–623. doi: 10.1037/0022-0663.87.4.611

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Blackwell, L. S., Trzesniewski, K. H., and Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Dev.* 78, 246–263. doi: 10.1111/j.1467-8624.2007.00995.x

Boaler, J. (2013). Ability and mathematics: The mindset revolution that is reshaping education. *Forum* 55, 143. doi: 10.2304/forum.2013.55.1.143

BrÅten, I., StrØms,Ø, H. I., and Samuelstuen, M. S. (2005). The relationship between internet-specific epistemological beliefs and learning within internet technologies. *J. Educ. Comput. Res.* 33, 141–171. doi: 10.2190/E763-X0LN-6NMF-CB86

Burgoyne, A. P., Hambrick, D. Z., and Macnamara, B. N. (2020). How firm are the foundations of mind-set theory? The claims appear stronger than the evidence. *Psychol. Sci.* 31, 258–267. doi: 10.1177/0956797619897588

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., and Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychol. Bull.* 139, 655–701. doi: 10.1037/a0029531

Byrnes, J. P., and Miller-Cotto, D. (2016). The growth of mathematics and reading skills in segregated and diverse schools: An opportunity-propensity analysis of a national database. *Contemp. Educ. Psychol.* 46, 34–51. doi: 10.1016/j.cedpsych.2016.04.002

Cheng, P.-Y., and Chiou, W.-B. (2010). Achievement, attributions, self-efficacy, and goal setting by accounting undergraduates. *Psychol. Rep.* 106, 54–64. doi: 10.2466/PR0.106.1.54-64

Clark, J. E. (2021). *Case Study of Growth Mindset and the Impact on Math Anxiety*. (Doctoral Thesis), Brenau University, Gainesville, GA, United States.

Cohen, L. D., and Rubinsten, O. (2017). Mothers, intrinsic math motivation, arithmetic skills, and math anxiety in elementary school. *Front. Psychol.* 8, 1939. doi: 10.3389/fpsyg.2017.01939

Costa, A., and Faria, L. (2018). Implicit theories of intelligence and academic achievement: A meta-analytic review. *Front. Psychol.* 9, 829. doi: 10.3389/fpsyg.2018.00829

Davis, J. L., Burnette, J. L., Allison, S. T., and Stone, H. (2010). Against the odds: Academic underdogs benefit from incremental theories. *Soc. Psychol. Educ.* 14, 331–346. doi: 10.1007/s11218-010-9147-6

De Castella, K., and Byrne, D. (2015). My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement. *Eur. J. Psychol. Educ.* 30, 245–267. doi: 10.1007/s10212-015-0244-y

Degol, J. L., Wang, M.-T., Zhang, Y., and Allerton, J. (2017). Do growth mindsets in math benefit females? Identifying pathways between gender, mindset, and motivation. *J. Youth Adolesc.* 47, 976–990. doi: 10.1007/s10964-017-0739-8

Dirzyte, A., Vijaikis, A., Perminas, A., and Rimasiute-Knabikiene, R. (2021). Associations between depression, anxiety, fatigue, and learning motivating factors in e-learning-based computer programming education. *Int. J. Env. Res. Pub. He.* 18, 9158. doi: 10.3390/ijerph18179158

Dong, L., and Kang, Y. (2022). Cultural differences in mindset beliefs regarding mathematics learning. *Curr. Opin. Behav. Sci.* 46, 101159. doi: 10.1016/j.cobeha.2022.101159

Dowker, A., Sarkar, A., and Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Front. Psychol.* 7, 508. doi: 10.3389/fpsyg.2016.00508

Dweck, C. S. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press.

Dweck, C. S. (2002). "The development of ability conceptions," in *The Development of Achievement Motivation*, eds A. Wigfield and J. Eccles (New York, NY: Academic Press), 57–88. doi: 10.1016/B978-012750053-9/50005-X

Dweck, C. S., Chiu, C., and Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychol. Inq.* 6, 267–285. doi: 10.1207/s15327965pli0604_1

Dweck, C. S., and Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychol. Rev.* 95, 256–273. doi: 10.1037/0033-295X.95.2.256

Dweck, C. S., and Licht, B. G. (1980). "Learned helplessness and intellectual achievement," in *Human Helplessness*, eds J. Garber and M.E.P. Seligman (New York, NY: Academic Press), 197–221.

Dweck, C. S., and Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspect. Psychol. Sci.* 14, 174569161880416. doi: 10.1177/1745691618804166

Essau, C. A., Leung, P. W. L., Conradt, J., Cheng, H., and Wong, T. (2008). Anxiety symptoms in Chinese and German adolescents: Their relationship with early learning experiences, perfectionism, and learning motivation. *Depr. Anxiety* 25, 801–810. doi: 10.1002/da.20334

Geary, D. C., Hoard, M. K., Nugent, L., and Scofield, J. E. (2021). In-class attention, spatial ability, and mathematics anxiety predict across-grade gains in adolescents' mathematics achievement. *J. Educ. Psychol.* 113, 754–769. doi: 10.1037/edu00 00487

Gonida, E., Kiosseoglou, G., and Leondari, A. (2006). Implicit theories of intelligence, perceived academic competence, and school achievement: Testing alternative models. *Am. J. Psychol.* 119, 223–238. doi: 10.2307/20 445336

Gottfried, A. E. (2019). "Academic intrinsic motivation: Theory, assessment, and longitudinal research," in *Advances in Motivation Science*, ed A. J. Elliot (Cambridge: Academia Press), 71–109. doi: 10.1016/bs.adms.2018.11.001

Hackett, G., and Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *J. Res. Math. Educ.* 20, 261–273. doi: 10.2307/749515

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* 21, 33–46. doi: 10.2307/749455

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M.-S., and Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *J. Pers. Soc. Psychol.* 77, 588–599. doi: 10.1037/0022-3514.77.3.588

Huang, X., Zhang, J., and Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: the gender moderating effect. *Eur. J. Psychol. Educ.* 34, 621–640. doi: 10.1007/s10212-018-0403-z

Hunsley, J. (1987). Cognitive processes in mathematics anxiety and test anxiety: The role of appraisals, internal dialogue, and attributions. *J. Educ. Psychol.* 79, 388–392. doi: 10.1037/0022-0663.79.4.388

Jones, B. D., Wilkins, J. L. M., Long, M. H., and Wang, F. (2011). Testing a motivational model of achievement: How students' mathematical beliefs and interests are related to their achievement. *Eur. J. Psychol. Educ.* 27, 1–20. doi: 10.1007/s10212-011-0062-9

Karlen, Y., Suter, F., Hirt, C., and Maag Merki, K. (2019). The role of implicit theories in students' grit, achievement goals, intrinsic and extrinsic motivation, and achievement in the context of a long-term challenging task. *Learn. Individ. Differ.* 74, 101757. doi: 10.1016/j.lindif.2019.101757

Kaskens, J., Segers, E., Goei, S. L., van Luit, J. E. H., and Verhoeven, L. (2020). Impact of Children's math self-concept, math self-efficacy, math anxiety, and teacher competencies on math development. *Teach. Teach. Educ.* 94, 103096. doi: 10.1016/j.tate.2020.103096

Komarraju, M., and Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learn. Individ. Differ.* 25, 67–72. doi: 10.1016/j.lindif.2013.01.005

Leondari, A., and Gialamas, V. (2002). Implicit theories, goal orientations, and perceived competence: Impact on students' achievement behavior. *Psychol. Sch.* 39, 279–291. doi: 10.1002/pits.10035

Li, Q., Cho, H., Cosso, J., and Maeda, Y. (2021). Relations between students' mathematics anxiety and motivation to learn mathematics: A meta-analysis. *Educ. Psychol. Rev.* 33, 1017–1049. doi: 10.1007/s10648-020-0 9589-z

Li, Y., and Bates, T. C. (2019). You can't change your basic ability, but you work at things, and that's how we get hard things done: Testing the role of growth mindset on response to setbacks, educational attainment, and cognitive ability. *J. Exp. Psychol.* 148, 1640–1655. doi: 10.1037/xge0000669

Martin, A. J., Bostwick, K., Collie, R. J., and Tarbetsky, A. L. (2017). "Implicit theories of intelligence," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill and T. Shackelford (New York, NY: Springer), 1. doi: 10.1007/978-3-319-28099-8_980-1

Molden, D. C., and Dweck, C. S. (2000). "Meaning and motivation," in *Intrinsic and Extrinsic Motivation. The Search for Optimal Motivation and Performance*, eds C. Sansone and J. M. Harackiewicz (San Diego, CA: Academic Press), 131–159. doi: 10.1016/B978-012619070-0/50028-3

Murayama, K., Pekrun, R., Lichtenfeld, S., and vom Hofe, R. (2012). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Dev.* 84, 1475–1490. doi: 10.1111/cdev.12036

Organisation for Economic Co-operation and Development (2010). *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I).* Paris: OECD Publishing.

Organisation for Economic Co-operation and Development and Programme for International Student Assessment (2019). *PISA 2018 results (Volume III): What School Life Means for Students' Lives.* Paris: OECD Publishing.

Pajares, F., and Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemp. Educ. Psychol.* 20, 426–443. doi: 10.1006/ceps.1995.1029

PISA 2009 Shanghai Committee (2016). *Quality and Equity: Research report of the Programmer for International Student Assessment 2009 in Shanghai [in Chinese].* Shanghai: Shanghai Educational Publishing House.

Priess-Groben, H. A., and Hyde, J. S. (2016). Implicit theories, expectancies, and values predict mathematics motivation and behavior across high school and college. *J. Youth Adolesc.* 46, 1318–1332. doi: 10.1007/s10964-016-0579-y

Radišić, J., Videnović, M., and Baucal, A. (2014). Math anxiety—contributing school and individual level factors. *Eur. J. Psychol. Educ.* 30, 1–20. doi: 10.1007/s10212-014-0224-7

Ramirez, G., Shaw, S. T., and Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educ. Psychol.* 53, 145–164. doi: 10.1080/00461520.2018.1447384

Richardson, F. C., and Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *J. Couns. Psychol.* 19, 551–554. doi: 10.1037/h0033456

Robins, R. W., and Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self Identity* 1, 313–336. doi: 10.1080/15298860290106805

Ryan, R. M., and Deci, E. L. (2009). "Promoting self-determined school engagement: Motivation, learning and well-being," in *Handbook of Motivation at School*, eds K. R. Wentzel and A. Wigfield (New York, NY: Taylor Francis), 171–195.

Samuel, T. S., and Warner, J. (2019). 'I Can Math!': Reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Comm. Coll. J. Res. Prac.* 45, 205–222. doi: 10.1080/10668926.2019.1666063

Schleicher, A. (2019). *PISA 2018 Insights and Interpretations*. OECD. Available online at: https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf (accessed August 24, 2022).

Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *J. Educ. Psychol.* 73, 93–105. doi: 10.1037/0022-0663.73.1.93

Siegle, D., Rubenstein, L. D. V., Pollard, E., and Romey, E. (2009). Exploring the relationship of college freshmen honors students' effort and ability attribution, interest, and implicit theory of intelligence with perceived ability. *Gifted Child Quart.* 54, 92–101. doi: 10.1177/0016986209355975

Silver, W. S., Mitchell, T. R., and Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organ. Behav. Hum. Dec.* 62, 286–299. doi: 10.1006/obhd.1995.1051

Smiley, P. A., Buttitta, K. V., Chung, S. Y., Dubon, V. X., and Chang, L. K. (2016). Mediation models of implicit theories and achievement goals predict planning and withdrawal after failure. *Motiv. Emot.* 40, 878–894. doi: 10.1007/s11031-016-9575-5

Smith, T. F., and Capuzzi, G. (2019). Using a mindset intervention to reduce anxiety in the statistics classroom. *Psychol. Learn. Teach.* 18, 326–336. doi: 10.1177/1475725719836641

Su, A., Wan, S., He, W., and Dong, L. (2021). Effect of intelligence mindsets on math achievement for chinese primary school students: Math self-efficacy and failure beliefs as mediators. *Front. Psychol* 12, 640439. doi: 10.3389/fpsyg.2021.640349

Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., et al. (2014). A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemp. Educ. Psychol.* 39, 342–358. doi: 10.1016/j.cedpsych.2014.08.002

Tempelaar, D. T., Rienties, B., Giesbers, B., and Gijselaers, W. H. (2014). The pivotal role of effort beliefs in mediating implicit theories of intelligence and achievement goals and academic motivations. *Soc. Psychol. Educ.* 18, 101–120. doi: 10.1007/s11218-014-9281-7

Todor, I. (2014). Investigating 'the old stereotype' about boys/girls and mathematics: Gender differences in implicit theory of intelligence and mathematics self-efficacy beliefs. *Procd. Soc. Behv.* 159, 319–323. doi: 10.1016/j.sbspro.2014.12.380

van Aalderen-Smeets, S. I., Walma van der Molen, J. H., and Xenidou-Dervou, I. (2018). Implicit STEM ability beliefs predict secondary school students' STEM self-efficacy beliefs and their intention to opt for a STEM field career. *J. Res. Sci. Teach.* 56, 465–485. doi: 10.1002/tea.21506

Wang, Y., Peng, H., Huang, R., Hou, Y., and Wang, J. (2008). Characteristics of distance learners: research on relationships of learning motivation, learning strategy, self-efficacy, attribution and learning results. *J. Open Dist. e-Learn.* 23, 17–28. doi: 10.1080/02680510701815277

Wang, Z., Lukowski, S. L., Hart, S. A., Lyons, I. M., Thompson, L. A., Kovas, Y., et al. (2015). Is math anxiety always bad for math learning? The role of math motivation. *Psychol. Sci.* 26, 1863–1876. doi: 10.1177/0956797615602471

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educ. Psychol.* 45, 28–36. doi: 10.1080/00461520903433596

Wigfield, A., Tonks, S., and Klauda, S. L. (2009). "Expectancy-value theory," in *Handbook of Motivation at School*, eds K. R. Wentzel and A. Wigfield (New York, NY: Taylor Francis), 55–75.

Xu, X., Zhang, Q., Sun, J., and Wei, Y. (2022). A bibliometric review on latent topics and research trends in the growth mindset literature for mathematics education. *Front. Psychol.* 13, 7496. doi: 10.3389/fpsyg.2022.1039761

Yeager, D. S., and Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educ. Psychol.* 47, 302–314. doi: 10.1080/00461520.2012.722805

Yeager, D. S., and Dweck, C. S. (2020). What can be learned from growth mindset controversies? *Am. Psychol.* 75, 1269–1284. doi: 10.1037/amp0000794

Young, A. J., and Urdan, T. C. (1993). "The influence of perceived classroom goals and prior beliefs on aspects of student motivation," in *Poster presented in the Annual Meeting of the American Educational Research Association, Atlanta, GA*. Available online at: https://files.eric.ed.gov/fulltext/ED366622.pdf (accessed January 20, 2023).

Young, E. S., and Dyess, S. R. (2021). Supporting prospective teachers in problem solving: Incorporating mindset messaging to overcome math anxiety. *Math. Teach. Educ.* 10, 9–28. doi: 10.5951/MTE.2020-0047

Zhang, Y., Liu, S., Yang, L., Feng, T., and Yang, X. (2022). Coparenting matters: The mediating effect of implicit theories of intelligence and depression on the relation between coparental conflict and STEM self-efficacy among Chinese adolescents. *Curr. Psychol.* 41, 6687–6699. doi: 10.1007/s12144-022-03118-7

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemp. Educ. Psychol.* 25, 82–91. doi: 10.1006/ceps.1999.1016

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 06 April 2023
DOI 10.3389/fpsyg.2023.1061027

OPEN ACCESS

EDITED BY
Luis J. Fuentes,
University of Murcia, Spain

REVIEWED BY
Changgen Pei,
Southwest University, China
Milagros Elena Rodriguez,
Universidad de Oriente, Venezuela
Viacheslav Osadchyi,
Borys Grinchenko Kyiv University, Ukraine

*CORRESPONDENCE
Lidong Wang
✉ wldyaoyuan@163.com

SPECIALTY SECTION
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

RECEIVED 04 October 2022
ACCEPTED 06 March 2023
PUBLISHED 06 April 2023

CITATION
Zuo H and Wang L (2023) The influences of mindfulness on high-stakes mathematics test achievement of middle school students. *Front. Psychol.* 14:1061027.
doi: 10.3389/fpsyg.2023.1061027

The influences of mindfulness on high-stakes mathematics test achievement of middle school students

Haode Zuo[1] and Lidong Wang[2]*

[1]College of Mathematical Science, Yangzhou University, Yangzhou, China, [2]Collaborative Innovation Center of Assessment of Basic Education Quality, Beijing Normal University, Beijing, China

Research has shown that mindfulness can reduce students' negative emotions associated with high-stakes tests and thereby improve test performance. This study explored the association between mindfulness-based intervention (MBI) and high-risk math test scores of middle school students, which is noticeably slim in the domain of mathematics education, through a mediating process involving math-specific test anxiety and math self-efficacy. Using data from a sample of 45 students, age 12–13, we found empirical support for a significant positive correlation between mindfulness and middle school students' math achievement. Participants listened to a mindfulness audio every other week before a mathematics test. Weekly mathematics test scores, student group discussion data, and in-depth interview data were analyzed to explore how mindfulness affected students' mathematics test performance, which showed a statistically significant improvement after mindfulness compared to mathematics achievement without the intervention. Our results indicate that mindfulness can relieve mathematics anxiety symptoms, including physiological manifestations, test-unrelated thinking and worries, and problem-solving obstacles caused by mathematics anxiety. Also, mindfulness, especially its non-judgmental attitude, positively affects students' mathematical self-efficacy. The current research provides evidence of the mindfulness intervention's efficacy for improving middle school mathematics test performance but also identifies the complexities of implementing it with large numbers of students.

KEYWORDS

mindfulness, meditation, high-stakes test, test anxiety, mathematics anxiety, mathematics test anxiety, mathematics self-efficacy, mathematics achievement

1. Introduction

Middle school mathematics tests are high stakes because they directly impact students, teachers, and schools (Reback et al., 2014). In many countries, mathematics academic performance is directly linked with students' graduation and continued study (Ho, 2009; Stobart and Eggen, 2012; Suprapto, 2016), which is especially significant in the test-oriented education environment of East-Asian countries. As a result, students in East-Asian countries are probably more anxious to do well in high-stakes mathematics tests (Tan and Yates, 2011). Current meta-analyses have shown that high-stakes tests are a kind of high-cognitive-demand competition, assessing students' cognitive competencies and ability to cope with anxiety in high-stakes testing situations (Burke, 2010; Keng et al., 2011). High-stakes mathematics tests often create a high anxiety situation for students, with mathematics anxiety playing a particularly large role (Bellinger et al., 2015). Empirical studies (Ashcraft and Moore, 2009) show that 17% of the population has high mathematics anxiety levels, and 33% of 15-year-old students in 65 countries who participated in the 2012 Program for International Student Assessment (PISA) reported feeling anxious when solving mathematics problems (OECD, 2013).

Moreover, improperly addressed mathematics anxiety will lead students into a vicious circle, where high-stakes mathematics tests cause mathematics anxiety, which causes further difficulties with mathematics (Maloney and Beilock, 2012). A comprehensive review found converging evidence of higher mathematics achievement usually accompanies a reduction in mathematics anxiety (Hembree, 1988). Further, a large number of empirical studies show that students with moderate mathematical self-efficacy often have a correct evaluation of their mathematics ability, but those with too high or too low mathematical self-efficacy will wrongly estimate their capacity, leading to excessive mathematics anxiety (Betz, 1978; Cooper and Robinson, 1991; Pajares and Urdan, 1996).

Recognizing middle school students' mathematics anxiety and improper sense of self-efficacy related to high-stakes tests, we posit that introducing mindfulness, a mode of attending to present moment experiences without judgment or elaboration (Bishop et al., 2006), may help. Previous work has suggested that mindfulness effectively improves academic performance by supporting students' ability to stay psychological well-being in the context of high cognitive demand (Napoli et al., 2005; Shapiro et al., 2011; Meiklejohn et al., 2012). Individuals tend to be more anxious in situations requiring high cognitive demand (Sarason, 1984; Schwarzer and Quast, 1985), especially high-stakes test situations. For example, when taking a high-stakes test, students' mathematics anxiety will give rise to test-irrelevant thoughts, e.g., about how hard the math test is or the consequences of failing the test. A growing body of evidence (Bellinger et al., 2015; de Abreu Costa et al., 2019) suggests that mindfulness provides a means to cope with and effectively buffer against the deleterious effects of anxiety. Additionally, robust research evidence indicates that mindfulness positively affects self-efficacy (Vidic and Cherup, 2019) and is the most significant determinant contributing to it (Chan et al., 2021).

Mathematics is assumed to elicit more substantial anxiety than most other academic subjects, such as literacy (Punaro and Reeve, 2012). However, there remains a lack of a microscopic examination of mindfulness's effects on improving mathematics performance in natural high-stakes testing environments. Thus, the present research investigated the impact of mindfulness within the context of a high-stakes mathematics test. Few studies have examined associations between mindfulness and students' high-stakes mathematics test scores, and even fewer have investigated how mindfulness-based intervention (MBI) affects students' performance. This study was the first to explore these two questions.

2. Theoretical background

2.1. High-stakes tests

Testing is considered high stakes when test results are referenced in important decisions affecting students, teachers, administrators, communities, schools, and districts (Haladyna et al., 1991). For example, high-stakes test scores affect students' graduation and promotion (Amrein and Berliner, 2002) and, in some cases, teachers' and principals' salaries and tenure (Council, 1998), as teachers and principals are often held accountable for students' high-stakes test performance (Jones et al., 2003). The effects of high-stakes tests on students differ across countries and cultures, but play a considerable role in East-Asian education, a test-oriented culture (Leung, 2001). For example, mainland China emphasizes that screening students' talents is an important national task. Historically, China's talent selection began during the Spring and Autumn Period, as self-recommendation. However, the country's destruction (despite many followers, such as the Four Gentlemen of the Warring States Period) showed this method was not ideal. The Han Dynasty initiated the approach of recommending people based on their filial piety and moral records, and the talented people selection system of Jin Dynasty revolves around nepotism and cronyism. Such methods had disadvantages, such as inconsistent reference standards, huge costs, and long talent training cycles. They also limited the generation of talents for the aristocracy, seriously aggravating social stratification and leading to reduced social mobility and other inequalities. The earliest high-stakes tests appeared in the imperial examination system established during the Sui and Tang Dynasties. For thousands of years, the advantages of high-stakes tests—such as low cost, high efficiency, high universality, and a high degree of standardization—were vividly demonstrated in the imperial examination and its evolution into various levels of entrance examinations. Of course, it had obvious drawbacks: high-stakes exams focus on measuring results, and such a strong sense of purpose goes hand in hand with a strong sense of crisis.

2.2. Math-specific test anxiety

Anxiety is an overarching construct, generally conceptualized as a state of emotion underpinned by qualities of fear and dread (Lewis, 1970). Mathematics anxiety has become increasingly prevalent in the past six decades, and its negative consequences for mathematics learning are well-documented (Dowker et al., 2016).

Mathematics anxiety, defined as a feeling of apprehension and fear related to mathematics (Ashcraft, 2002), is closely related to students' negative attitudes toward mathematics. First, mathematics anxiety comprises different components, often termed "cognitive" and "emotional." Researchers have argued that mathematics anxiety's emotional dimension is more strongly negatively correlated with mathematics achievement than its cognitive dimension (Wigfield and Judith, 1988). Specific to high-stakes test situations, scholars argue that mathematics anxiety seems to be an aspect of "situational specific psychological distress" rather than an aspect of "mathematics" (Hembree, 1988). Neuropsychological data (Sheffield and Hunt, 2006) suggest that students' hearts may beat more quickly or firmly in high-stakes academic testing environments. Therefore, the relationship between mathematics anxiety and academic performance in a high-stakes testing environment needs careful consideration. Secondly, mathematics anxiety has various negative consequences, one of the most obvious being a decline or failure in academic achievement. A series of studies (Wu et al., 2012; Vukovic et al., 2013; Ching, 2017) has confirmed the negative link between mathematics anxiety and mathematics performance. Moreover, this negative correlation is particularly prominent among East-Asian students (Lin and Chen, 1995; Kirkpatrick and Zang, 2011).

Finally, Li (2001) found that individuals' thoughts when dealing with high-stakes mathematics exams commonly center on (a) the consequences of not performing satisfactorily and awkwardness when the actual situation does not meet their expectations, (b) feelings of inferiority and self-blame when they see other students' progress faster, and (c) clinging to familiar problem-solving methods. All these thoughts and behaviors, which are useless or harmful to solving test problems, can be concluded as thoughts unrelated to the current task and separated from the external environment (Stawarczyk et al., 2011).

Recent meta-analyses (Zhang et al., 2019; Barroso et al., 2021) provided findings to support that mathematics anxiety is largely in the form of math-specific test anxiety. Poor performance in a math test would lead to higher mathematics anxiety, whereas individuals who have higher levels of mathematics anxiety frequently get poor math performance. Overall, we believe the existing negative correlation between mathematics anxiety and mathematics performance is mainly talking in ways of math-specific test anxiety. In addition, as pointed out by empirical work (Bellinger et al., 2015), mindfulness might reduce specific manifestations of psychological problems raised by high-stakes testing situations. Thus, this paper explored mathematics anxiety in a high-stakes test context—math-specific test anxiety—which refers to abnormal mental experiences caused by excessive anxiety in mathematics test situations, especially high-risk mathematics test situations.

2.3. Mathematics self-efficacy

According to Bandura (2010), self-efficacy refers to the subjective conviction that one can successfully execute the behavior required to attain a desired outcome. He argued that people with high self-efficacy appraise their capabilities more favorably than those with low task self-efficacy, and this favorable self-appraisal leads them to outperform those with low self-efficacy.

One such kind of self-efficacy is mathematics self-efficacy, which refers to an individual's confidence in their ability to successfully perform or accomplish a specific mathematical task or problem (Betz and Hackett, 1983). Examples of mathematics self-efficacy include "I like to challenge mathematical problems," "I hope mathematics can play an important role in my future work," "When facing mathematical problems, I have the confidence and ability to solve them," and "I am praised by my classmates and teachers because of mathematics," etc. Studies support the theory that self-efficacy judgments are not mere reflections of past performance but thoughts about one's current situation or performance during the test (Schunk, 1982). Unsurprisingly, mathematics self-efficacy is negatively related to mathematics anxiety (Betz, 1978; Cooper and Robinson, 1991; Pajares and Urdan, 1996). First, a consistent finding of mathematics self-efficacy associated with mathematics anxiety is that individuals who experience mathematics anxiety express more negative attitudes about mathematics (Ashcraft, 2002), suggesting they lack confidence in their abilities. Further, Cassady and Johnson (2002) found that mathematics test anxiety contributes to increased worries and negative self-criticism, thus failing in high-stakes academic situations.

Unfortunately, prior research has taken a primarily static view of the relationship between mathematics anxiety and mathematics self-efficacy, without addressing the potential impact of change in external variables. Therefore, this study fills an existing research gap by exploring the interactive effects of mathematics self-efficacy and mathematics anxiety on students' mathematics achievement in high-stakes tests with the presence of the external variable, mindfulness.

2.4. Mindfulness

Beginning in the 1970s, international psychology and education scholars began to pay increased attention to mindfulness, leading to several definitions, such as "the awareness that emerges through paying attention on purpose, in the present moment, and non-judgmentally to the unfolding of experience moment by moment" (Kabat-Zinn, 2010, p. 145), and "a process of regulating attention in order to bring a quality of non-elaborative awareness to current experience and a quality of relating to one's experience within an orientation of curiosity, experiential openness, and acceptance" (Bishop et al., 2006, p. 234). The clinical psychology application of mindfulness has been altered and extended to many different fields through a variety of mindfulness courses, such as mindfulness workplace coaching (Ancona and Mendelson, 2014), childbirth and parenting (Duncan and Bardacke, 2010), competitive sports coaching (Scott-Hamilton et al., 2016), etc.

Of all possible applications, school-based mindfulness that combines students' psychosomatic characteristics and traditional school mental health curricula has received the most attention (Yu and Wenjie, 2017). School-based mindfulness has been widely used in recent years to help a growing number of students suffering from mental health problems, such as anxiety (Essau et al., 2011), depression (Jellinek and Snyder, 1998), and social withdrawal (Hipson and Coplan, 2017). Previous studies have confirmed the value of mindfulness applications for improving schoolwork. British and American scholars have found that mindfulness can alleviate students' depressive symptoms (Kuyken et al., 2013) and relieve stress (Wall, 2005) to a certain extent. Moreover, East-Asian scholars report that mindfulness improves students' concentration and performance in memory and other cognitive tasks (Ching et al., 2015; Lam, 2016). Likewise, scholars in mainland China have found that mindfulness training can improve middle school students' self-control (Changyu and Xiao, 2016), help relieve exam phobia, and prevent mental wandering (Shanshan and Zhun, 2018). As such, the common value of mindfulness applications across different fields is that they reduce the possibility of being disturbed by emotional factors in uncertain, painful, and anxious situations, such as job performance, pregnancy and childbirth, and high-intensity competition, thereby improving performance.

Most mindfulness therapies are categorized into two major branches based on their functional differences: mindfulness-based stress reduction (MBSR), which is used to help patients relieve anxiety, depression, and other emotional and psychological problems through meditation (Kabat-Zinn, 2003), and mindfulness-based cognitive therapy (MBCT). MBSR is a safe and effective treatment for reducing emotional deregulation, especially anxiety. Studies have found that MBSR outperformed in a group of individuals with a generalized anxiety disorder, suggesting it is superior in reducing anxiety symptoms (Evans et al., 2008; Vøllestad et al., 2011; Hoge et al., 2013). Recent studies have also examined the effects of MBSR intervention on positive states of mind and mindfulness self-efficacy, suggesting that mindfulness self-efficacy and positive states of mind are significantly higher after intervention (Chang et al., 2010). Similarly, MBCT has also been shown to effectively relieve anxiety (Chiesa and Serretti, 2011; Hofmann and Gòmez, 2017). Hence, mindfulness-based therapies are generally effective in reducing anxiety, stress, and depressive symptoms in adults (Baer, 2010) and children (Semple et al., 2010).

In addition, empirical studies and meta-analyses have found that mindfulness is the most significant determinant contributing to self-efficacy (Chan et al., 2021). In particular, Zeljka Vidic and colleagues found support for the impact of mindfulness on self-efficacy (Vidic and Cherup, 2019), indicating mindfulness can reduce test anxiety by improving students' sense of self-efficacy. Other findings suggest that a brief mindfulness intervention is an effective and practical means of enhancing academic self-efficacy and emotional well-being in university students, and a reliable and valid self-efficacy measure for mindfulness meditation practice has been developed (Birdee et al., 2020).

However, most school-based mindfulness studies rely on questionnaires (mainly self-reported student assessments) to assess the effects; few studies have measured the impact based on objective academic performance data. In particular, a common problem with questionnaires or self-report measures is that they may be affected by the accuracy and truthfulness of students' self-perceptions. Additionally, when students complete self-reports or questionnaires, they are not in an uncertain, painful, or anxious situation, limiting the research conclusions' validity. Therefore, studies are needed that obtain objective measures to assess the influence of mindfulness on student performance. Moreover, there has been little research on the efficacy of MBSR and MBCT interventions in high-stakes mathematics tests. We address this absence by using mindfulness interventions to ease students' math-specific test anxiety during high-stakes tests. Our research findings help explain how MBI enables students to get rid of math-specific test anxiety and enhance self-efficacy during testing, thus improving their academic performance.

2.5. The present study

As noted earlier, the current study's primary aim is to explore the process linking mindfulness to mathematics achievement in a high-stakes testing environment. Drawing on the extant literature (Franco et al., 2010; Bellinger et al., 2015; Samuel and Warner, 2021; Leppma and Darrah, 2022), we propose a theoretical framework that mindfulness interventions are beneficial for students' mathematics achievements in high-stakes exams situations by moderating the degree to which students were anxious about high-stakes mathematics tests and regulating mathematics self-efficacy to a normal level. The hypotheses for the present research study are:

1. Math-specific test anxiety will decrease in students receiving the mindfulness intervention.
2. Math self-efficacy (MSE) will increase in students receiving the mindfulness intervention.
3. The mindfulness intervention will improve students' high-stakes mathematics test achievements by reducing their math-specific test anxiety and increasing their math self-efficacy.

Two main research questions are addressed:

1. What is the relationship between mindfulness and students' high-stakes mathematics test achievements?
2. What are the underlying and intervening mechanisms in mindfulness and students' high-stakes mathematics test achievement relationships?

3. Research methods

3.1. Research participants

The stress resulting from exam crises is most apparent in middle school. At the primary school level, young students have not yet faced significant academic pressure; at the university stage, adult students have a relatively mature psychological bearing, examinations are typically not as intense, and there are opportunities to repeat exams after a poor performance. Therefore, the study was designed as a quasi-experiment, taking one natural class (45 students, ages 12–13 years, 30 boys and 15 girls) of eighth-graders in a junior middle school, conveniently selected from central Jiangsu province (China), as the subjects. As the sampled school is located in an urban area, participants largely identified as town dwellers. Consent was obtained from all participants, including their parents and relevant school personnel, before the study. The participants were similar to those students taking part in large-scale international comparative academic performance studies, such as The Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS). According to PISA 2018 (Schleicher, 2019), students in Jiangsu scored higher than the OECD average in mathematics. In addition, the overall mathematics level in the sampled class was relatively stable, and the scores of the recent two mathematics tests had not significantly fluctuated.

3.2. Research design

As students' academic performance followed (or approximately followed) a normal distribution, this study combined the characteristics of random variables subject to normal distribution. The G-power toolbox was used to calculate the required sample size to ensure sufficient statistical power. The class of 45 students was randomly chosen from a middle school, meeting G-power's sample size estimate (effect size = 0.5, 1 - α = 0.05, 1 - β = 0.95).

TABLE 1 Equivalent-time-sample design.

First cycle	Second cycle
Math scores without mindfulness X_0O_1	Math scores without mindfulness X_0O_3
Math scores with mindfulness XO_2	Math scores with mindfulness XO_4

Applying the equivalent-time-sampling method shown in Table 1, quantitative survey and objective data were used to test research hypotheses. The experiment lasted 4 weeks and was divided into the experimental period (X) and control period (X_0), each lasting 2 weeks. Data were collected weekly (i.e., four math test scores in 4 weeks (O_1, O_2, O_3, O_4). Students were not given mindfulness (X_0) before the exam in the first and third weeks (i.e., the students were not guided to any mindfulness practice); in the second and fourth weeks, the same students received the intervention (X) before the exam (i.e., the students were guided to practice mindfulness by listening to a recorded audio of mindfulness).

3.3. Research instruments

3.3.1. Mindfulness-based intervention

The intervention involved students sitting in the classroom as a group before the test, guided by audio mindfulness. None of the many ready-made versions of mindfulness audio is custom-made for a math test. The researchers re-contextualized the concept, putting it into the framework of student psychological problems that may occur during a stressful high-stakes math examination. As discussed above, as MBSR and MBCT are commonly-used means of MBI, their operational definitions (Williams et al., 2007) were highlighted to ensure the essential connotations of mindfulness—intentional focus, non-judgemental attitude, and the present moment—were not lost. The first author acted as the MBI instructor. Having completed an online mindfulness training practicum hosted by Dr. Huiqi Tong, a teaching partner and chief Chinese translator of Kabat-Zinn, founder of the "MBSR" program, he possessed a basic level of training and sufficient proficiency to deliver MBI to the students. The researchers' audio recordings lasted 15 min; evidence (Zeidan et al., 2014) has shown that short-term mindfulness intervention is also effective for subjects without prior meditation experience. The main content of MBI in this current study is summarized below: At the beginning of the session, the instructor asked students to sit in a graceful, comfortable, and conscious posture, with their bodies straight, eyes slightly closed, shoulders down, and hands relaxed. The instructor then guided the students to breathe naturally, pay attention to their breath (without consciously controlling or adjusting it), be aware of the feeling it brought to their body's organs, focus on keeping their inhalation and exhalation natural, and gradually calm their heart. When students found themselves involuntarily distracted from their breathing, the instructor would tell them: *"Don't panic; you should know that distraction or mind-wandering is a natural occurrence, be aware of where it is, let go, and bring your mind back to your breathing."*

Then, the instructor directed students to turn their attention to the body and feel the sensation of breathing on all body parts. If the student felt tense or uncomfortable during this process, the instructor would tell the student, *"Make sure they're there, don't try to change them, just put them down and get your attention back on your breath."*

Next, the instructor led the students through an awareness exercise. At the beginning of the exercise, the instructor told the students: *"If you find yourself distracted, don't panic. That's normal. We can use the third person's perspective to see where our attention is going and whether it is becoming your thought. Just like our awareness of breathing, we only need to observe the emergence, change, and disappearance of thoughts. Now imagine that we are sitting in an exam room taking a math test. During the test, I would calculate how many points I got while being afraid of how many points I would lose if I made a mistake in this question. Then I began to think about how much my ranking would drop if I lost these points and what would happen after that. We all know it's unnecessary, but we can't get rid of it. Now, I'd like you to look at this thought and name it. Is it a plan, a recollection, a worry, or a fanciful image? You would find that it's a 'worry', it's our habitual way of thinking, and it's not real. After naming, the thought usually loosens, disintegrates, and disappears. Notice the naming and the disappearance of the thought. Then, bring your attention back to your breath, to the present moment."*

Through this process, the instructor used mindfulness to intervene in students' compulsive checking and being swayed by considerations of gain and loss, commonly seen when students participate in high-stakes tests. This achieved the "contextualization of mindfulness in high-stakes mathematics test situations" by using mindfulness's "intentional attention" to keep students focused on the test itself and the "non-judgemental attitude" and "present moment" in mathematics self-efficacy to tell students that worries are only thoughts, not facts, thus getting rid of expectations about how well they will do on the activity: *"For another example, when I took the test, I found that the questions were far from what I expected; it's different than usual. I began to scratch my head and lose my mind. At this point, we should also look at the thought, classify it, and discover that it's a plan, that it's not a fact, that I should turn my attention back to my breath, back to the present."*

Through this process, the instructor used mindfulness to intervene in students' common problem of fixating on difficult questions, leading to a complete loss of confidence. He used mindfulness' "intentional attention" and "present moment" to tell students that "plans" are only thoughts, not facts, and ultimately improve their sense of self-efficacy, making "re-contextualized" mindfulness easier for students to accept, allowing them to participate in the intervention process actively.

Throughout the practice, the instructor guided students to observe their thoughts in a non-judgmental manner, telling them that instead of thinking about the exam results or expecting a challenging exam, it would be better to focus on the present moment without presupposing the results to enhance self-efficacy. Whether the thought itself or the emotion it caused were pleasant or unpleasant or there were feelings attached, there was no need to dwell on the content, just watch one thought after another as they rose and disappeared, and then continue.

Finally, the instructor led the students through an observing emotions exercise. Students were asked to take three deep breaths, watching for and patiently observing the emergence of any

TABLE 2 Reliability analysis.

Cronbach's alpha	Numbers
0.864	8

TABLE 3 Difficulty and discrimination value.

Week	Difficulty	Discrimination
1	0.671	0.584
2	0.723	0.520
3	0.627	0.604
4	0.686	0.555

emotions as they breathed. Sometimes, several emotions would appear simultaneously; other times, there were no obvious feelings. The instructor told the students: *"Recognize and confirm any emotion, don't judge it but just experience the feelings it brings to your body; it's just a mental phenomenon and doesn't represent the truth. Don't try to change it. Similarly, imagine sitting in an exam room and taking a math test. During the test, I saw that other students solved problems quickly and reached the next page, but I still stayed on the previous page. I began to feel inferior and self-blamed, blindly following their speed. Now, let's look at this emotion and give it a name. Is it fear, insecurity, or loneliness? We find that it's insecurity. It's just a phenomenon of our mind and is not the same as a fact—we're not the same as the emotion. After naming, the emotion usually changes and goes away. Notice this process, and then bring the attention back to the breath and back to the present."*

Throughout this process, the instructor used mindfulness to intervene in students' common problems of self-abasement and self-reproach, using mindfulness's "intentional attention," "non-judgemental attitude," and "present moment" to tell students that insecurity is only an emotion, not a fact. After being named, the emotion would change and disappear.

At the end of the intervention, the instructor prompted students to bring their minds back to reality and try to bring a relaxed, secure, and comfortable state of mindfulness into the upcoming exam.

3.3.2. Math test

Most studies rely on questionnaire measures to assess for effects (mainly student self-report) and do not include follow-up assessments, which remains a limitation in the existing literature (Liehr and Diaz, 2010; Lau and Hue, 2011; Lagor et al., 2013; Metz et al., 2013; Felver et al., 2016). This study obtained student achievement and grades as objective data on student educational or behavioral outcomes. The experimental class's math teacher was responsible for making the test, compiling the questions, correcting the completed tests, and compiling the grades. The test consisted of free-response questions and was graded on a 10-point scale, the scoring rubric was inter-rater checked by the first author and the class's math teacher. Tests are ebbed in the participants' weekly quizzes, which are linked with their unit assessment, and may further imply their graduation assessment. And the tests are taken in the classroom, where the participants could receive peer pressure during the test-taking (Beilock et al., 2004; DeCaro et al., 2010). The math test used in this study was based on students' actual learning progress and chosen or adapted from the textbook or related exercises to ensure content validity. An analysis of student responses to each question showed a Cronbach's alpha of 0.864, indicating high internal consistency reliability and suggesting the math test had qualified and sufficient homogeneity reliability (Table 2).

To ensure that the math tests used across the 4-week experiment were parallel, the difficulty and differentiation of four sets of questions were analyzed, and their difficulty was calculated using an extreme grouping method that ranked the weekly math scores in descending order. Starting from the highest score, 27% of the total volume was sampled as the high score; starting from the lowest score, 27% of the total volume was sampled as the low score. P_H and P_L represent the difficulty value of the test for high and low group candidates, respectively, and the difficulty value of the test was expressed by $P = \frac{P_H + P_L}{2}$. The results are shown in Table 3. The difficulty of the four groups of questions was between 0.6 and 0.8. $D = P_H - P_L$ was used to represent the tests' degree of distinction. The results are shown in Table 3; the four groups' discrimination degree was above 0.4, indicating that the difficulty and differentiation of the four groups of questions were consistent. Therefore, the mathematical tests adopted in this 4-week experiment were equivalent in terms of homogeneity, equal reliability, fairness, etc.

3.3.3. Group discussion and in-depth interviews

At the end of the 4-week experiment, researchers interviewed students to understand whether they perceived the mindfulness intervention as helpful and, if so, why. A class meeting was held, attended by the 45 students. Students were given the math papers they had marked and were asked to review the mindfulness intervention's whole process and confirm their answers. Then they were asked to discuss "whether the mindful intervention had been helpful in the test" based on their test situation in the first 4 weeks. The students were divided into four groups, with one researcher chairing each group's discussion. The discussion process was recorded and included the following questions:

1. What psychological problems have been alleviated and improved after mindfulness practice?
2. What do you think the mindfulness practice did to your ability to successfully perform or accomplish a specific mathematical task or problem during the examination?
3. When you need to deal with these psychological problems again in a future exam, will you have a better way and more choices to deal with it?
4. Do you think the mindfulness practice was useless for you? If so, explain why.
5. What aspects could be improved in implementing mindfulness to make it easier for you to accept?

Each group was asked the questions in sequence; all participating students expressed their opinions.

After the group discussions, to better explore the impact and limitations of mindfulness, the researchers chose four students—two who had significantly improved and two who had not—to complete an in-depth interview on "what psychological factors

TABLE 4 Interviewee characteristics.

Students	Gender	Scores				Improvement rate in Week 4 compared to Week 1
		Week 1	Week 2	Week 3	Week 4	
A_1	Male	9	10	9	10	11.11%
B_1	Male	5	10	7	7.5	50%
A_2	Female	9	10	9	10	11.11%
B_2	Male	5	6	9	9	80%

mindfulness can influence regarding middle school students' math test scores." Interviewee characteristics are shown in Table 4.

Before the formal interview, the students were introduced to the study's purpose and promised their information would be kept confidential. Then the students were guided to review the mindfulness process and the math exam. One researcher asked questions based on the interview outline. For those who believed "mindfulness is useful to me" the main questions were about mindfulness' influence on their psychological problems under the stressful conditions of a high-stakes mathematics test and their achievement motivation during the exam. For those who believed "mindfulness is almost useless to me," the main questions concerned why they could not integrate into the mindfulness environment. The other researchers were responsible for recording students' responses.

Interviews were transcribed verbatim with no analysis cuts through three processes: extraction, coding, and classification. The second author checked the transcript for accuracy, and the researchers manually coded it. We used a two-step coding system to capture the themes and codes to describe the mechanisms transmitting mindfulness' effects on students' mathematics achievements in high-stakes testing environments, first converting informal language into academic language (i.e., researchers first read through the transcript to grasp the answers, inductively deriving codes), then coding high-frequency key nouns (i.e., "relieve," "enhance," and "promote"). Constant comparison coding was conducted in which the researchers read the texts and highlighted related information about the underlying relationships among mindfulness, math-specific test anxiety, and mathematics self-efficacy, based on the literature and the researchers' understandings of mindfulness' effects in school contexts. Next, thematic categories were constructed based on the connections identified in the open coding process and aligned with the implemented theoretical framework. The first author and two research assistants conducted the coding process independently, resolving any coding disagreements through discussion (Cohen et al., 2017). The resulting coding themes and categories are presented in Table 5.

4. Results

This study employed mixed methods, with quantitative data as the main material and qualitative data as auxiliary material, to explore the mechanism of mindfulness' influence on students' high-stakes mathematics test performance, based on the same-time-sampling design applied to a middle school sample.

4.1. Changes in students' high-stakes test scores before and after mindfulness practice

The average scores of all students from week 1 to week 4 were 6.66, 6.91, 6.80, and 7.31, respectively. The results are depicted graphically in Figure 1, where there are clear trends of increasing after receiving mindfulness intervention and a trend of decreasing after removing the intervention.

A one-sided Z-test was conducted on math score means from parallel tests in weeks 1 and 4 to examine whether mindfulness intervention could account for students' improvement in scores of tests. The results are shown in Table 6.

Differences were deemed significant when confidence limits exceeded 95% ($p < 0.05$). The results showed that the scores of d that followed mindfulness significantly improved compared to the tests' scores without the mindfulness intervention.

4.2. What mindfulness brings for math-specific test anxiety

Based on thematic analysis and synthesis of the qualitative data collected from group discussion and in-depth interviews, three broad themes were identified regarding mindfulness's influence on students' mathematics test anxiety in the high-stakes testing environment.

4.2.1. Reduces physiological manifestations of MTA

The participants believed the mindfulness practice lowered their heart rate, mainly the guided breath frequency and body scan in the audio. In the beginning, the audio guided participants to breathe based on its rhythms (*"Close your eyes and take some deep, slow breaths in and out; with every breath in, relax your body a little bit more"*), and then asked them to *"feel the surface below you, relax your face, your tongue, your neck and your arms and your fingers...while breathing out relax your back, your chest, your hips, your legs, and your feet"*. The MBI taught students to pay attention

TABLE 5 Coding themes and categories.

Math-specific test anxiety (MTA)	Mathematics self-efficacy (MSE)
Physiological manifestations:	**Past experiences dictate students' opinions**
(a) Heart rate	**concerning their personal ability in mathematics:**
(b) Tight body	(a) Comparing self-performance to peers
	(b) Overconfidence or under-confidence in test performance
Test-irrelevant thinking and worry:	
(a) Excessive consideration of the	**Mathematics self-efficacy-related problem-solving obstacles:**
consequences of failure in this math exam	(a) Comparing problem-solving speed to peers, breaking own
(b) Sorrow for their parents	time arrangement for answering the test paper
(c) Undue frustration by running out of the exam time	(b) Excessive pursuit of ingenious problem-solving methods
	(c) Habitually walk away from any challenging
	mathematics problem
Mathematics test anxiety-related problem-solving obstacles:	
(a) Clinging on to some familiar problem-solving method	
(b) Obsessing on some unsolved problems, unable to	
adjust mind in time to solve the next problem	
(c) Excessive upset when the real difficulty of the test is	
inconsistent with the anticipated difficulty	

to their bodies, making them feel their heart rate slow and assume a more relaxing posture before the exam, as some students stated in the group discussion:

> *"I think it's helpful when I listen to the audio, like, you know, it makes me feel comfortable and lowers my heart rate. And the rhythm of my breath seems to be enough to change my brain, a move from anxious, more toward wellness..."*
>
> *"...the moment in which during the body scan, I have a better awareness of my body organs. I have a better sense of my head, my neck, my shoulders, my chest... I can feel my arms by my side, feel my feet on the ground. Sometimes, I kept my shoulders shrugging when I was doing some homework. Listening to the audio, I feel my body's stretched, I can feel my arms by my side, feel my feet on the ground. In sum, I am able to find a comfortable position, I can find my body awareness, the sensory awareness in my body, I relax myself from head to foot, and it makes me feel at ease, relaxed, less anxious. I really enjoy this feeling."*

4.2.2. Reduces test-irrelevant thinking and worry

Aside from reducing the physiological manifestations of MTA, students shared that the MBI also reduced test-irrelevant thinking and worry. Analyzing the qualitative data describes three aspects of this effect:

- Reducing excessive consideration of the consequences of failure in these math exams
- Relieving sorrow for their parents
- Reducing undue frustration by running out of the exam time.

First, during the exam, students were obsessive about attaining a correct outcome and the consequences of failing the test. The mindfulness audio guided them to reduce MTA by refocusing on the pencil in their hand rather than worrying and obsessing about outcomes and consequences. The participants mentioned how test-irrelevant thinking and worrying might be reduced by the MBI:

> *"I often go down the alarming path of thinking. I just cannot get rid of 'what happens if?' What happens if I have miscalculated? What happens if my teacher penalizes me for a proof mistake or a misplaced symbol? The audio describes it as that our lives are continuous exceptions of ups and downs, twists and turns of agreeable and disagreeable situations. I noticed that lives are based on the present moment. So my focus should be on making the present moment the best we can. I gradually free myself from these disturbing and worrying thoughts, I find a kind of concentration arise, and I am not that anxious."*
>
> *"Every time I hear my mother's relatives say some other children are good at math and look proud, I wonder when I can make my parents so happy. And every time, it seems the more I want to satisfy my parents by performing well on a math test, the more disappointed my parents may get. I already feel dumb instead of anxious about this... however, after receiving the MBI, I realize that being in the past ruminating or regretting events is not helpful at all. Hence, I set aside thoughts of sorrow for my parents, and I pay attention to the test moment in an open, accepting way to what I am experiencing."*
>
> *"I am kind of a slowpoke person. When I was in grade one or two, it was not until ten in the evening before I could finish writing my homework. After entering junior high school, it is increasingly common for me to be unable to complete the math*

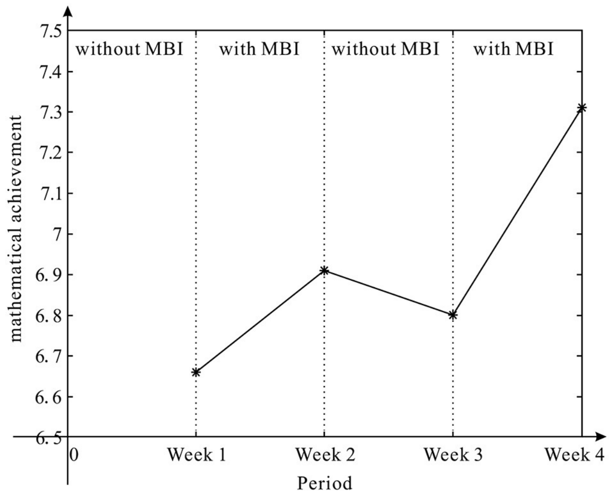


FIGURE 1
The average score of students in math test.

TABLE 6 Mean Z-test for math scores in weeks 1 and 4 (α = 0.05).

Mean		Mean		95%CI	
Week 1	Week 4	Week 1	Week 4	p	Cohen's d
6.66	7.31	[5.95,7.36]	[6.60,8.02]	0.035	0.27

paper in regular time. This situation almost made me in a vicious circle. The more frustrated I am by running the exam time, the less I can concentrate on problem-solving. With the help of the MBI, instead of frustrating for not having enough time, I started to spend longer concentrating on the present problem-solving and became more focused and engaged. Finally, to my surprise, I did finish the exam paper before the bell rang. It comes to me that maybe because if I always think about the due time, my focus is constantly on the future, and not on the work that needs to be done at that very moment."

4.2.3. Reduces MTA-related problem-solving obstacles

MBI can also be seen as essential in helping students surmount mathematics test anxiety-related problem-solving obstacles. Participants said they thought that MBI could stop them from:

1. clinging to familiar problem-solving methods;
2. obsessing over unsolved problems and help them adjust their minds in time to solve the next problem;
3. being excessively upset when the test is more difficult than anticipated.

B_1 and B_2 all particularly mentioned that these mindfulness exercises are helpful in mathematics problem-solving:

> B_1: *"In solving the problem 'Given the height and distance of two straight rods, the distance between the top of the rod and a certain point is equal, find the horizontal distance between the bottom of the rod and that point,' I no longer had the thought of clinging on to some familiar problem-solving methods, for example, proving the congruence of two triangles. I told myself that I should focus my attention on the present problem, not on the previous problem or not on congruent triangles on seeing line segments equal. And then, when I got down to it, I realized that I could solve this with the Pythagorean Theorem... 'Given the perpendicular bisector of the hypotenuse of a right triangle, find the distance from its intersection to the vertex of the right angle'; this problem, which we rarely encountered in our daily study, belongs to the new type of question. Faced with this new type of question, I was no longer confused, no longer subconsciously comparing it with the old type of question and judging how much I could score. Instead, I broke my expectation about score and began to consciously focus my attention on the question itself and look for solving strategies. Finally, I got the answer by adding auxiliary lines and combining the Pythagorean Theorem and vertical bisectors' properties."*
>
> B_2: *"In solving the problem 'Given the height and distance of two straight rods, the distance between the top of the rod and a certain point is equal, find the horizontal distance between the bottom of the rod and that point,' I found that I could not prove the congruence of triangles with the methods I had mastered, nor did I know of any other methods. Also, I found myself spending too much time on this problem, so I decided to take the plunge. I guessed the answer according to the existing numbers in the question, and I told myself that the answer which I think is correct is probably the right one. And it would be difficult for the teacher to be random when setting the question. Instead of trying to solve this question, I would rather pay attention to the details in the question stem and guess the answer first. Finally, my guess turned out to be the right answer."..."I didn't know where to start when I answered the question 'draw a specified square and triangle in a square grid,' and then I chose to do the next one first. At this point, I will just focus on the present problem and not think about the previous problem. When I finish this problem, I will use the rest of the time to think back to the previous problem."*

4.3. What mindfulness brings for math self-efficacy

Based on the analytic framework (Table 2), two general themes emerged from our group discussion and in-depth interviews regarding mindfulness' effects on participants' MSE, as identified below.

4.3.1. Gain a better understanding of the judgment and evaluation of MSE

According to the participants, an non-judgemental attitude—an essential element of mindfulness practice—enabled them to recognize and understand their actual MSE concerning their experience. Two themes came to light in terms of the MBI's perceived effects:

1. students stopped comparing their performance to their peers', and,
2. students reduced their overconfidence and underconfidence regarding math test performance.

Students suggested, in the group discussion, how MBI might help them maintain a more general and practical understanding of their MSE:

> *"Part of my mind is always comparing. If I compare myself to people better than me, I gonna be unhappy... Sometimes, I use harsh self-criticism, based on my results of the comparison with peer students, to motivate myself to study... The awareness and the wisdom I gain from mindfulness work together, helping to reduce the time spent in judgment and evaluation of the past experiences dictate my opinions concerning my ability in mathematics."*
>
> *"I believe myself to be better than others in mathematics. Therefore, I may sometimes be overconfident in a math test, and it is more likely for me to make some stupid mistakes in the exams. This kind of mathematics self-efficacy sometimes annoys me. I have been punished for my carelessness in both math tests and math homework. My carelessness used to cost me first place on the test several times… through mindfulness, I have an awareness of knowing what my mind is doing. I slow down and read the questions fully, and I create an outline for the geometrical proof, which is good for the organization before I start writing. This may help me to avoid making stupid mistakes or random mistakes on exams."*
>
> *"For me, studying mathematics is a negative experience. I used to think that I was not a math person. I used to have a weak will in exams because of my slow thinking. When I found that people were all the better than me in mathematics, I would become more cynical and have low exam results expectations...At every moment, and although I know that I do not have any control over what has happened in the past. Instead of comparing myself to others, feeling guilt, resentment, bitterness, and sadness on my previous mathematics experience, I shall constantly focus on what I will achieve in this present math test."*

4.3.2. Reduces MSE-related problem-solving obstacles

The following three themes emerged from our qualitative research regarding how students' judgments about their ability to accomplish the math test (mathematics self-efficacy-related) affected the obstacles to their mathematics problem-solving. Specifically, students:

1. compared their problem-solving speed to their peers', breaking their time arrangements for answering the test paper;
2. excessively pursued ingenious problem-solving methods; and,
3. habitually walked away from challenging mathematics problems.

Respondents indicated that MBI could help them through the above difficulties:

> *"I'm a little slow at how to find if triangles are congruent, and in the past, I'd start noticing if everyone around me had already constructed congruent triangles, and then I'd blame myself for not being able to solve them. After taking part in the mindfulness intervention, I stopped judging my and others' math ability and focused my attention on the present through the senses. I no longer pay attention to the speed of the people around me, no longer compare myself to others, and question and deny my ability based on others' progress. Instead, I only pay attention to the problem and identify the appropriate one from the five rules to prove triangles congruent."*
>
> *"I think I've always been good at math. Therefore, I was accustomed to comparing and analyzing various solutions before starting to solve a test question. For example, when facing the item 'draw a triangle with sides $3,\sqrt{10},\sqrt{13}$ of in the square grid (Week 3 test),' though some conventional methods could solve it, I just want to find a different triangle that could make my solution remarkable. And it turns out not gratifying because this wastes me a lot of time. To be honest, I sometimes pursue more skilful solutions, which could distinguish me from other students. Somehow, it is tough to find an ingenious solution in the examination situation, and it seems to be the loss outweighs the gain. Mindfulness told me that I should find my inner power and constructively work with what I think is the best at each given moment instead of constantly striving toward the new and better based on my judgment."*
>
> *"I tend to put more weight on my past failures when challenging the challenging question in the test. This negative bias keeps me habitually running away from those difficult problems, usually the second question of the item. Then mindfulness told me that the outcome of any particular situation is not determined by the factors we went into. If we change the factors, we can continuously change the results. Instead of judging events that haven't happened yet, and, frankly, may never happen, we should be open and accepting. Hence, I have a good attitude toward those challenging items and challenged the last item's second question on the Week 4 test. Though I didn't get it totally done, I still get some points on this item."*

5. Discussion and conclusion

This study has examined the relationship between mindfulness and middle school students' high-stakes mathematics test achievement by exploring the association between mindfulness, MTA, and mathematics self-efficacy. The *p*-value and effect size values revealed significant positive correlations between mindfulness and middle school students' mathematics achievement. Semi-structured interview results provided evidence of the mediational mechanism between mindfulness and MTA. The relationship between mindfulness and mathematics achievement was partially mediated through MTA and mathematics self-efficacy. Our results are somewhat similar to those reported by previous studies, though these are noticeably slim in the domain of mathematics education.

First, the findings support previous research indicating positive associations between mindfulness and academic achievements

TABLE 7 The classification of participants' ideas according to the zone patterns of the "past," the "present," and the "future."

Zone patterns		
Past	Present	Future
1. Sorrow for their parents. 2. Comparing self-performance to peers. 3. Clinging on to some familiar problem-solving method. 4. Excessive upset when the real difficulty of the test is inconsistent with the anticipated difficulty. 5. Habitually walk away from any challenging mathematics problem. 6. Obsessing on some unsolved problems, unable to adjust mind in time to solve the next problem.	Paying attention in the moment, in an open and accepting way, to what you are experience in high stakes math test.	1. Excessive consideration of the consequences of failure in this math exam. 2. Excessive pursuit of ingenious problem-solving methods. 3. Comparing problem-solving speed to peers, breaking own time arrangement for answering the test paper.

(Beauchemin et al., 2008; Lu et al., 2017; Miralles-Armenteros et al., 2021). In particular, Bellinger et al. (2015), the first to investigate the connection between mindfulness and mathematics performance in high-stakes academic testing environments, found that mindfulness indirectly benefited math performance by attenuating the harmful effects of test anxiety. Also, consistent with Franco et al. (2010) and Samuel and Warner (2021), the relationship between mindfulness and academic performance in this study was mediated by anxiety and self-efficacy. Although there is no quantitative support for this mediating mechanism, our research has highlighted robust qualitative supporting evidence. More specifically, mindfulness appears to clear up symptoms of MTA disorders, including physiological manifestations, test-irrelevant thinking and worry, and mathematics test anxiety-related problem-solving obstacles. Also, mindfulness, especially adopting a non-judgmental attitude, moderated students' mathematics self-efficacy. Mindfulness can make students more cognizant of and confident in their ability to perform or accomplish specific mathematical tasks or problems; in other words, students with a non-judgmental attitude can think of their abilities better and stay focused on the "present-moment" in high-stakes tests, thereby better accessing their mathematics self-efficacy. Table 7 may largely subsume participants' ideas according to the zone patterns of the "past," the "present," and the "future," which may reveal the underlying and intervening mechanisms in mindfulness and students' high-stakes mathematics test achievement relationships.

Second, the current research does not explore how working memory may apply to understanding the phenomenon of mathematics anxiety as reported in previous studies (Mandler and Sarason, 1952; Eysenck and Calvo, 1992; Ashcraft and Moore, 2009; Berggren and Derakshan, 2013). Many studies have found it (Hitch, 1978; Gathercole and Pickering, 2000; Swanson and Sachse-Lee, 2001) over the years, indicating that the deficit of working memory would have a powerful effect on arithmetic. One reason could be that the math test used in this study was on geometry rather than arithmetic, which requires a more robust understanding of geometrical theorems (i.e., the Pythagorean Theorem, the Condition of Triangle Congruence, and the Property of Congruent Figures). This, in turn, requires less working memory and might interfere less with the mathematics anxiety associated with working memory deficits.

Third, in-depth interviews with students who did not show significant improvement also showed they viewed mindfulness as "mysterious" and "magical." They believed mindfulness was equivalent to Buddhism and could not fully devote themselves to the intervention during the experiment; "The whole audio gave me a sense of mystery. As a result, I don't feel fully immersed in the environment that the audio provides." Therefore, mindfulness needs to be demystified and re-contextualized for the educational context so students can more easily understand and accept the intervention. "Demystifying" involves integrating the Buddhist dharma of mindfulness into students' daily school life in a language that maintains its original meaning without mentioning the term "Buddha dharma." "Re-contextualization in the educational context" refers to comprehending mindfulness in an educational context. The connotation of mindfulness is explained based on the difference between Buddhist and educational contexts. Through attentional intervention, students are taught to be aware of their "bad" psychology and behaviors in their daily studies and examinations, treat them with a non-judgemental attitude, accept them, and stay in the present moment.

Furthermore, mindfulness showed effects related to religious concepts on students' exams:

> *"During the exam, I would tell myself that the result is a foregone conclusion. What I need to do at this moment is to try my best to do well in the questions that I know the answer. As long as I did my best in the exam, the score is not what I need to worry about. After I had got involved in the intervention, I found that my expectation for the score would not disturb my rhythm anymore."*

Although researchers emphasized demystifying mindfulness and re-contextualizing it in an educational context, mindfulness cannot eliminate the Buddhist context in some students' minds. Accordingly, identifying ways to handle this aspect should be a priority in future research.

As with all research, the current research has both strengths and limitations. From a methodological viewpoint, a major benefit of our study is its mindfulness instrument. The mindfulness audio in this study was more tailored for a mathematics test than those used in previous studies in that it targeted students' math test performance in an actual high-stakes test situation. However, the study's methodological limitations should be considered when interpreting its findings, particularly its lack of a randomized design; participants were limited to a convenient sample drawn from a single school, which limits generalizability. Future research should employ more extensive and diverse samples to yield results that can be more readily generalized. Another potential weakness involves mathematical content bias. Since only the students' recent math study was used to compile the test, the results do not reflect mindfulness' effect on their participation in high-stakes tests on other mathematical content.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Yangzhou University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

HZ contributed to the study conceptualization and design, produced the mindfulness audio, and conducted the data collection. HZ and LW collaborated in preparing the initial draft of the manuscript. LW edited the final manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Jiangsu Province University Philosophy and Social Science Fund Project (2019SJA1798), Project of the 14th Five Year Plan of Educational Science in Jiangsu Province of China (C/2022/01/72), Project of the 14th Five Year Plan of Educational Science of Nanjing (LZ/2021/122), Project of Teaching Reform in Yangzhou University (YZUJX2019-18B), and the Qinglan Project of Yangzhou University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amrein, A. L., and Berliner, D. C. (2002). High-stakes testing and student learning. *Educ. Policy Anal. Arch.* 10, 18. doi: 10.14507/epaa.v10n18.2002

Ancona, M. R., and Mendelson, T. (2014). Feasibility and preliminary outcomes of a yoga and mindfulness intervention for school teachers. *Adv. Sch. Ment. Health Promot.* 7, 156–170. doi: 10.1080/1754730X.2014.920135

Ashcraft, M. H. (2002). Math anxiety: personal, educational, and cognitive consequences. *Curr. Dir. Psychol. Sci.* 11, 181–185. doi: 10.1111/1467-8721.00196

Ashcraft, M. H., and Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *J. Psychoeduc. Assess.* 27, 197–205. doi: 10.1177/0734282908330580

Baer, R. (2010). *Assessing Mindfulness and Acceptance Processes in Clients: Illuminating the Theory and Practice of Change*. Oakland, CA: New Harbinger Publications.

Bandura, A. (2010). "Self-efficacy," in *The Corsini Encyclopedia of Psychology*, eds I. Weiner and W. E. Craighead (Hoboken, NJ: John Wiley & Sons), 1–3. doi: 10.1002/9780470479216.corpsy0836

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., and Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychol. Bull.* 147, 134–168. doi: 10.1037/bul0000307

Beauchemin, J., Hutchins, T. L., and Patterson, F. (2008). Mindfulness meditation may lessen anxiety, promote social skills, and improve academic performance among adolescents with learning disabilities. *Complement. Health Pract. Rev.* 13, 34–45. doi: 10.1177/1533210107311624

Beilock, S. L., Kulp, C. A., Holt, L. E., and Carr, T. H. (2004). More on the fragility of performance: choking under pressure in mathematical problem solving. *J. Exp. Psychol. Gen.* 133, 584–600. doi: 10.1037/0096-3445.133.4.584

Bellinger, D. B., DeCaro, M. S., and Ralston, P. A. (2015). Mindfulness, anxiety, and high-stakes mathematics performance in the laboratory and classroom. *Conscious. Cogn.* 37, 123–132. doi: 10.1016/j.concog.2015.09.001

Berggren, N., and Derakshan, N. (2013). Attentional control deficits in trait anxiety: why you see them and why you don't. *Biol. Psychol.* 92, 440–446. doi: 10.1016/j.biopsycho.2012.03.007

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *J. Couns. Psychol.* 25, 441–448. doi: 10.1037/0022-0167.25.5.441

Betz, N. E., and Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *J. Vocat. Behav.* 23, 329–345. doi: 10.1016/0001-8791(83)90046-5

Birdee, G. S., Wallston, K. A., Ayala, S. G., Ip, E. H., and Sohl, S. (2020). Development and psychometric properties of the self-efficacy for mindfulness meditation practice scale. *J. Health Psychol.* 25, 2017–2030. doi: 10.1177/1359105318783041

Bishop, S. R., Lau, M., Shapiro, S., Carlson, L., Anderson, N. D., Carmody, J., et al. (2006). Mindfulness: a proposed operational definition. *Clin. Psychol. Sci. Pract.* 11, 230–241. doi: 10.1093/clipsy.bph077

Burke, C. A. (2010). Mindfulness-based approaches with children and adolescents: a preliminary review of current research in an emergent field. *J. Child Fam. Stud.* 19, 133–144. doi: 10.1007/s10826-009-9282-x

Cassady, J. C., and Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemp. Educ. Psychol.* 27, 270–295. doi: 10.1006/ceps.2001.1094

Chan, S. H. W., Yu, C. K., and Li, A. W. O. (2021). Impact of mindfulness-based cognitive therapy on counseling self-efficacy: a randomized controlled crossover trial. *Patient. Educ. Couns.* 104, 360–368. doi: 10.1016/j.pec.2020.07.022

Chang, V. Y., Palesh, O., Caldwell, R., Glasgow, N., Abramson, M., Luskin, F., et al. (2010). The effects of a mindfulness-based stress reduction program on stress, mindfulness self-efficacy, and positive states of mind. *Stress Health* 20, 141–147. doi: 10.1002/smi.1011

Changyu, S., and Xiao, C. (2016). The effect of mindfulness meditation training on middle school students 'self-control.' *Chin. J. Health Psychol.* 24, 1359–1363.

Chiesa, A., and Serretti, A. (2011). Mindfulness based cognitive therapy for psychiatric disorders: a systematic review and meta-analysis. *Psychiatry Res.* 187, 441–453. doi: 10.1016/j.psychres.2010.08.011

Ching, B. H.-H. (2017). Mathematics anxiety and working memory: longitudinal associations with mathematical performance in chinese children. *Contemp. Educ. Psychol.* 51, 99–113. doi: 10.1016/j.cedpsych.2017.06.006

Ching, H., Koo, M., Tsai, T., and Chen, C. (2015). Effects of a mindfulness meditation course on learning and cognitive performance among university students in Taiwan. *Evid. Complement. Altern. Med.* 2015, 254358. doi: 10.1155/2015/254358

Cohen, L., Manion, L., and Morrison, K. (2017). *Research Methods in Education*. New York, NY: Routledge. doi: 10.4324/9781315456539

Cooper, S. E., and Robinson, D. A. (1991). The relationship of mathematics self-efficacy beliefs to mathematics anxiety and performance. *Meas. Eval. Counsel. Dev.* 24, 4–11.

Council, N. (1998). *High Stakes: Testing for Tracking, Promotion, and Graduation.* Washington, DC: National Academy Press.

de Abreu Costa, M., D'Alò de Oliveira, G. S., Tatton-Ramos, T., Manfro, G. G., and Salum, G. A. (2019). Anxiety and stress-related disorders and mindfulness-based interventions: a systematic review and multilevel meta-analysis and meta-regression of multiple outcomes. *Mindfulness* 10, 996–1005. doi: 10.1007/s12671-018-1058-1

DeCaro, M. S., Rotar, K. E., Kendra, M. S., and Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving. *Quart. J. Exp. Psychol.* 63, 1619–1630. doi: 10.1080/17470210903474286

Dowker, A., Sarkar, A., and Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years? *Front. Psychol.* 7, 508. doi: 10.3389/fpsyg.2016.00508

Duncan, L. G., and Bardacke, N. (2010). Mindfulness-based childbirth and parenting education: promoting family mindfulness during the perinatal period. *J. Child Fam. Stud.* 19, 190–202. doi: 10.1007/s10826-009-9313-7

Essau, C. A., Sasagawa, S., Anastassiou-Hadjicharalambous, X., Guzmàn, B. O., and Ollendick, T. H. (2011). Psychometric properties of the spence child anxiety scale with adolescents from five european countries. *J. Anxiety Disord.* 25, 19–27. doi: 10.1016/j.janxdis.2010.07.001

Evans, S., Ferrando, S., Findler, M., Stowell, C., and Haglin, D. (2008). Mindfulness-based cognitive therapy for generalized anxiety disorder. *J. Anxiety Disord.* 22, 716–721. doi: 10.1016/j.janxdis.2007.07.005

Eysenck, M. W., and Calvo, M. G. (1992). Anxiety and performance: the processing efficiency theory. *Cognit. Emot.* 6, 409–434. doi: 10.1080/02699939208409696

Felver, J. C., Hoyos, C. E. C., Tezanos, K. M., and Singh, N. N. (2016). A systematic review of mindfulness-based interventions for youth in school settings. *Mindfulness* 7, 34–45. doi: 10.1007/s12671-015-0389-4

Franco, C., Mañas, I., Cangas, A. J., and Gallego, J. (2010). "The applications of mindfulness with students of secondary school: results on the academic performance, self-concept and anxiety," in *Knowledge Management, Information Systems, E-Learning, and Sustainability Research: Third World Summit on the Knowledge Society, WSKS 2010, Corfu, Greece, September 22–24, 2010, Proceedings, Part I*, eds M. D. Lytras, P. O. De Pablos, A. Ziderman, A. Roulstone, H. Maurer, and J. B. Imber (Berlin; Heidelberg: Springer), 83–97. doi: 10.1007/978-3-642-16318-0_10

Gathercole, S. E., and Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *Brit. J. Educ. Psychol.* 70, 177–194. doi: 10.1348/000709900158047

Haladyna, T. M., Nolen, S. B., and Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educ. Res.* 20, 2–7. doi: 10.3102/0013189X020005002

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Rev. Educ. Res.* 58, 47–77. doi: 10.3102/00346543058001047

Hipson, W. E., and Coplan, R. J. (2017). "Social withdrawal in childhood," in *Encyclopedia of Evolutionary Psychological Science*, eds T. Shackelford and V. Weekes-Shackelford (Cham: Springer), 1–10. doi: 10.1007/978-3-319-16999-6_2447-1

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognit. Psychol.* 10, 302–323. doi: 10.1016/0010-0285(78)90002-6

Ho, E. S.-c. (2009). Characteristics of east asian learners: what we learned from PISA. *Educ. Res. J.* 24, 327–348.

Hofmann, S. G., and Gòmez, A. F. (2017). Mindfulness-based interventions for anxiety and depression. *Psychiatr. Clin. North Am.* 40, 739–749. doi: 10.1016/j.psc.2017.08.008

Hoge, E. A., Bui, E., Marques, L., Metcalf, C. A., and Simon, N. M. (2013). Randomized controlled trial of mindfulness meditation for generalized anxiety disorder: effects on anxiety and stress reactivity. *J. Clin. Psychiatry* 74, 786–792. doi: 10.4088/JCP.12m08083

Jellinek, M. S., and Snyder, J. B. (1998). Depression and suicide in children and adolescents. *Pediatr. Rev.* 19, 255–64. doi: 10.1542/pir.19-8-255

Jones, M. G., Jones, B. D., and Hargrove, T. Y. (2003). *The Unintended Consequences of High-stakes Testing*. Lanham, MD: Rowman & Littlefield.

Kabat-Zinn, J. (2003). Mindfulness-based stress reduction (MBSR). *Psychother. Psychosom. Medizin. Psychol.* 61, 328–332. doi: 10.1055/s-0031-1276859

Kabat-Zinn, J. (2010). Mindfulness-based interventions in context: past, present, and future. *Clin. Psychol. Sci. Pract.* 10, 144–156. doi: 10.1093/clipsy.bpg016

Keng, S.-L., Smoski, M. J., and Robins, C. J. (2011). Effects of mindfulness on psychological health: a review of empirical studies. *Clin. Psychol. Rev.* 31, 1041–1056. doi: 10.1016/j.cpr.2011.04.006

Kirkpatrick, R., and Zang, Y. (2011). The negative influences of exam-oriented education on chinese high school students: backwash from classroom to child. *Lang. Test. Asia* 1, 36. doi: 10.1186/2229-0443-1-3-36

Kuyken, W., Weare, K., Ukoumunne, O. C., Vicary, R., Motton, N., Burnett, R., et al. (2013). Effectiveness of the mindfulness in schools programme: non-randomised controlled feasibility study. *Brit. J. Psychiatry* 203, 126–131. doi: 10.1192/bjp.bp.113.126649

Lagor, A. F., Williams, D. J., Lerner, J. B., and Mcclure, K. S. (2013). Lessons learned from a mindfulness-based intervention with chronically ill youth. *Clin. Pract. Pediatr. Psychol.* 1, 146–158. doi: 10.1037/cpp0000015

Lam, K. (2016). School-based cognitive mindfulness intervention for internalizing problems: pilot study with Hong Kong elementary students. *J. Child Fam. Stud.* 25, 3293–3308. doi: 10.1007/s10826-016-0483-9

Lau, N. S. E., and Hue, M. T. (2011). Preliminary outcomes of a mindfulness-based programme for Hong Kong adolescents in schools: well-being, stress and depressive symptoms. *Int. J. Child. Spiritual.* 16, 315–330. doi: 10.1080/1364436X.2011.639747

Leppma, M., and Darrah, M. (2022). Self-efficacy, mindfulness, and self-compassion as predictors of math anxiety in undergraduate students. *Int. J. Math. Educ. Sci. Technol.* 2022, 1–16. doi: 10.1080/0020739X.2022.2054740

Leung, F. K. S. (2001). In search of an east asian identity in mathematics education. *Educ. Stud. Math.* 47, 35–51. doi: 10.1023/A:1017936429620

Lewis, A. (1970). The ambiguous word 'anxiety'. *Int. J. Psychiatry* 9, 62–79.

Li, J. (2001). Ten kinds of psychology that affect the result of mathematics examination. *Math. Teach. Ref. Middle Sch.* 6, 53–54.

Liehr, P., and Diaz, N. (2010). A pilot study examining the effect of mindfulness on depression and anxiety for minority children. *Arch. Psychiatr. Nurs.* 24, 69–71. doi: 10.1016/j.apnu.2009.10.001

Lin, J., and Chen, Q. (1995). Academic pressure and impact on students' development in China. *McGill J. Educ.* 30, 149–168.

Lu, S., Huang, C.-C., and Rios, J. (2017). Mindfulness and academic performance: an example of migrant children in China. *Child. Youth Serv. Rev.* 82, 53–59. doi: 10.1016/j.childyouth.2017.09.008

Maloney, E. A., and Beilock, S. L. (2012). Math anxiety: who has it, why it develops, and how to guard against it. *Trends Cognit. Sci.* 16, 404–406. doi: 10.1016/j.tics.2012.06.008

Mandler, G., and Sarason, S. B. (1952). A study of anxiety and learning. *J. Abnorm. Soc. Psychol. Educ.* 47, 166–173. doi: 10.1037/h0062855

Meiklejohn, J., Phillips, C., Freedman, M. L., Griffin, M. L., Biegel, G., Roach, A., et al. (2012). Integrating mindfulness training into k-12 education: fostering the resilience of teachers and students. *Mindfulness* 3, 291–307. doi: 10.1007/s12671-012-0094-5

Metz, S. M., Frank, J. L., Reibel, D., Cantrell, T., Sanders, R., and Broderick, P. C. (2013). The effectiveness of the learning to breathe program on adolescent emotion regulation. *Res. Hum. Dev.* 10, 252–272. doi: 10.1080/15427609.2013.818488

Miralles-Armenteros, S., Chiva-Gómez, R., Rodrí-guez-Sánchez, A., and Barghouti, Z. (2021). Mindfulness and academic performance: the role of compassion and engagement. *Innov. Educ. Teach. Int.* 58, 3–13. doi: 10.1080/14703297.2019.1676284

Napoli, M., Krech, P. R., and Holley, L. C. (2005). Mindfulness training for elementary school students: the attention academy. *J. Appl. Sch. Psychol.* 21, 99–125. doi: 10.1300/J370v21n01_05

OECD (2013). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs, Vol. III, Preliminary Version*. Paris: OECD.

Pajares, F., and Urdan, T. (1996). Exploratory factor analysis of the mathematics anxiety scale. *Meas. Eval. Counsel. Dev.* 29, 35–47.

Punaro, L., and Reeve, R. (2012). Relationships between 9-year-olds' math and literacy worries and academic abilities. *Child Dev. Res.* 2012, 359089. doi: 10.1155/2012/359089

Reback, R., Rockoff, J. E., and Schwartz, H. L. (2014). Under pressure: job security, resource allocation, and productivity in schools under no child left behind. *Amer. Econ. J. Econ. Policy* 6, 207–241. doi: 10.1257/pol.6.3.207

Samuel, T. S., and Warner, J. (2021). "I can math!": reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Commun. Coll. J. Res. Pract.* 45, 205–222. doi: 10.1080/10668926.2019.1666063

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *J. Pers. Soc. Psychol.* 46, 929–938. doi: 10.1037/0022-3514.46.4.929

Schleicher, A. (2019). *Pisa 2018: Insights and Interpretations.* Paris: OECD Publishing.

Schunk D. H. (1982). Progress self-monitoring. *J. Exp. Educ.* 51, 89–93. doi: 10.1080/00220973.1982.11011845

Schwarzer, R., and Quast, H. (1985). Multidimensionality of the anxiety experience: evidence for additional components. *Adv. Test Anx. Res.* 4, 3–14.

Scott-Hamilton, J., Schutte, N. S., and Brown, R. F. (2016). Effects of a mindfulness intervention on sports - anxiety, pessimism, and flow in Competitive cyclists. *Appl. Psychol. Health Well Being* 8, 85–103. doi: 10.1111/aphw.12063

Semple, R. J., Lee, J., Rosa, D., and Miller, L. F. (2010). A randomized trial of mindfulness-based cognitive therapy for children: promoting mindful attention to enhance social-emotional resiliency in children. *J. Child Fam. Stud.* 19, 218–229. doi: 10.1007/s10826-009-9301-y

Shanshan, Z., and Zhun, G. (2018). "A study on the effect of mindfulness training on mental wandering of college students," in *Abstract of the 21st National Psychological Academic Conference*. Beijing: Chinese Psychological Association.

Shapiro, S. L., Brown, K. W., and Astin, J. (2011). Toward the integration of meditation into higher education: a review of research evidence. *Teach. Coll. Rec.* 113, 493–528. doi: 10.1177/016146811111300306

Sheffield, D., and Hunt, T. (2006). How does anxiety influence maths performance and what can we do about it? *MSOR Connect.* 6, 19. doi: 10.11120/msor.2006.06040019

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., and D'Argembeau, A. (2011). Mind-wandering: phenomenology and function as assessed with a novel experience sampling method. *Acta Psychol.* 136, 370–381. doi: 10.1016/j.actpsy.2011.01.002

Stobart, G., and Eggen, T. (2012). "High-stakes testing-value, fairness and consequences," *Assess. Educ. Princip. Policy Pract.* 19, 1–6. doi: 10.1080/0969594X.2012.639191

Suprapto, N. (2016). "What should educational reform in Indonesia look like? - Learning from the pisa cience scores of east-asian countries and Singapore," in *Asia-Pacific Forum on Science Learning Teaching*, Vol. 17, 1–21.

Swanson, H. L., and Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: both executive and phonological processes are important. *J. Exp. Child Psychol.* 79, 294–321. doi: 10.1006/jecp.2000.2587

Tan, J. B., and Yates, S. (2011). Academic expectations as sources of stress in asian students. *Soc. Psychol. Educ.* 14, 389–407. doi: 10.1007/s11218-010-9146-7

Vidic, Z., and Cherup, N. (2019). Mindfulness in classroom: effect of a mindfulness-based relaxation class on college students' stress, resilience, self-efficacy and perfectionism. *Coll. Stud. J.* 53, 130–144.

Vøllestad, J., Sivertsen, B., and Nielsen, G. H. (2011). Mindfulness-based stress reduction for patients with anxiety disorders: evaluation in a randomized controlled trial. *Behav. Res. Ther.* 49, 281–288. doi: 10.1016/j.brat.2011.01.007

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., and Harari, R. R. (2013). Mathematics anxiety in young children: concurrent and longitudinal associations with mathematical performance. *Contemp. Educ. Psychol.* 38, 1–10. doi: 10.1016/j.cedpsych.2012.09.001

Wall, R. B. (2005). Tai chi and mindfulness-based stress reduction in a boston public middle school. *J. Pediatr. Health Care* 19, 230–237. doi: 10.1016/j.pedhc.2005.02.006

Wigfield, A., and Judith, L. M. (1988). Math anxiety in elementary and secondary school students. *J. Educ. Psychol.* 80, 210–216. doi: 10.1037/0022-0663.80.2.210

Williams, M., Teasdale, J., Segal, Z., and Kabat-Zinn, J. (2007). *The Mindful Way Through Depression: Freeing Yourself from Chronic Unhappiness*. New York, NY: The Guilford Press.

Wu, S., Amin, H., Barth, M., Malcarne, V., and Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Front. Psychol.* 3, 162. doi: 10.3389/fpsyg.2012.00162

Yu, F., and Wenjie, D. (2017). The basic forms and characteristics of school mindfulness intervention. *Chin. J. Clin. Psychol.* 25, 991–994.

Zeidan, F., Martucci, K. T., Kraft, R. A., McHaffie, J. G., and Coghill, R. C. (2014). Neural correlates of mindfulness meditation-related anxiety relief. *Soc. Cognit. Affect. Neurosci.* 9, 751–759. doi: 10.1093/scan/nst041

Zhang, J., Zhao, N., and Kong, Q. P. (2019). The relationship between math anxiety and math performance: a meta-analytic investigation. *Front. Psychol.* 10, 1613. doi: 10.3389/fpsyg.2019.01613

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 12 May 2023
DOI 10.3389/fpsyg.2023.1125589

Check for updates

OPEN ACCESS

EDITED BY
Shuhua An,
California State University,
Long Beach, United States

REVIEWED BY
Pinaki Chakraborty,
Netaji Subhas University of Technology, India
Shuang Song,
Capital Normal University, China

*CORRESPONDENCE
Rangmei Li
rangmei_li@mail.bnu.edu.cn

RECEIVED 16 December 2022
ACCEPTED 26 April 2023
PUBLISHED 12 May 2023

CITATION
Shang X, Li R and Li Y (2023) The effect of structured stepwise presentations on students' fraction learning: an eye-tracking study. *Front. Psychol.* 14:1125589. doi: 10.3389/fpsyg.2023.1125589

The effect of structured stepwise presentations on students' fraction learning: an eye-tracking study

Xiaoqing Shang[1], Rangmei Li[2]* and Yangping Li[3]

[1]Faculty of Education, Shaanxi Normal University, Xi'an, Shaanxi, China, [2]School of Mathematical Sciences, Beijing Normal University, Beijing, China, [3]Hangzhou International Urbanology Research Center and Zhejiang Urban Governance Studies Center, Hangzhou, Zhejiang, China

The structured stepwise presentation is based on the segmenting and cueing principles. The main purpose of the study was to examine the effect of the structured stepwise presentations on students' attention and fraction learning. A total of 100 primary pupils participated in this study. They were divided into three parallel groups and were, respectively, applied three kinds of presentation types (structured and stepwise, no structure and stepwise, and structure and no stepwise) of the teaching content to learn the fraction concept. A stable eye tracker was used to record students' visual attention during learning, the first fixation duration and total fixation duration of students were recorded, and the regression time was also calculated within correspondent relative elements. After the experiment, through a one-way ANOVA test, we found significant differences among the three groups in students' attention. The learning performance of these three groups also differed. The results showed that structured stepwise presentation played an important role in attention guidance during fraction teaching. It better guided students' attention to connecting relative elements and resulted in better learning performance in fraction learning. The findings suggested the importance of structured stepwise presentations during teaching practices.

KEYWORDS
structured presentation, stepwise presentation, attention guidance, eye-tracking, fraction learning

1. Introduction

Multimedia technology provides opportunities to optimize teaching and learning due to its multiple and dynamic information presentation features. Especially for primary school teaching, presentations based on multimedia technology are beneficial for promoting deep understanding, since abstract concepts can be illustrated in multiple methods and visualized by demonstrating dynamic sequential processes (Rieber, 1990; Park and Hopkins, 1992), which draw students' attention to key information, thus effectively prevent them from being distracted by external factors (Boucheix and Guignard, 2005; Wouters et al., 2008).

Cueing and segmenting in multimedia presentations have been successfully used to guide students' attention and promote their learning (De Koning et al., 2007; Mayer, 2014; Van Gog, 2014; Rey et al., 2019). Cueing refers to the manipulation of visuospatial characteristics of instructional materials (Mautone and Mayer, 2001), such as distinctive colors, arrows, labels, picture references, and mode of arrangement, which can draw students' attention to related content and decrease the associated extraneous cognitive load (De Koning et al., 2009; Wang et al., 2013). Segmenting refers to the process of breaking the instructional content into individual parts (Rey et al., 2019). Combining cueing with segmenting to reflect structural relationships and framework of knowledge is called structured presentations in our study (De Koning et al., 2009), which can enable students to connect corresponding information easily. In addition, stepwise is a kind of dynamic cue based on segmentation, which is used to gradually highlight each segment accompanied by oral guidance (Chen et al., 2016). The stepwise presentation allows students to learn at their own pace and gives them sufficient time to integrate information, thus reducing and compensating for potential split attention (Stiller et al., 2011; Rey et al., 2019). Therefore, guiding students' attention through multimedia techniques is key to optimizing students' learning.

Research has consistently shown that fractions are challenging for primary students to learn (Lin et al., 1996; Lortie-Forgues et al., 2015; Reinhold et al., 2020). It is hard for students to make connections between natural numbers and fractions, as the fraction concept involves high-element interactivity (Reinhold et al., 2020). The inappropriate teaching designs, such as scenario fiction, formalization of process, and improper connection of prior knowledge, further increase students' cognitive load and make them feel struggling to understand (Lin et al., 1996). To our knowledge, research on combining structured presentations and stepwise presentations to guide students' attention and its impact on students' fraction learning, remains unexplored. Eye-tracking technology can scientifically interpret learners' learning processes through their eye movements, and, especially in multimedia learning, fixation duration and regression time are key indicators of students' attention and cognitive processes (Ozcelik et al., 2010; Jamet, 2014; Jian et al., 2019; Pi et al., 2020). Therefore, in this study, we attempted to employ eye-tracking technology to investigate the effect of teaching presentations on students' learning of fractions by integrating structured presentations and stepwise presentations into teaching to attract students' attention.

2. Theoretical background

2.1. Cues and segmentation adjust students' cognitive load

People's psychological resources are limited, and they cannot deal with excessive amounts of information simultaneously. The greater the visual difference between a perceived object and the background, the more noticeable and more easily it is perceived by an individual (Treisman and Schmidt, 1982). For fraction learning with high-element interactivity, simply using text sequence descriptions is difficult to provide the visual difference for students, and would waste students' cognitive resources to visualize information by themselves (Seufert and Brünken, 2006; Richter et al., 2016). According to the cueing principle, cues can decrease students' extraneous cognitive load and draw their attention to key teaching points (Seufert and Brünken, 2006; Doolittle and Altstaedter, 2009; Van Gog, 2014). The empirical studies showed that the total and first fixation durations of students on signal presentations were longer than on the no-signal teaching materials (Boucheix and Lowe, 2010; Ozcelik et al., 2010; Pi et al., 2020). Thus, adding cues to the text presentations of fraction concepts could strengthen students' retention of knowledge and promote learning efficiency (Wang et al., 2013; Schneider et al., 2018).

Segmenting–a form of temporal cueing–involves hidden signals that increase the salience of the natural boundaries between events in a process. Unblocked information presentation confuses key and background information, hinders students from selecting and organizing key information, and results in cognitive overload, reducing students' available cognitive capacity for dealing with essential information (Cierniak et al., 2009; Mayer and Pilegard, 2014). The principle of segmenting holds that complex content should be broken into smaller, manageable, learner-controlled units (Mayer and Pilegard, 2014) to maximize students' limited working memory. Segmenting in a presentation can reduce the viewers' cognitive processing burden (Kurby and Zacks, 2008; Spanjers et al., 2010) and decrease their extraneous cognitive load, because the information is clearly organized and minimizes the time spent searching for related information. Therefore, teaching presentations organized by segmentation enable learners to extract key information effectively, deeply understand the content at their own pace (Rey et al., 2019), and avoid split attention.

Segmentation can also balance students' intrinsic cognitive load by organizing multiple types of information (Spanjers et al., 2010; Rey et al., 2019), especially when presenting intricate content. The concept of fractions comes from daily life and undergoes a transformation from life context to mathematical symbols, so the rich information presented in different ways could arose students' learning interests and facilitate their understanding (Mitchell and Miller, 2010). However, teachers often transmitted the fraction concept directly to students, instead of connecting real-life scenarios with it (Lin et al., 1996). Using segmentation, the fraction concept can be broken into simple components, with sub-blocks to represent sub-concepts related to key content that students can learn and recall more easily, enabling them to experience organizing individual segments. At the same time, multiple representations including symbols and words can be used to express the fraction concept (Obersteiner et al., 2015; Reinhold et al., 2020). Thus, segmentation is beneficial for reserving cognitive capacity to enhance the perception of essential information, providing a method for learning fractions.

2.2. Structured presentations provide cues to draw attention to segments

In terms of multimedia techniques, structured presentations usually combine cueing and segmenting. First, segmenting splits complex content into known content, clearly distinguishing each sub-concept for identification by students. Cueing further reveals the logic of knowledge generation by building bridges between blocks and organizing each block into a coherent representation (De Koning et al., 2009). Combining cueing with segmenting supports the selection and organization of learning content, which is beneficial for attracting students' attention and decreasing their cognitive load. Especially for learning fractions, the content structure is complex and students need to conduct relational reasoning to understand. The structured presentation is helpful for enhancing relational reasoning about concepts by visually emphasizing the structure of the content (Kalra et al., 2020). This helps learners extract key information from static information and process it quickly (Tversky et al., 2008; Boucheix et al., 2013), enabling them to grasp the entire content (Gross and Harmon, 2009). Guided by a structured presentation, students can easily integrate a new topic with previous knowledge and comprehend the structure of a concept, which not only decreases students' cognitive load but also increases their confidence about new teaching content (Bransford et al., 2000; Smith and Shimeld, 2014).

According to previous studies on the effects of segmenting and cueing, structured presentations also improve knowledge retention (Schneider et al., 2018; Rey et al., 2019). Previous research has proved that event-structured knowledge is understandable and easily recalled (Carter, 1994) because it reflects the logical relationships within teaching content and facilitates the identification and subsequent representation of the material's structural organization (De Koning et al., 2009; Lee et al., 2018). Based on proven techniques (Boucheix and Guignard, 2005; De Koning et al., 2009; Boucheix et al., 2013; Richter et al., 2016), key information can be highlighted, and its relationships with relevant knowledge can be presented by using different signals to visualize organized knowledge elements, development logic, and structural hierarchy. Visualization of the knowledge structure can support learners in integrating the elements between and within representations into a coherent whole (Wouters et al., 2008; De Koning et al., 2010), which is beneficial for constructing schemata of the knowledge and committing them to memory.

2.3. Stepwise presentations add dynamic stimuli to attract students' attention

Stepwise is regarded as a key dynamic cue for presenting teaching content. Dynamic cues are privileged by the human visual system (Wolfe and Horowitz, 2004). Because of their dynamic nature, these cues can present interactive relationships among knowledge elements in a way that static signals cannot. Particularly for abstract cognitive processes, dynamic signals can facilitate the externalization of cognitive processes better than verbal descriptions (Wouters et al., 2008). Combining visual animation with narration encourages students to process information at a deeper level than narration or on-screen text alone (Dunsworth and Atkinson, 2007). Thus, the stepwise presentation can guide students' attention to dynamic elements by presenting different stimuli or posing different questions in each step. Under the guidance of the stepwise presentation, a new object, as the learning input can be quickly captured by students' perceptual system (Chen et al., 2016; Lei et al., 2017), and meanwhile, students' attention can be attracted to the location where the new objects occurred (Yantis and Jonides, 1996).

The stepwise presentation is also based on segmentation, which breaks the complex content into simple elements to decrease interactivity (Chen et al., 2016; Lee et al., 2018). As we all know, fraction concepts involve multiple elements that interact with each other, and presenting them all at once will increase students' cognitive load (Mayer and Pilegard, 2014). Thus, during teaching practice, decomposing the fraction concept is a common approach to reduce the complexity. Segmentation, as the first step of the stepwise presentation, can break the fraction concept into simple components, each unit representing a natural number that the students learned before. Combined with oral guidance and coloring, the stepwise presentation can attract students' attention to the key point step by step (Jamet et al., 2008). This can also provide scaffolding for students to build relationships between components. Therefore, stepwise presentations can help students construct knowledge actively.

3. Research questions and hypotheses

Existing studies have shown that structured presentations and stepwise presentations individually enhance students' learning (Miao et al., 2000; Tso et al., 2011; Kalra et al., 2020), which play different roles in teaching guidance. The purpose of the structured presentation is to visualize the overall static structure of mathematics content itself, and stepwise, as a kind of dynamic teaching guidance technique, aims to strengthen the generating logic of mathematical structure. However, the application of combing structured presentations and stepwise presentations in fraction teaching still remains unexplored. Especially regarding the effect of integrating these two methods on students' attention has received little attention from researchers. Therefore, in this study, we aimed to improve students' learning efficiency by combining different visual presentation approaches, including structured presentations, and stepwise presentations, which we called structured stepwise presentations, to draw students' attention. The research question asked whether students' attention and learning performance are influenced by structured stepwise presentations of fraction concepts. Based on prior studies, the hypotheses in this study were as follows:

> *H1*: Structured stepwise presentations can draw students' attention during the learning of fractions.
>
> *H2*: Structured stepwise presentations are advantageous for encouraging students to connect interactive elements between fraction concepts.
>
> *H3*: Students in the group using structured stepwise presentations can learn fraction concepts more easily than other groups.

4. Methods

4.1. Participants

A total of 100 third-grade students from a primary school (44 female and 56 male), none of whom had studied fractions before, were selected from 10 classes. The age of these students is eight or nine. Considering that neither structured nor stepwise presentation of fractions is relatively rare in teaching practice, we only focused on three conditions in this experiment. Initially, the participants were randomly assigned to one of these three conditions: a structured and stepwise (SaS) group, $n = 33$; an unstructured and stepwise (UaS) group, $n = 33$; and a structured and not stepwise (SaN) group, $n = 34$. After the experiment, five participants were excluded because their eye-tracking rates were below 80%. Finally, 95 participants (SaS group $n = 31$, UaS group $n = 32$, SaN group $n = 32$) were included in the analysis. We examined these participants' prior knowledge in the school before studying fractions. The test lasted 1 h and the total score was 100. The content was mainly related to the algebra knowledge that students have learned before, including multi-digit addition and subtraction, one-digit multiplication and division, the concept of decimals, the addition and subtraction of one-digit decimals, and the application of these concepts. After using analysis of variance (ANOVA), we found no significant differences among the three groups: F (2, 93) = 0.614, $p = 0.544 > 0.05$, which indicated the same level of mathematical ability of these students.

4.2. Instruments

We used a Tobii Pro X-60 stationary eye tracker to record the participants' eye-movement information. We installed it at the bottom of a 24-inch computer screen (see Figure 1), with a display resolution of 1,024 × 768 pixels, which we used to present different stimuli to the participants. We collected the participants' binocular eye-movement data at a 60-Hz sample rate. Before starting the experiment, we calibrated the eye tracker for each participant using a nine-point calibration.

After the experiment, each participant in the three groups was given 15 min to finish the post-test in order to examine the learning effect. Five questions tested the participants' understanding of the concept of fractions, and the total possible score was 30. Three of the questions tested the students' recollection of the presented teaching content, including a realistic explanation of a fraction and its function in mathematics. The other two questions tested students' simple relational reasoning about fractions.

4.3. Procedure

The two key variables in this experiment were *structured* and *stepwise*. In the structured knowledge presentation, we used a row-and-column organization of two-dimensional space to establish corresponding relationships. For example, to present the fraction concept, we divided the presentation into three columns: the left column presented key information about a real-life scenario, the middle column presented the symbolic mathematical elements, and the right column presented the literal mathematical elements. The information relating to the numerator, score line, and denominator was presented in rows and arranged according to the corresponding relationship. Thus, structured connections between fraction concepts could be visualized through correlations between real-life information, mathematical symbols, and mathematical elements or through abstract correlations between the key elements of the fraction. Additionally, In the stepwise presentation, each page presented different visual stimuli, distinguished information with distinct colors, and posed different problems to encourage the students to think. The stepwise action was prompted by "clicking," combined with oral guidance. Therefore, we assigned the participants to the three experimental groups based on the above variables, namely the SaS group, UaS group, and SaN group, as previously outlined. The presentation content, the teacher's explanations, and the teaching method were the same for all three groups; only the PowerPoint (PPT) presentation modes differed.

In the SaS group, the teaching presentation was structured and stepwise. The presentation slides were synchronized with the teacher's oral guidance, and each slide presented only one piece of crucial information, marked in red, which would turn black in the next step. We arranged the teaching content according to its structure. After all the slides had been presented, the students could see the structure of the entire content.

In the UaS group, the presentation was designed unstructured but stepwise with color. We arranged everything linearly in the order that it should be dealt with. The number of slides in the UaS group was the same as in the SaS group, but the arrangement of the PPT content differed.

In the SaN group, the presentation was structured, but there were no steps. The content was presented on one slide before the teacher gave oral guidance. Therefore, the number of slides was smaller than for the other groups.

The flow diagram of the eye-tracking experiment using the SaS group as an example is shown in Figure 2. The teaching for each group consisted of four phases, introduction, and then presenting three videos (videos A–C). The introduction part was used to allow students to experience the concept of average by dividing apples in a real-life situation. After that, videos A–C were played consecutively at the same interval and each video was preceded by a guiding question. The total length of these three videos was 8 min and 15 s. For the SaS group and the UaS group, the numbers of slides in videos A, B, and C were 7, 6,

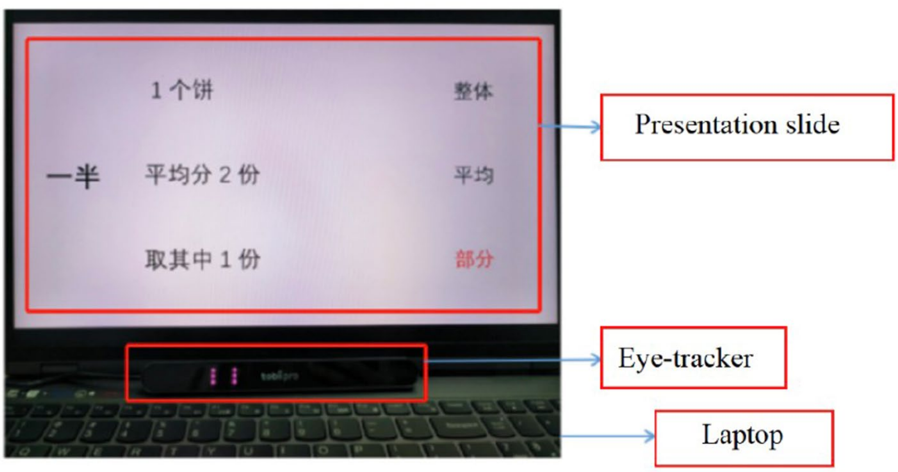


FIGURE 1
The device used in this study.

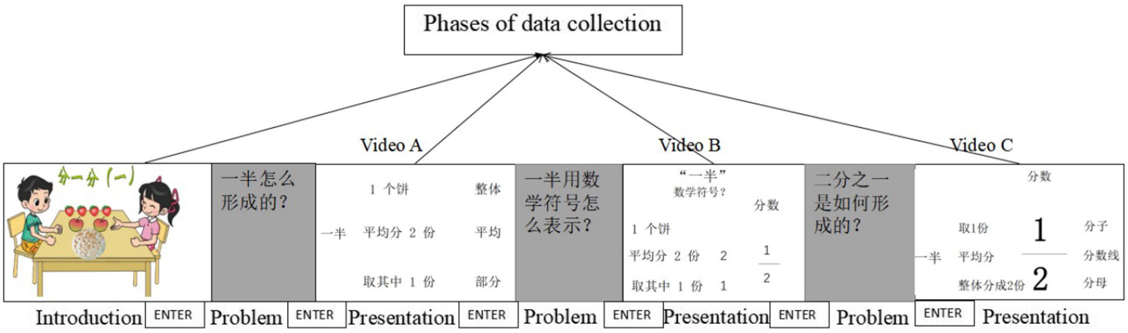


FIGURE 2
Flow diagram of the eye-tracking experiment (SaS example).

and 11, respectively, but there was only one slide in each video for the SaN group. To measure the instantaneous effect of PPT design on students' learning, all students were only given one chance to watch the videos and were not allowed to review them.

4.4. Data analysis

In this experiment, the independent variable was the technological presentation mode, and students' eye movements and learning performance were the dependent variables. To measure students' visual attention, we adopted the following eye-movement indicators: the first and total durations of fixation in the areas of interest (AOIs) and the regression time, which referred to the total fixation duration in which students related the current AOI to another AOI (Rayner et al., 2009; Jian et al., 2019).

Before analyzing the eye movements, we first defined the AOIs in each video for the three groups, as shown in Figure 3. Each AOI represented a keyword or key sentence in the teaching content. There were 24 AOIs for each group, which were marked AOI(A1), …, AOI(A7); AOI(B1), …, AOI(B6); and AOI(C1), …, AOI(C11). The capital letter in parentheses represents the name of the video (A, B, or C), and the number represents the serial number of the AOI in that video. Although there shows similar presenting content in some AOIs, it conveyed different meanings during teaching. The first video was used to introduce the real-life information of one half; the second video was used to guide students to extract mathematical information from life language, and convert them into mathematical symbols; the third video was used to connect real-life information and fraction elements, in order to foster students' deep understanding of the fraction concept.

In other words, there are diverse ways to express the fraction concept by relating various elements from different perspectives. In this study, the related elements were formed by two kinds of AOIs: one used literal mathematical language, and the other consisted of real-life information or symbolic mathematical language. Taking video A as an example, the fraction concept had three mathematical components: the whole AOI(A3), the average AOI(A5), and the part AOI(A7), which, respectively, connected to the students' own life experience, one piece of cake (AOI(A2)) divided into two equal parts (AOI(A4)), and one part removed (AOI(A6)). Thus, we obtained three related elements: AOI(A2–A3), AOI(A4–A5), and AOI(A6–A7). Finally, for three videos, there were 11 related elements, including AOI(A2–A3), AOI(A4–A5), AOI(A6–A7), AOI(B2–B3), AOI(B4–B5), AOI(C2–C3), AOI(C4–C5), AOI(C6–C7), AOI(C3–C10), AOI(C5–C9), and AOI(C7–C11). To analyze whether students made connections between two related AOIs, we used the regression time to characterize the duration of students' looking back to the former AOI when the teaching guided the students to the later AOI. For example, the fixation duration for AOI(A2–A3) represented the amount of time that students' attention dwelled on the former AOI(A2) when the teaching pointed to the later AOI(A3). The specific calculation approach was as follows: when the first fixation time on AOI(A2) preceded fixation on AOI (A3), the regression time for AOI (A2–A3) was the total fixation duration on AOI(A2) minus the first fixation duration on AOI(A2). When the first fixation time on AOI(A2) lagged behind that on AOI(A3), we used the total fixation duration on AOI(A2) to depict the regression time on AOI(A2–A3). Thus, regression time could be used to reveal students' ability to understand and integrate information (Schotter et al., 2014).

We gathered data about students' learning performance from the test that followed the experiment. All students were given as much time as needed to answer the questions and finish the test. The reliability values of the test scores for the three groups were 0.925 (SaS group), 0.890 (UaS group), and 0.917 (SaN group), respectively. We analyzed the first fixation duration, the total fixation duration, and the regression time of the students' eye movements across AOIs during learning. We used IBM® SPSS® 27.0 software to conduct the quantitative analysis. The significance level α was set at 0.05, and we used a one-way ANOVA to check for differences between the three groups.

5. Results

5.1. Significant differences in the attention of the three groups

The statistical data for the first and total fixation durations of eye movement confirmed that the structured stepwise presentation had a strong attention-guiding effect, which was consistent with H1.

Table 1 shows the students' first fixation duration for each AOI in the three groups. A one-way ANOVA showed that the first fixation duration of the three groups differed significantly.

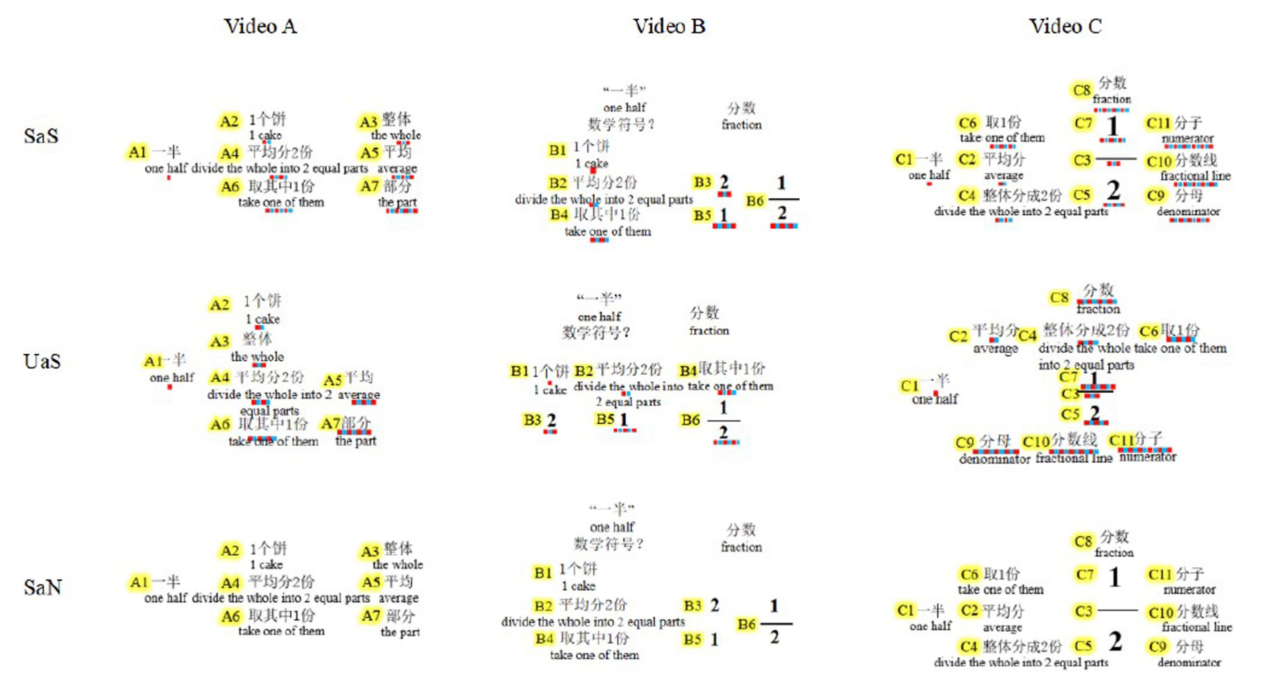


FIGURE 3
AOI Labels representing teaching content in each video for three groups. Color marking below each AOI for the SaS group and UaS group in the figure represents the sequence of AOI steps; there is no color marking in the SaN group, as the whole contents are presented at once with no steps.

Considering that several independent statistical tests were performed simultaneously, we conducted the Bonferroni correction by taking the alpha value for each comparison equal to 0.05/24 (0.002). The results showed significant differences for the following 12 AOIs: AOI(A1), $F(2, 93) = 13.691$, $p < 0.002$, $\eta^2 = 0.229$; AOI(A2), $F(2, 93) = 11.698$, $p < 0.002$, $\eta^2 = 0.203$; AOI(A3), $F(2, 93) = 13.405$, $p < 0.002$, $\eta^2 = 0.226$; AOI(A6), $F(2, 93) = 6.443$, $p = 0.002$, $\eta^2 = 0.123$; AOI(C1), $F(2, 93) = 18.207$, $p < 0.002$, $\eta^2 = 0.284$; AOI(C2), $F(2, 93) = 10.472$, $p < 0.002$, $\eta^2 = 0.185$; AOI(C5), $F(2, 93) = 9.498$, $p < 0.002$, $\eta^2 = 0.171$; AOI(C6), $F(2, 93) = 7.617$, $p = 0.001$, $\eta^2 = 0.142$; AOI(C8), $F(2, 93) = 7.868$, $p = 0.001$, $\eta^2 = 0.146$; AOI(C9), $F(2, 93) = 11.661$, $p < 0.002$, $\eta^2 = 0.202$; AOI(C10), $F(2, 93) = 8.690$, $p < 0.002$, $\eta^2 = 0.159$; and AOI(C11), $F(2, 93) = 16.686$, $p < 0.002$, $\eta^2 = 0.266$. Considering the equal variances assumed, we used the Least Significant Difference test (LSD) for multiple comparisons to test which group differs, and the results were marked with subscript letters in Table 1 at $p < 0.002$. The first fixation duration of the students in the SaS group was considerably longer than that of the other groups. The next longest duration was in the UaS group, and the first fixation time of the SaN group was the lowest among the three groups. Additionally, the number of AOIs with significant differences in the first fixation duration between the SaS group and the SaN group was higher than the number of AOIs with differences between other groups. Furthermore, it was hard to find significant differences between groups in video B. Therefore, in the SaS group, the students were stimulated by structured stepwise presentations, and their first attention was better than that of the other two groups.

Table 2 shows the total fixation duration of the students for each AOI. After the Bonferroni correction, there showed significant differences among the three groups for 19 AOIs at $p < 0.002$: AOI(A1), AOI(A2), AOI(A3), AOI(A4), AOI(A5), AOI(A6), AOI(A7), AOI(B2), AOI(C1), AOI(C2), AOI(C3), AOI(C4), AOI(C5), AOI(C6), AOI(C7), AOI(C8), AOI(C9), AOI(C10), and AOI(C11). There were no significant differences for the following AOIs: AOI(B1), AOI(B3), AOI(B4), AOI(B5), AOI(B6), and AOI(C1). We further performed the LSD test for multiple comparisons, as shown in Table 2 with subscript letters. There were significant differences in the total fixation duration of most AOIs between any two groups, especially between the SaS group and the SaN group. Specifically, the total fixation duration for most AOIs of the SaS group was significantly longer than that of the other groups, and the total fixation duration for all AOIs of the SaN group was the lowest among the three groups. In other words, the effect of stepwise presentations was higher than the effect of structured presentations. Even for the content with a low cognitive load, the stepwise presentation also worked to attract students' visual attention.

In sum, the structured stepwise presentation efficiently attracted the students' attention in fraction learning, thereby verifying H1 to a certain extent. There could be some possible reasons for no significant differences in the first fixation durations and total fixation durations of some AOIs among the three groups. Firstly, the average time of the first fixation duration was too short (less than 1 s), which may cause no statistically significant difference between the three groups. This also resulted in more AOIs with no significant difference in first fixation duration than that in total fixation duration. Secondly, most AOIs with no significant difference were from video B, which involves fewer abstract concepts and a lower intrinsic cognitive load required than the other two videos. In another word, this indicates that the structured stepwise presentation is particularly effective for the contents with high intrinsic cognitive load.

TABLE 1 One-way ANOVA of the first fixation duration for AOIs (time in milliseconds).

	SaS group		UaS group		SaN group		*F*	*p*	η^2
	M	SD	*M*	SD	*M*	SD			
Video A									
A1	0.818a	0.510	0.290b	0.530	0.335b	0.220	13.691	<0.001	0.229
A2	0.787a	0.532	0.635a	0.641	0.206b	0.213	11.698	<0.001	0.203
A3	0.612a	0.384	0.269b	0.231	0.265b	0.280	13.405	<0.001	0.226
A4	0.751a	0.438	0.749a	0.992	0.414a	0.331	2.783	0.067	0.057
A5	0.540a,b	0.416	0.586a	0.612	0.224b	0.282	5.905	0.004	0.114
A6	0.651a	0.442	0.560a,b	0.453	0.294b	0.327	6.443	0.002	0.123
A7	0.439a	0.261	0.431a	0.521	0.262a	0.309	2.185	0.118	0.045
Video B									
B1	0.359a	0.201	0.389a	0.292	0.353a	0.327	0.156	0.856	0.003
B2	0.685a	0.660	0.525a	0.485	0.390a	0.502	2.244	0.112	0.047
B3	0.324a	0.380	0.258a	0.364	0.091a	0.183	4.408	0.015	0.087
B4	0.692a	0.836	0.666a	0.734	0.353a	0.654	2.040	0.136	0.042
B5	0.314a	0.312	0.277a	0.458	0.056a	0.142	5.670	0.005	0.110
B6	0.737a	0.604	0.603a	0.742	0.759a	0.885	0.401	0.671	0.009
Video C									
C1	0.448a	0.443	0.109b	0.201	0.048b	0.084	18.207	<0.001	0.284
C2	0.662a	0.596	0.579a	0.657	0.095b	0.261	10.472	<0.001	0.185
C3	0.599a	0.682	0.464a,b	0.427	0.173b	0.346	5.915	0.004	0.114
C4	0.938a	0.722	0.696a,b	1.093	0.122b	0.242	9.378	<0.001	0.169
C5	0.433a,b	0.541	0.688a	0.751	0.097b	0.168	9.498	<0.001	0.171
C6	0.530a	0.485	0.217a,b	0.709	0.042b	0.133	7.617	0.001	0.142
C7	0.566a	0.784	0.570a	0.919	0.157a	0.528	3.103	0.050	0.063
C8	0.549a	0.787	0.256a,b	0.248	0.063b	0.219	7.868	0.001	0.146
C9	0.622a	0.497	0.403a,b	0.482	0.118b	0.201	11.661	<0.001	0.202
C10	0.823a	0.709	0.514a,b	0.526	0.249b	0.350	8.690	<0.001	0.159
C11	0.651a	0.492	0.279b	0.421	0.087b	0.217	16.686	<0.001	0.266

Values with different subscript letters in a row differ significantly at $p < 0.002$.

5.2. Significant differences in connecting corresponding AOIs

In three videos, we designed 11 indices to show the structured connections between corresponding AOIs in the presentation and used the regression time to evaluate the significance of attention across the two corresponding AOIs. The one-way ANOVA results showed significant differences between the three groups for the following seven related elements, as shown in Table 3, and the alpha value was set at 0.004 (0.05/11): AOI(B2–B3), $F(2, 93) = 11.094$, $p < 0.004$, $\eta^2 = 0.194$; AOI(B4–B5), $F(2, 93) = 9.531$, $p < 0.004$, $\eta^2 = 0.172$; AOI(C2–C3), $F(2, 93) = 12.665$, $p < 0.004$, $\eta^2 = 0.216$; AOI(C4–C5), $F(2, 93) = 12.949$, $p < 0.004$, $\eta^2 = 0.220$; AOI(C6–C7), $F(2, 93) = 6.961$, $p = 0.002 < 0.004$, $\eta^2 = 0.131$; AOI(C5–C9), $F(2, 93) = 7.241$, $p < 0.004$, $\eta^2 = 0.136$; and AOI(C7–C11), $F(2, 93) = 10.622$, $p < 0.004$, $\eta^2 = 0.188$. After the LSD test (see Table 3), the results showed significant differences in the regression time for more than half of the related AOIs between the SaS group and the UaS group or the SaN group. Specifically, the regression time of the SaS group was longer than that of the UaS group or the SaN group. That is to say, the effect of the structured presentations on the regression time is almost as significant as the effect of the stepwise presentations. Therefore, the structured stepwise presentations were helpful in guiding students to establish relationships between the corresponding elements. Furthermore, most of the regression time differences were found in videos B and C, in which students were experiencing the challenging process of expressing "half" in mathematical language. This indicates that students can make connections between symbolic mathematical language and real-life information or literal mathematical language with step oral guidance. In comparison, video A was mainly used to recall prior knowledge and introduce the concepts, namely whole, average, and part, which not involving different languages for students to relate. Thus, the differences in the regression time between the three groups were not significant as the differences in the fixation duration between them.

TABLE 2 One-way ANOVA of the total fixation duration for AOIs (time in milliseconds).

	SaS group		UaS group		SaN group		*F*	*p*	η^2
	M	SD	*M*	SD	*M*	SD			
Video A									
A1	4.364a	2.614	0.535b	0.754	0.766b	0.428	58.183	<0.001	0.558
A2	1.471a	0.817	1.049a	0.804	0.31b	0.306	23.391	<0.001	0.337
A3	1.616a	0.892	0.658b	0.749	0.494b	0.471	21.999	<0.001	0.324
A4	3.151a	1.136	1.893b	1.238	1.017c	0.938	29.316	<0.001	0.389
A5	1.777a	1.142	1.252a,b	1.128	0.488b	0.705	12.992	<0.001	0.220
A6	2.465a	1.046	1.372b	1.046	0.613c	0.717	30.320	<0.001	0.397
A7	2.067a	1.874	1.591a,b	1.626	0.612b	1.137	7.043	0.001	0.133
Video B									
B1	1.143a	0.773	1.171a	0.946	0.780a	0.727	2.051	0.134	0.043
B2	2.119a	1.161	1.347a,b	1.238	0.696b	0.74	14.006	<0.001	0.233
B3	0.744a	0.832	0.393a	0.645	0.217a	0.66	4.408	0.015	0.087
B4	1.779a	1.412	1.248a	1.26	0.836a	0.992	4.636	0.012	0.092
B5	0.672a	0.799	0.500a,b	0.822	0.091b	0.277	6.113	0.003	0.117
B6	2.679a	2.139	1.467a	1.899	1.508a	1.405	4.401	0.015	0.087
Video C									
C1	0.771a	0.701	0.160b	0.368	0.041b	0.078	23.121	<0.001	0.335
C2	1.465a	0.753	0.939a	0.849	0.139b	0.372	29.655	<0.001	0.392
C3	1.881a	1.324	0.908b	0.859	0.201b	0.367	25.848	<0.001	0.360
C4	2.239a	0.991	1.171b	1.294	0.189c	0.388	35.430	<0.001	0.435
C5	0.799a	0.878	0.889a	0.868	0.120b	0.207	10.854	<0.001	0.191
C6	1.462a	1.017	0.267b	0.729	0.042b	0.133	34.865	<0.001	0.431
C7	1.593a	1.027	0.958a	1.029	0.217b	0.691	17.351	<0.001	0.274
C8	1.237a	1.208	0.481b	0.609	0.063b	0.219	17.993	<0.001	0.281
C9	1.795a	0.941	0.799b	0.806	0.213b	0.460	34.741	<0.001	0.430
C10	1.775a	1.005	0.892b	0.886	0.388b	0.598	21.672	<0.001	0.320
C11	1.903a	1.230	0.338b	0.573	0.088b	0.217	48.906	<0.001	0.515

Values with different subscript letters in a row differ significantly at $p < 0.002$.

5.3. Structured stepwise presentation achieved better learning performance

The total test score for students' learning performance was 30. All the students' answers were reliable and valid. Unfortunately, the one-way ANOVA results showed no statistically significant differences among the three groups, $F(2, 92) = 2.939$, $p = 0.058$, $\eta^2 = 0.06$. There could be other important factors that influence students' learning performance, such as emotions, metacognitive abilities, and personality traits. We have to consider the delayed effect of the concentrating process on the final learning achievement. Despite this fact, the average score of students in the SaS group ($M = 23.566$, SD = 6.612) was still higher than that in the UaS group ($M = 19.606$, SD = 7.697), and the SaN group ($M = 19.875$, SD = 7.469). But the score for the UaS group was close to the score for the SaN group. In another word, the students in the SaS group who paid more attention to key information and related elements achieved better learning performance to some extent. This indicates that the structured stepwise presentation could have a beneficial effect on the students' fraction learning.

6. Discussion and conclusion

In this study, we investigated the effect of structured stepwise presentations on students' learning of fractions using eye-tracking technology. The results showed that the structured stepwise presentations played an important synergistic role in directing students' attention and promoting learning performance.

6.1. Dynamic stimulation to guide and strengthen students' attention

The results of this study showed that the mean fixation duration of students in the SaS and UaS groups, who learned with stepwise presentations, was greater than that of the students in the SaN group.

TABLE 3 One-way ANOVA of the regression time between corresponding AOIs (time in milliseconds).

	SaS group		UaS group		SaN group		F	p	η^2
	M	SD	M	SD	M	SD			
Video A									
A2–A3	0.475a	0.541	0.198a	0.287	0.383a	0.477	3.165	0.047	0.064
A4–A5	1.137a	0.906	1.358a	1.116	0.633a	0.740	5.075	0.008	0.099
A6–A7	0.973a	1.698	0.703a	0.734	0.620a	0.940	0.752	0.474	0.016
Video B									
B2–B3	0.702a	0.654	0.912a	1.027	0.102b	0.203	11.094	<0.001	0.194
B4–B5	0.138a	0.231	0.557a	0.749	0.081a	0.238	9.531	<0.001	0.172
Video C									
C2–C3	0.902a	0.756	0.481a,b	0.554	0.171b	0.363	12.665	<0.001	0.216
C4–C5	0.588a	0.652	0.156a	0.356	0.048a	0.218	12.949	<0.001	0.220
C6–C7	0.174a	0.309	0.027a	0.109	0.013a	0.042	6.961	0.002	0.131
C3–C10	0.083a	0.322	0.015a	0.050	0.019a	0.110	1.165	0.316	0.025
C5–C9	0.211a	0.287	0.104a,b	0.202	0.016b	0.052	7.241	0.001	0.136
C7–C11	0.625a	0.796	0.273a,b	0.432	0.021b	0.092	10.622	<0.001	0.188

Values with different subscript letters in a row differ significantly at $p < 0.004$.

In other words, more attention was allocated to the AOIs in the stepwise presentation groups. These results confirmed that stepwise presentations as dynamic signals favorably guided students' attention, which was consistent with the existing research (De Koning et al., 2009; Tso et al., 2011).

In fact, the stepwise effect was enhanced by dynamic stimulation. The stepwise format intensified the dynamic nature of the presentation, which stimulated students' attention. Since a new object was generated in each step, the learning input could be sequentially captured by students' attention and effectively prevent them from being distracted by external factors (Boucheix and Guignard, 2005; Wouters et al., 2008; Lowe and Schnotz, 2014). Taking the experimental group's teaching as an example, since the teacher presented each slide in sequence, key information was shown to the students gradually, which provided dynamic signals for the students and functioned as a stimulus to attract the students' attention to the cues. Thus, in the SaS and UaS groups, the students' first fixation duration on each AOI was longer than that of students in the SaN group. As prior studies have shown, dynamic signaling not only captures students' attention but also focuses their attention on key processes and prevents distraction (Wouters et al., 2008; Boucheix et al., 2013).

Stepwise presentations can effectively support the teaching of complex content and provide scaffolding for students' knowledge construction by segmenting the content into several related parts (Lee et al., 2018). In the SaS group, the teaching presentation showed multiple representations of the fraction concept. Each scaffolding sub-concept involved key elements for learning fractions, including real-life applications, literal mathematical language, and symbolic mathematical language, which were presented in videos B and C. As Table 3 showed, the effect of relating the corresponding elements was better in the SaS group than that in the SaN group. Thus, stepwise presentations can help students to identify the connection between related information blocks, which provided a constructive learning process that transformed real-life information into mathematical language and presented an overall concept-forming process (Chen et al., 2016; Rey et al., 2019). The eye-tracking data in this study suggested that the stepwise presentations were effective to direct students' attention. In fact, the stepwise effect can release more cognitive resources for students to construct knowledge (Rey et al., 2019). In each stepwise connecting block, the teacher provided a construction cue using dynamic stimuli. Thus, differences between the upper and lower blocks could be easily observed by the students. Therefore, stepwise presentations for abstract information can be understood in terms of effective connections between accessible sub-elements, and problems in the learning of fractions, such as situation distortion, monotonous content, and lack of a constructive process, can be solved.

6.2. Structured presentations are conducive to the formation of concepts

This study showed that structured presentations better encourage students to make connections between related elements. A structured presentation visualizing the potential conceptual relationships, especially in videos B and C, had a significant effect on students' attention to connect related elements in the SaS group, as shown in Table 3. It revealed that students' attention was attracted by structured cueing, which helped them integrate related elements (Schotter et al., 2014; Eskenazi and Folk, 2017). The results confirm the effectiveness of the signaling principle for visualizing semantic and grammatical rules. Structured presentations convert seeing into understanding, allowing semantics and grammar to be distinguished by visualizing the relationships between learning concepts (Miao et al., 2000; De Koning et al., 2010). Thus, in the SaS group, students easily identified related elements and had a stronger disposition to fixate on them than in the other groups. The results confirmed that effective visual signals enable connections to be made and knowledge to be clearly visualized, reducing the burden of interpretation (De Koning et al., 2010).

A structured presentation also guides students to form concepts step-by-step and supports them in a concept-forming process that includes exploration, reflection, and discovery. Visual structured knowledge representations not only help students internalize and reflect on their knowledge but also support negotiation and exploration (Miao et al., 2000). In this study, the SaS group had better attention and more effectively made connections between corresponding AOIs, indicating that these students experienced a fruitful learning process. This aligns with research showing that regression time is informational and beneficial for students' understanding and connection-making (Rayner et al., 2009; Schotter et al., 2014; Jian et al., 2019). Furthermore, based on the cueing principle, according to presented visual stimuli, such as color, information blocks, and identifiers, students constantly adjust the direction of their attention to understand the content and reflect on it, which is beneficial for promoting self-regulated learning (Ferrara and Butcher, 2011). Therefore, students' active construction of knowledge can be supported by structured presentations.

In addition, structured presentations give students a deep impression of their knowledge. In this study, regression time reflected the cognitive process of integrating, rather than simply extracting information (Rayner et al., 2009; Jian et al., 2019), and the regression for relevant information was applied more often by high-level learners (Mason et al., 2013). In this study, compared to other groups, the students in the SaS group obtained higher scores on the test that included several recall questions, and this result indicated that structured knowledge is well remembered (Carter and Kathy, 1994). The prior study suggested that visual reinforcement stimulation is more conducive to processing text and symbols than language stimulation during the process of short-term memory formation (Liang, 2014). Structured presentations, by visually presenting the structure of knowledge and imaginable characteristics, help students form refined short-term memory and knowledge schemata, which are easily extracted, retained, and recalled. Dedicated short-term memory supports students' effective long-term memory storage, which can provide the basis for forming strong knowledge structures. Therefore, structured knowledge presentations help students realize the fine processing of knowledge and the effective construction of knowledge schemata under conditions of limited working memory.

6.3. Conclusion

In this study, we confirmed that the structured stepwise presentation drew students' attention, guided them to understand the cognitive process of learning about fractions, and promoted their learning performance. Furthermore, the results showed that stepwise presentations worked better than structured presentations to attract students' visual attention. Structured presentations are effective for complex concepts with high-element interactivity, but are not as remarkable as for contents with low cognitive load. Furthermore, stepwise presentations can accelerate the effect of structured presentations, which are adaptive to the content with different cognitive levels. The findings of this study reveal that a good presentation is vital for helping students construct knowledge and guiding them to deeply understand new content. In order to achieve high-quality fraction teaching, we suggested that teachers should combine structured presentations and stepwise presentations together to provide dynamic cues in a well-organized way, and further clarify the relationships between different to support students in forming clear knowledge schemata.

In sum, we examined the rationale of structured stepwise presentation based on the principle of segmenting and cueing. However, there are some limitations in this study. We used only eye-movement indicators to analyze the students' attention. Future studies could find other ways to evaluate students' attention and investigate whether an extraneous cognitive load can be decreased by structured stepwise presentations. In terms of the control of research variables, it is necessary to investigate the learning characteristics of each research object in the future and explore which kind of students is more effective under the guidance of structured stepwise presentations. Regarding the learning test, this study examined students' learning performance based on memory understanding of fraction concepts. Whether structured stepwise presentations can be used to solve more complex cognitive problems and promote students' advanced thinking should be examined in future research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XS designed and carried out the experiment. XS and RL analyzed the data and wrote the manuscript text. YL provided ideas for this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Humanities and Social Science Foundation of Ministry of Education of China (22XJA880009), by the Science and Technology Foundation Project of Shaanxi (2022QFY01-05).

Acknowledgments

We thank all students who participated in this study. We are grateful to Prof. Mingjang Chen at Hsinchu Chiao Tung University for his advice on research design, Prof. Zhongling Pi at Shaanxi Normal University for proofreading, and Associate Prof. Dong Yang for his assistance with the data collection. In addition, we would like to thank the reviewers for their constructive comments on previous drafts of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Boucheix, J. M., and Guignard, H. (2005). What animated illustrations can improve technical document comprehension in young students? Format, signaling, and control of the presentation. *Eur. J. Psychol. Educ.* 20, 369–388. doi: 10.1007/BF03173563

Boucheix, J. M., and Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learn. Instr.* 20, 123–135. doi: 10.1016/j.learninstruc.2009.02.015

Boucheix, J. M., Lowe, R. K., Putri, D. K., and Groff, J. (2013). Cueing animations: dynamic signaling aids information extraction and comprehension. *Learn. Instr.* 25, 71–84. doi: 10.1016/j.learninstruc.2012.11.005

Bransford, J., Brophy, S., and Williams, S. (2000). When computer technologies meet the learning sciences: issues and opportunities. *J. Appl. Dev. Psychol.* 21, 59–84. doi: 10.1016/S0193-3973(99)00051-9

Carter, K. (1994). Preservice teachers' well-remembered events and the acquisition of event-structured knowledge. *J. Curric. Stud.* 26, 235–252. doi: 10.1080/0022027940260301

Chen, M. J., Lee, C. Y., Lei, K. H., and Yang, C. Y. (2016). The effects of using stepwise attention-guiding multimedia instruction to learn properties of tangents of circles. *Taiwan J. Math. Educ.* 3, 1–30. doi: 10.6278/tjme.20161005.001

Cierniak, G., Scheiter, K., and Gerjets, P. (2009). Explaining the split-attention effect: is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Comput. Hum. Behav.* 25, 315–324. doi: 10.1016/j.chb.2008.12.020

De Koning, B. B., Tabbers, H. K., Rikers, R. M., and Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Appl. Cogn. Psychol.* 21, 731–746. doi: 10.1002/acp.1346

De Koning, B. B., Tabbers, H. K., Rikers, R. M., and Paas, F. (2009). Towards a framework for attention cueing in instructional animations: guidelines for research and design. *Educ. Psychol. Rev.* 21, 113–140. doi: 10.1007/s10648-009-9098-7

De Koning, B. B., Tabbers, H. K., Rikers, R. M., and Paas, F. (2010). Attention guidance in learning from a complex animation: seeing is understanding? *Learn. Instr.* 20, 111–122. doi: 10.1016/j.learninstruc.2009.02.010

Doolittle, P. E., and Altstaedter, L. L. (2009). The effect of working memory capacity on multimedia learning: does attentional control result in improved performance? *J. Res. Innov. Teach.* 2, 7–25.

Dunsworth, Q., and Atkinson, R. K. (2007). Fostering multimedia learning of science: exploring the role of an animated agent's image. *Comput. Educ.* 49, 677–690. doi: 10.1016/j.compedu.2005.11.010

Eskenazi, M. A., and Folk, J. R. (2017). Regressions during reading: the cost depends on the cause. *Psychon. Bull. Rev.* 24, 1211–1216. doi: 10.3758/s13423-016-1200-9

Ferrara, L., and Butcher, K. R.. (2011). Visualizing feedback: using graphical cues to promote self-regulated learning. In: *Proceedings of the thirty-third annual conference of the cognitive science society* No. 30, p. 33.

Gross, A. G., and Harmon, J. E. (2009). The structure of power point presentations: the art of grasping things whole. *IEEE Trans. Prof. Commun.* 52, 121–137. doi: 10.1109/TPC.2009.2020889

Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Comput. Hum. Behav.* 32, 47–53. doi: 10.1016/j.chb.2013.11.013

Jamet, E., Gavota, M., and Quaireau, C. (2008). Attention guiding in multimedia learning. *Learn. Instr.* 18, 135–145. doi: 10.1016/j.learninstruc.2007.01.011

Jian, Y. C., Su, J. H., and Hsiao, Y. R. (2019). Differentiated processing strategies for science reading among sixth-grade students: exploration of eye movements using cluster analysis. *Comput. Educ.* 142:103652. doi: 10.1016/j.compedu.2019.103652

Kalra, P. B., Hubbard, E. M., and Matthews, P. G. (2020). Taking the relational structure of fractions seriously: Relational reasoning predicts fraction knowledge in elementary school children. *Contemp. Educ. Psychol.* 62:101896. doi: 10.1016/j.cedpsych.2020.101896

Kurby, C. A., and Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends Cogn. Sci.* 12, 72–79. doi: 10.1016/j.tics.2007.11.004

Lee, C. Y., Lei, K. H., Chen, M. J., Tso, T. Y., and Chen, I. P. (2018). Enhancing understanding through the use of structured representations. Eurasia journal of mathematics science and technology education, 14(5), 1875-1886. *Contemp. Educ. Psychol.* 14:101896. doi: 10.29333/ejmste/85424

Lei, K. H., Chen, M. J., Lee, C. Y., Tso, T. Y., and Lin, S. L. (2017). Multimedia instruction presented by integrated context to enhance understanding of compass-and-straightedge construction. *Eur. J. Math. Sci. Technol. Educ.* 13, 3735–3752. doi: 10.12973/eurasia.2017.00756a

Liang, N. J. (2014). *Contemporary cognitive psychology*. Shanghai: Shanghai Education Press.

Lin, F. L., Huang, M. F., and Leu, Y. C. (1996). A developmental study on learning and teaching of beginning fractions. *Chin. J. Sci. Educ.* 4, 161–196.

Lortie-Forgues, H., Tian, J., and Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Dev. Rev.* 38, 201–221. doi: 10.1016/j.dr.2015.07.008

Lowe, R. K., and Schnotz, W. (2014). "Animation principles in multimedia learning," in *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning*. ed. R. E. Mayer (New York: Cambridge University Press), 513–546.

Mason, L., Tornatora, M. C., and Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Comput. Educ.* 60, 95–109. doi: 10.1016/j.compedu.2012.07.011

Mautone, P. D., and Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *J. Educ. Psychol.* 93, 377–389. doi: 10.1037/0022-0663.93.2.377

Mayer, R. E. (2014). "Principles for managing essential processing in multimedia learning: segmenting, pretraining, and modality principles" in *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning*. ed. R. E. Mayer (New York, NY: Cambridge University Press), 345–367.

Mayer, R. E., and Pilegard, C. (2014). "Principles for managing essential processing in multimedia learning: segmenting, pre-training, and modality principles" in *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning*. ed. R. E. Mayer (New York: Cambridge University Press), 316–344.

Miao, Y., Holst, S., Holmer, T., Fleschutz, J., and Zentel, P.. (2000). *An activity-oriented approach to visually structured knowledge representation for problem-based learning in virtual learning environments*. International Conference on the Design of Cooperative Systems, pp. 303–318

Mitchell, C. E., and Miller, L. D. (2010). Tech prep academics: using real life connections to develop scientific and mathematical literacy. *Sch. Sci. Math.* 95, 417–422. doi: 10.1111/j.1949-8594.1995.tb10195.x

Obersteiner, A., Hoof, J. V., Verschaffel, L., and Dooren, W. V. (2015). Who can escape the natural number bias in rational number tasks? A study involving students and experts. *Br. J. Psychol.* 107, 537–555. doi: 10.1111/bjop.12161

Ozcelik, E., Arslan-Ari, I., and Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Comput. Hum. Behav.* 26, 110–117. doi: 10.1016/j.chb.2009.09.001

Park, O. C., and Hopkins, R. (1992). Instructional conditions for using dynamic visual displays: a review. *Instr. Sci.* 21, 427–449. doi: 10.1007/BF00118557

Pi, Z., Xu, K., Liu, C., and Yang, J. (2020). Instructor presence in video lectures: eye gaze matters, but not body orientation. *Comput. Educ.* 144:103713. doi: 10.1016/j.compedu.2019.103713

Rayner, K., Pollatsek, A., Liversedge, S. P., and Reichle, E. D. (2009). Eye movements and non-canonical reading: comments on. *Vis. Res.* 49, 2232–2236. doi: 10.1016/j.visres.2008.10.013

Reinhold, F., Hoch, S., Werner, B., Richter-Gebert, J., and Reiss, K. (2020). Learning fractions with and without educational technology: what matters for high-achieving and low-achieving students? *Learn. Instr.* 65, –101264. doi: 10.1016/j.learninstruc.2019.101264

Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T. H., and Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educ. Psychol. Rev.* 31, 389–419. doi: 10.1007/s10648-018-9456-4

Richter, J., Scheiter, K., and Eitel, A. (2016). Signaling text-picture relations in multimedia learning: a comprehensive meta-analysis. *Educ. Res. Rev.* 17, 19–36. doi: 10.1016/j.edurev.2015.12.003

Rieber, L. P. (1990). Animation in computer-based instruction. *Educ. Technol. Res. Dev.* 38, 77–86. doi: 10.1007/BF02298250

Schneider, S., Beege, M., Nebel, S., and Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educ. Res. Rev.* 23, 1–24. doi: 10.1016/j.edurev.2017.11.001

Schotter, E. R., Tran, R., and Rayner, K. (2014). Don't believe what you read (only once): comprehension is supported by regression during reading. *Psychol. Sci.* 25, 1218–1226. doi: 10.1177/0956797614531148

Seufert, T., and Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Appl. Cogn. Psychol.* 20, 321–331. doi: 10.1002/acp.1248

Smith, B., and Shimeld, S. (2014). Using pictorial mnemonics in the learning of tax: a cognitive load perspective. *Higher Educ. Res. Dev.* 33, 565–579. doi: 10.1080/07294360.2013.841652

Spanjers, I. A. E., van Gog, T., and van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educ. Psychol. Rev.* 22, 411–423. doi: 10.1007/s10648-010-9135-6

Stiller, K. D., Petzold, K., and Zinnbauer, P. (2011). Presentation time concerning system-paced multimedia instructions and the superiority of learner pacing. *Australas. J. Educ. Technol.* 27, 693–708. doi: 10.14742/ajet.945

Treisman, A., and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cogn. Psychol.* 14, 107–141. doi: 10.1016/0010-0285(82)90006-8

Tso, T. Y., Lu, F. L., Tzeng, S. C., Wu, H. M., Chen, M. J., and Tan, T. (2011). The effect of reducing task complexity on the reading of geometric proofs by experts and novices in a segmented way. N.C. *J. Educ. Psychol.* 43, 291–314. doi: 10.6251/BEP.20110517

Tversky, B., Heiser, J., Mackenzie, R., Lozano, S., and Morrison, J. (2008). "Enriching animations," in *Learning with animation: Research implications for design*. eds. R. K. Lowe and W. Schnotz (Cambridge, UK: Cambridge University Press), 263–285.

Van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. R. E. Mayer, *Cambridge Handbook of Multimedia Learning*, Cambridge, UK: Cambridge University Press, 263–278.

Wang, F., Duan, Z., and Zhou, Z. (2013). Attention guidance in multimedia learning: the role of cueing. *Adv. Psychol. Sci.* 21, 1430–1440. doi: 10.3724/SP.J.1042.2013.01430

Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501. doi: 10.1038/nrn1411

Wouters, P., Paas, F., and van Merriënboer, J. J. (2008). How to optimize learning from animated models: a review of guidelines based on cognitive load. *Rev. Educ. Res.* 78, 645–675. doi: 10.3102/0034654308320320

Yantis, S., and Jonides, J. (1996). Attentional capture by abrupt onsets: new perceptual objects or visual masking? *J. Exp. Psychol. Hum. Percept. Perform.* 22, 1505–1513. doi: 10.1037//0096-1523.22.6.1505

TYPE Original Research
PUBLISHED 12 May 2023
DOI 10.3389/fpsyg.2023.1145442

OPEN ACCESS

EDITED BY
Shuhua An,
California State University, Long Beach, United States

REVIEWED BY
Lisa Bendixen,
University of Nevada, Las Vegas, United States
Gregory Siy Ching,
National Chengchi University, Taiwan

*CORRESPONDECE
Chang-Hua Chen
✉ cchen72@cc.ncue.edu.tw

RECEIVED 16 January 2023
ACCEPTED 26 April 2023
PUBLISHED 12 May 2023

CITATION
Lin C-H and Chen C-H (2023) Development, testing, and application of a mathematics learning scale of self-direction.
Front. Psychol. 14:1145442.
doi: 10.3389/fpsyg.2023.1145442

Development, testing, and application of a mathematics learning scale of self-direction

Chia-Hui Lin[1] and Chang-Hua Chen[2]*

[1]Office of Teacher Education and Careers Service, National Taichung University of Education, Taichung, Taiwan, [2]Graduate Institute of Science Education, National Changhua University of Education, Changhua, Taiwan

Many countries' curriculum reforms focus on developing the next generations' competencies of self-directed learning (SDL) to address rapid social changes and sustainable environmental development. Taiwan's curriculum reform corresponds with the global trend in education. The latest curriculum reform, which proposed a 12-year basic education, was implemented in 2018 and included SDL explicitly in its guidelines. The reformed curriculum guidelines have been followed for over 3 years. Thus, it is necessary to conduct a large-scale survey to examine its impact on Taiwanese students. However, existing research instruments help provide a generalized analysis of SDL and have yet to be designed specifically for SDL of mathematics. Therefore, we developed a mathematics SDL scale (MSDLS) and examined its reliability and validity in this study. Subsequently, MSDLS was utilized to investigate Taiwanese students' SDL of mathematics. The MSDLS consists of four sub-scales with 50 items. It has acceptable reliability, validity, and measurement invariance across gender and grade groups. The MSDLS was administered online to 5,575 junior high school students, and 5,456 valid responses were collected. The findings highlight the gender and grade differences in SDL of mathematics. Male students are higher than female students in many factors. It is noted that the SDL in mathematics does not increase with grade. In sum, the MSDLS is a helpful instrument for examining secondary school students' SDL of mathematics.

KEYWORDS

learning scale, mathematics education, self-directed learning (SDL), self-direction, self-reguated learning

1. Introduction

In response to the rapid changes in society and the environment in the 21st century, countries around the world are pursuing quality education and social justice. They consider what and how the next generation should learn (Senge et al., 2000). For example, the New Zealand Curriculum identifies five key competencies. Among the five key competencies, managing self means that students should establish personal learning goals, set high standards, make plans, manage projects, and have strategies for meeting challenges (New Zealand Ministry of Education, 2007). Hong Kong has been promoting the "Learning to Learn" curriculum reform since 2001, updated the curriculum framework to "Learning to Learn 2.0" in 2017, and added self-directed learning (SDL) abilities (H.K. Curriculum Development Council, 2017), hence enabling students to become independent and self-directed learners. The Organization for Economic Cooperation and Development (OECD) has included self-direction in the mathematical assessment framework of the Program for International Student Assessment (PISA) as a key 21st century skill (OECD, 2018). In other words, SDL has gradually become the common language for curriculum reform in many countries or regions.

Taiwan's curriculum reform corresponds with the global trend in education. This is the first time that SDL has been included in the Curriculum Guidelines of 12-Year Basic Education: General Guidelines (referred to as general guidelines) and SDL is considered a prerequisite for lifelong learning and whole-person education (National Academy for Educational Research, 2018). The general guidelines use the core competencies as the basis of curriculum development. These competencies are divided into three broad dimensions: spontaneity, communication and interaction, and social participation. SDL functions as the purpose and process of the development of core competencies. As an element of core competencies, SDL demonstrates learners' knowledge, skills, and attitudes related to learning after completing 12-year basic education and lays the foundation for lifelong learning and career development. In addition, SDL becomes an indispensable process for the development of core competencies. Students set learning goals and adopt strategies to achieve these goals through SDL. In this process, they continuously evaluate the gap between the current scenario and the goals and revise their strategies to reduce the gap and move towards the goals.

The general guidelines encourage elementary and secondary schools to integrate SDL into school-based curricula (National Academy for Educational Research, 2018, pp. 11, 13) and alternative learning periods (pp. 21, 28). Schools and teachers should guide students in learning how to learn, including general learning strategies, domain learning strategies, and metacognitive strategies (National Academy for Educational Research, 2018, p. 48). High schools should include the spirit and practice of SDL in school-based curriculum development and key items of school evaluation and school visits. In Taiwan's curriculum reform, the general guidelines serve as a guiding document of curriculum development and guide the construction and design of domain-specific curriculum guidelines (referred to as domain guidelines). Therefore, the spirit of SDL is explicitly included or incorporated into the domain guidelines. For example, the core competencies of the mathematics curriculum guidelines at the junior high school level are "students should be able to identify the connection between real-life problems and mathematics, develop problem-solving strategies from multiple perspectives, and apply them in real-life scenarios" (Ministry of Education, 2018, p. 3). Including SDL in the mathematics curriculum guidelines prepares students for university study and career development and addresses the problem of low motivation, low self-confidence, and negative attitudes of Taiwanese students in previous international mathematics learning achievement assessments.

Although Taiwan's curriculum reform emphasizes SDL, SDL is not clearly defined in the general and domain guidelines. Furthermore, although PISA (OECD, 2018) includes self-direction in the 2022 mathematical literacy assessment framework, it does not provide the definition and related aspects of SDL in mathematics. Therefore, this study aimed to construct a mathematics SDL scale and examine its reliability. This study conducted a survey in Taiwan and examined the differences in SDL in mathematics across demographic variables. As the curriculum reform of 12-year basic education was implemented more than 3 years ago, it is necessary to develop a research instrument to investigate the current scenario of SDL in mathematics among Taiwanese students. This paper aims to clarify the concept of SDL in mathematics in the context of curriculum reform and propose a measurement tool to facilitate the understanding and discussion of this 21st century skill in the academic world. From a practical perspective, this paper identifies possible problems in implementing the new curriculum guideline and proposes some educational policy recommendations.

Thus, the objectives of this investigation are dual-fold: (a) fashioning a mathematical learning scale of self-direction and evaluating its reliability and validity and (b) exploring the SDL variance in mathematics with respect to gender and scholastic levels.

2. Literature review

2.1. Definitions of self-directed learning (SDL)

SDL has received much academic interest over the past four decades (Panadero, 2017) and has had various definitions (de Bruin and van Merriënboer, 2017; Brandt, 2020). Knowles (1975) defined SDL as a means for individuals to actively diagnose their own learning requirements, set learning goals, find relevant learning resources, select and apply appropriate learning strategies, and evaluate learning outcomes. SDL is often confused with self-regulated learning (SRL), but SDL has been considered a broader concept encompassing SRL (Saks and Leijen, 2014). The general guidelines also consider SRL as a sub-process of SDL and emphasize its development in elementary schools (National Academy for Educational Research, 2018). SRL provides learners with the self-direction to use their mental abilities to regulate their learning behaviors, transform their mental abilities into academic performance, and perform self-reflection after learning tasks, thereby preparing themselves for the next learning task (Zimmerman, 2008). Therefore, SDL is closely related to metacognition. Some scholars, such as Mevarech et al. (2018), believed that the two terms can be considered synonymous in mathematics classrooms.

Although academic perspectives on SDL are divergent, commonalities exist among the views that focus on how learners organize, guide, and monitor individual learning activities. Steffens (2015) believed that the object of SDL theory is not the learning itself, and it is closer to a meta-theory focusing on how to learn. It involves learning motivation, learning strategies, learning regulation, and learning resource management (National Academies of Sciences, Engineering, and Medicine, 2018). Zimmerman (2000) believed that the key to SDL is self-regulation, which is a three-stage cycle that includes motivation and self-efficacy, strategy development and implementation, and self-monitoring and self-adjusting. Self-directed learners demonstrate excellent self-control abilities in cognitive and behavioral strategies, and these abilities are reflected in their active participation in learning activities and excellent academic achievements (Schunk and Rice, 1987, 1991). In addition, self-directed learners can manage learning resources and seek help when they encounter challenges (Schunk, 2012).

The mathematics education community has been exploring SDL for a long time, focusing on planning, monitoring, evaluating, and reflecting in problem-solving. In the influential book, "How to Solve It," Polya (1957) proposed a four-step model for problem-solving: understanding the problem, developing a problem-solving strategy, implementing the strategy, and reviewing the solution. First, the problem solver should understand the problem, identify the type of problem by its conditions, devise a problem-solving plan, select a problem-solving strategy, implement the problem-solving plan, and monitor the effectiveness of the strategy. After obtaining the solution, the problem solver should check whether the solution is correct and whether a better solution exists. Schoenfeld (1985) extended Polya's view of problem-solving by proposing a six-step model: reading the problem, analyzing the problem, exploring the known conditions and goals of the problem, devising a problem-solving plan, implementing and monitoring the plan, and evaluating the solution. He believed that mathematical problem-solving includes four elements: knowledge resources (resources), strategy exploration (heuristics), control (control), and belief systems (belief systems). The control is self-regulation, which is the key to the mathematical problem-solving process (Schoenfeld, 1992). Schoenfeld's (1985) research became the theoretical basis for many subsequent studies related to SDL in mathematics (Mevarech et al., 2018).

2.2. Previous views on SDL

SDL covers a wide range of concepts. Some scholars discuss it from a holistic perspective and others focus on a particular direction (Brandt, 2020). Bandura's (1977) social cognitive theory proposed the concept of SDL at an early stage. He believed that SDL is the result of individuals' external behaviors and the process by which individuals monitor, evaluate, and modify their cognition, motivation, emotions, and behaviors. Some scholars (such as Winne, 1996) focused on cognitive processes, such as whether learners have the necessary knowledge, abilities, and strategies to achieve goals. Self-directed learners demonstrate excellent self-control abilities in cognitive and behavioral strategies, and these abilities are reflected in their active participation in learning activities and excellent academic achievements (Schunk and Rice, 1987, 1991). Some scholars (such as Corno, 2001) indicated that many students fail to perform SDL not because they lack the relevant knowledge and ability, but because they lack the willingness to use it in the classroom. Therefore, motivation and attitude have an important role in SDL. When learners enter a classroom and begin learning, their personal goals, self-efficacy, values, and emotions determine how they approach the task (namely learning strategies) and how much effort they apply to it (Schunk and Mullen, 2013; Wiliam, 2018). Most studies on SDL use self-report scales (Saks and Leijen, 2014), and the commonly used scales include the self-directed learning readiness scale (SDLRS) developed by Guglielmino (1977) and the motivated strategies for learning questionnaire (MSLQ) developed by Pintrich et al. (1991). The SDLRS and MSLQ assess general SDL skills. They have been translated into multiple languages and implemented in many countries and have excellent reliability and validity.

Many studies indicated SDL abilities can be acquired and improved (Gabrielle et al., 2006; Amey, 2008; Dignath et al., 2008). Chen et al. (2021) conducted a large-scale survey using SDLRS and observed that the 12-year basic education curriculum assisted in improving the SDL readiness of high school students. Teaching students SDL skills can enhance their mathematical reasoning (Schoenfeld, 1992; Mevarech and Kramarski, 2014). Note that although SDL is considered a domain-general skill, it can only be effectively taught and developed in a domain-specific context, e.g., mathematics (Kirschner and Hendrick, 2020; Schunk, 2020). The more mathematical knowledge and strategies one has, the more effective one can self-monitor and self-regulate in mathematics learning (Schoenfeld, 2014). Developing and applying SDL skills in domain-specific learning, such as using general learning strategies and problem-solving strategies in mathematics learning, can improve mathematics learning performance and facilitate learning transfer (Schunk, 2012). Research demonstrates that K-12 students and adults who plan, monitor, evaluate, and reflect when solving mathematical problems perform better in problem-solving (Stillman and Mevarech, 2010). In mathematics learning, self-regulation is highly correlated with academic achievement, particularly in solving complex, non-routine, and unfamiliar mathematical problems (Mevarech et al., 2018). Gender and grade differences have been observed in SDL. Females have a lower self-efficacy in solving mathematical problems than males and often attribute low performance to uncontrollable factors (low ability and difficult tasks) (Vermeer et al., 2000). SDL abilities may increase as grade level increases (Chen et al., 2021), but some dimensions (such as motivation and self-confidence) may decrease as grade level increases (Mok et al., 2007).

In sum, the importance of SDL in mathematics education is evident, as these skills have been identified as crucial factors in improving student performance. It is noted that there need to be more well-developed learning scales to assess and measure self-directed learning skills in mathematics education effectively. While existing self-report scales, such as SDLRS and MSLQ, have been widely used and validated, they primarily assess general SDL skills. This research gap presents an opportunity for researchers to create targeted and reliable instruments that will enhance our understanding of the role of self-directed learning in mathematics and support educational strategies to foster these skills in students.

3. Research methods

3.1. Study participants

This study was divided into a pretest sample and a final sample. The pretest sample was used for item analysis and exploratory factor analysis (EFA). The final sample was used for confirmatory factor analysis (CFA), cross-validation, and measurement invariance tests. Crocker and Algina (2006) suggested that the number of participants should be at least five times the number of items in item analysis. If the number of participants is 10 times the number of items, the results will be more stable. Therefore, this study invited students from a private junior high school that accepts applications from all students in Taiwan to complete the online survey and used the data as a pretest sample. As a NAER partner school, it has been implementing an SDL curriculum since the development of the general guidelines. Additionally, it is the target of NAER's survey on the effectiveness of the implementation of the new curriculum guideline and it is representative. Students from this school could answer the online survey from September to mid-October 2021. There were 681 valid responses, and the return rate was 99.5%.

The final sample was obtained from six other NAER partner schools, which are also the targets of NAER's survey on the effectiveness of the implementation of the new curriculum guideline.

These schools have different sizes and are in the northern, central, southern, and eastern regions of Taiwan; hence, they are regionally representative. The researcher invited all students to complete the online survey, but some schools only permitted seventh and eighth graders to participate in the survey because the ninth graders were preparing for the high school entrance exams. A total of 5,575 students completed the online survey between the end of October and December 2021, and 5,456 valid responses were collected. The return rate was 97.8%.

3.2. Scale development and validation

The scale was developed in two stages: scale development and scale validation. The first stage involved constructing the items of the scale based on the curriculum guideline document and literature review of SDL. The second stage was divided into three sub-stages: content validity testing, pretest analysis, and formal test analysis, as shown in Figure 1.

3.2.1. Scale development

Because SDL is both domain-specific and cross-disciplinary, this study first defined the SDL in junior high school mathematics based on Schoenfeld's (1985) and Zimmerman's (2000) theories and the 12-year national mathematics curriculum guidelines. Subsequently, this study used the Delphi technique to examine and modify the definition. Delbecq et al. (1975) suggested that the number of experts in the Delphi technique should be 15 to 30 when the homogeneity of experts is high and can be reduced to five to ten when the homogeneity is low. However, if the number of experts in the Delphi technique is more than ten, the error of the expert group is minimized, and the group is most credible (Dalkey, 1969). Therefore, this study invited 15 experts, including the general guideline development committee, the mathematics curriculum guideline development committee, mathematics education scholars, mathematicians, scholars who conduct tests and assessments, and expert mathematics teachers to participate in the expert group of the Delphi technique. These experts provided advice from various professions and perspectives. After two Delphi techniques, all experts agreed that the definition of SDL in mathematics at the junior high school level is, "A junior high school student who has the SDL ability in mathematics is motivated to learn, can assess his/her learning needs, set learning goals, use strategies and resources to help achieve learning goals, identify and adjust his/her learning status, solve mathematical problems effectively, and reflect on the effectiveness of mathematical problem-solving strategies." Students who have completed three-year junior high school mathematics learning of the 12-year basic education should possess SDL competencies.

After establishing the definition of SDL in mathematics, the researcher used the definition to develop the scale from four dimensions: learning motivation, learning strategies, mathematics self-regulation, and learning resource management. According to experts' recommendations, the researcher adopted a four-point Likert scale that included strongly agree, agree, disagree, and strongly disagree. The researcher referred to SDL-related literature and scales (such as SDLRS and MSLQ; Mevarech and Kramarski, 2014), selected items from international surveys such as TIMSS and PISA, and modified the items. Additionally, the researcher compiled some items based on the definition to expand the item database of the scale.

3.2.2. Scale validation

In this stage, this study examined the content validity. Referring to Cheng et al. (2010), the researcher invited the 15 experts to evaluate each item individually and independently using a four-point Likert scale. The scoring criteria were as follows: (i) Appropriateness: this item reflects the definition of SDL in mathematics and is appropriate for assessing junior high school students and (ii) Clarity: this item is clear and easy to understand for junior high school students. After evaluating each item, the experts provided a score of 4 for the item being highly appropriate, clear, and precise, 3 for the item being appropriate and requiring minor revisions, 2 for the item being less appropriate and requiring major revisions, and 1 for the item being inappropriate and should be deleted. During the examination and

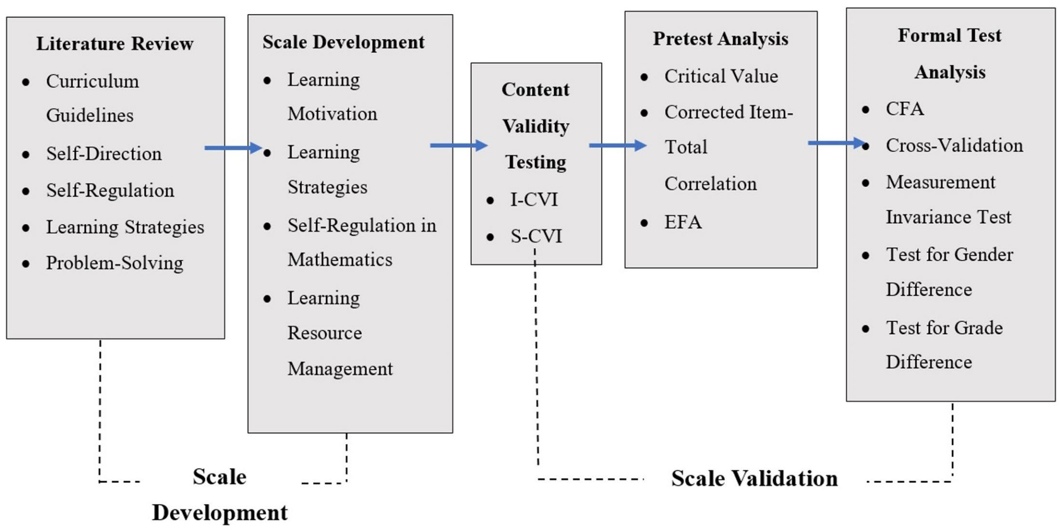


FIGURE 1
Framework for the development and validation of the mathemetics SDL scale.

grading process, the experts provided advice for the revision of the items if necessary. After the experts rated the items, the researcher calculated the content validity index (CVI) based on their ratings. The CVI value of each item (I-CVI) is the number of experts scoring three or more divided by the total number of experts. The CVI value of the scale (S-CVI) is the average of the I-CVI values within the scale (Polit and Beck, 2006). After the scale and the items passed the content validity test, a first draft of 68 items was completed.

This study used 68 items in the pretest and conducted item analysis and EFA. The researcher removed items with low-quality scores from the first draft based on the critical values and corrected item-total correlation (Spector, 1992) to ensure the quality of the items on the scale. Subsequently, this study conducted EFA, extracted factors using the principal axis factoring (PAF) method, and determined the number of factors based on the Kaiser eigenvalues greater than one and Cattell's scree plot. Because the factors were correlated with each other, in terms of the factor rotation methods, this study used the Promax method of oblique rotation to perform the rotation. The selection criterion was that the factor loading of the pattern matrix was greater than 0.40 (Wu, 2012), and 50 items were selected.

Subsequently, the researcher conducted the formal test, CFA, cross-validation, and measurement invariance tests across gender and grade. This study used the structural equation models to construct measurement models for the observed variables based on the EFA results and analyzed the results using AMOS 22.0 software and the maximum likelihood (ML) estimation method to confirm the composite reliability and convergent validity of the data. However, the aforementioned model may be rejected in CFA owing to the large sample size and model complexity (Bentler and Bonett, 1980; Marsh and Hocevar, 1985; Marsh et al., 1988). The scale consisted of 50 items in the formal test, and Bentler and Chou (1987) suggested that when the population distribution is unknown, the number of participants for the CFA should be larger than 10 times the number of items. Therefore, we randomly selected 1,014 students from the 5,456 samples from six schools based on the proportion of students in each school and randomly divided them into two groups. The first group was selected as the calibration sample ($N_1 = 507$) for CFA. AMOS 22.0 was used to examine the degree of consistency between the factor structure and the theory. For the CFA, all factors were allowed to covary (The Pearson correlation coefficient matrix for all study items is available at https://doi.org/10.5281/zenodo.7094563). After the model was fitted well, the second group was used as the validation sample ($N_2 = 507$) for cross-validation. To examine gender and grade differences, this study tested the measurement invariance of the scale across gender and grade using cohort analysis.

3.3. Analysis of gender and grade differences

This study examined gender and grade differences after measurement invariance was established. This study used the independent sample *t*-test to examine the latent mean difference across gender and used the analysis of variance to examine the latent mean difference across grades. If significant differences were observed between students in different grades, further comparisons were made. The data were analyzed using the software SPSS 22.0.

4. Research findings

4.1. Content validity test

This scale consisted of four subscales: learning motivation, learning strategies, mathematical self-regulation, and learning resource management. After the first round of expert evaluation, the I-CVI values of the items on the scale were between 0.6 and 1, and the S-CVI was 0.9. Polit and Beck (2006) suggested that an I-CVI greater than 0.78 and an S-CVI greater than 0.9 are acceptable when the number of experts is large. Although the experts approved the content validity of the first draft of the scale, some items required to be deleted or revised. The researcher deleted the items with an I-CVI below 0.78 and revised some items according to experts' suggestions. After the second round of expert review, the study calculated the CVI values. The S-CVI was 0.97, the CVI values for each subscale were between 0.96 and 0.99, and the CVI values for each item were between 0.87 and 1.00. The items that passed the content validity test were included in the first draft of the scale to conduct the pretest.

4.2. Item analysis

Item analysis refers to the process of selecting appropriate items based on critical values and corrected item-total correlation (DeVellis, 1998). The critical value method was used to divide the pretest sample into high-score and low-score groups (27% each). Subsequently, the two groups were used as independent variables, and the scores of individual items were used as dependent variables for an independent sample t-test. The scores of discriminating items should be significantly different in the two groups. In this study, the significance level was set at $\alpha = 0.05$. The results showed that the scores of all items were statistically significant ($p < 0.05$), meaning that all items on the scale demonstrated good discriminatory power. The corrected item-total correlation was calculated using the Pearson product–moment correlation coefficient between each item and the total score of the subscale (excluding the score of the item). The criterion for selecting items for this study was that the corrected item-total correlation coefficient must be above 0.3. The results showed that the correlation coefficients of all items were above 0.3, which meant that they were homogeneous. In other words, all items satisfied the criterion in terms of discriminatory power and corrected item-total correlation coefficient. Therefore, all items were retained for the EFA.

4.3. Exploratory factor analysis (EFA)

To confirm whether the data were suitable for factor analysis, this study used the Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) to examine the correlation coefficients between the variables and used Bartlett's spherical test values to examine whether the correlation coefficients in the correlation coefficient matrix were significantly higher than zero. The results showed that the KMO values of each subscale of the junior high school mathematics SDL scale were above 0.60. The Bartlett's spherical test values were 15433.47, 5593.17, 19157.58, 606.57, and 5379.56, which were all significant ($p < 0.001$) (Table 1). This indicated that the sample and subscales were suitable for the factor analysis.

TABLE 1 KMO and Bartlett's spherical test values for each subscale of the mathematics SDL scale.

		Learning motivation	Cognitive strategies	Mathematical self-regulation	Learning resource management
KMO measure of sampling adequacy		0.959	0.915	0.977	0.927
Bartlett's spherical test	χ^2	15433.47***	50.38.42***	15554.24***	5379.56***
	df	276	36	210	45
	p	<0.001	<0.001	<0.001	<0.001

***$p < 0.001$.

TABLE 2 Summary of factor analysis for the learning motivation pretest Subscale.

Item no.	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Q22	0.99	−0.05	0.01	−0.09	0.05
Q24	0.90	0.03	0.07	−0.06	−0.11
Q23	0.86	−0.04	−0.04	0.09	−0.02
Q21	0.85	−0.04	−0.02	0.06	0.07
Q17	0.05	0.64	0.01	−0.03	0.10
Q18	0.09	0.54	0.11	0.05	0.07
Q14	0.12	0.50	0.15	−0.09	0.20
Q7	−0.07	0.03	0.92	−0.01	−0.04
Q8	−0.02	−0.05	0.92	−0.08	0.09
Q9	0.05	−0.05	0.75	0.14	0.05
Q10	0.11	0.17	0.57	0.11	−0.17
Q2	0.08	0.04	−0.01	0.83	−0.09
Q1	0.24	−0.08	−0.03	0.79	0.01
Q4	0.27	−0.07	−0.06	0.73	0.05
Q6	0.04	0.10	0.14	0.47	0.04
Q13	−0.04	0.07	−0.06	−0.03	0.80
Q12	0.05	0.06	0.04	0.02	0.65
Q11	0.03	−0.04	0.35	−0.02	0.48
Factor name	Self-efficacy	Identified motivation	Extrinsic motivation (Work values)	Intrinsic motivation	Extrinsic motivation (Achievement-oriented)

Chen and Chang (2007) indicated that because the factors are correlated with each other, it is not possible to list the commonality of the items and the percentage explained by each factor. Therefore, only the factor loadings are presented in this paper.

4.3.1. Learning motivation pretest subscale

As shown in Table 2, the learning motivation pretest subscale included five factors. For Factor 1, four items satisfied the standard with factor loadings between 0.85 and 0.99, and the factor was called self-efficacy based on the theme of the items. For Factor 2, three items satisfied the standard with factor loadings between 0.50 and 0.64, and the factor was called the identified motivation based on the theme of items. For Factor 3, four items satisfied the standard with factor loadings between 0.57 and 0.92, and the factor was called extrinsic motivation (work values) based on the theme of items. For Factor 4, four items satisfied the standard with factor loadings between 0.47 and 0.83 and the factor is named intrinsic motivation based on the theme of items. For Factor 5, three items satisfied the standard with factor loadings between 0.48 and 0.80, and the factor was called extrinsic motivation (achievement-oriented) based on the theme of items.

4.3.2. Learning strategy pretest subscale

As shown in Table 3, the learning strategy pretest subscale consisted of three factors. For Factor 1, four items satisfied the standard with factor loadings between 0.53 and 0.73, and the factor was called refined strategies based on the theme of items. For Factor 2, three items satisfied the standard with factor loadings between 0.56 and 0.95, and the factor was called rehearsal strategies based on the theme of items. For Factor 3, two items satisfied the standard with factor loadings between 0.61 and 0.87, and the factor was called critical thinking strategies based on the theme of items.

4.3.3. Mathematics self-regulation strategy pretest subscale

As shown in Table 4, the mathematics self-regulation strategy pretest subscale consisted of two factors. For Factor 1, ten items

TABLE 3 Summary of factor analysis for the learning strategy pretest subscale.

Item no.	Factor 1	Factor 2	Factor 3
Q28	0.73	0.07	0.04
Q29	0.68	0.24	−0.03
Q31	0.53	0.04	0.22
Q25	−0.03	0.95	−0.04
Q26	0.33	0.59	−0.05
Q27	−0.10	0.56	0.22
Q32	0.00	0.06	0.87
Q33	0.29	−0.01	0.61
Factor name	Refined strategies	Rehearsal strategies	Critical thinking strategies

TABLE 5 Summary of factor analysis for the learning resource management pretest subscale.

Item	Factor 1	Factor 2
Q57	0.95	−0.17
Q56	0.83	0.05
Q55	0.79	0.03
Q59	0.48	0.23
Q63	0.16	0.62
Q62	0.30	0.52
Q60	0.32	0.47
Factor name	Time management and study environment	Interpersonal interaction and help-seeking

TABLE 4 Summary of factor analysis for the mathematics self-regulation strategy pretest subscale.

Item no.	Factor 1	Factor 2
Q49	0.98	−0.17
Q50	0.87	−0.01
Q47	0.85	−0.02
Q48	0.80	0.03
Q51	0.79	0.03
Q54	0.74	0.06
Q53	0.74	0.11
Q45	0.74	0.11
Q52	0.55	0.27
Q46	0.44	0.37
Q35	0.00	0.77
Q38	0.05	0.71
Q40	0.16	0.70
Q42	0.19	0.69
Q41	0.18	0.67
Q43	0.32	0.54
Q37	0.25	0.54
Factor name	Self-regulation in problem-solving	SRL in mathematics

satisfied the standard with factor loadings between 0.44 and 0.98, and the factor was called self-regulation in problem-solving based on the theme of the items. For Factor 2, seven items satisfied the standard with factor loadings between 0.54 and 0.77, and the factor was called SRL in mathematics based on the theme of items.

4.3.4. Learning resource management pretest subscale

As shown in Table 5, the learning resource management pretest subscale consisted of two factors. For Factor 1, four items satisfied the standard with factor loadings between 0.48 and 0.95, and the factor was called time management and study environment based on the theme of items. For Factor 2, three items satisfied the standard with factor loadings between 0.47 and 0.62, and the factor was called interpersonal interaction and help-seeking based on the theme of items.

In brief, this study established a mathematics SDL scale encompassing four subscales and 50 items. EFA revealed that the scale consisted of 12 factors. Table 6 demonstrates the dimensions, factors, and number of the items in the scale.

4.4. Confirmatory factor analysis

Before initiating the confirmatory factor analysis, the data was scrutinized to determine whether the data exhibit multivariate normality. Bollen (1989) contends that for data to be deemed multivariate normal, the multivariate kurtosis (i.e., Mardia's coefficient), must not exceed $p \times (p + 2)$, where p represents the number of observed items, and the critical ratio should be confined to a range of 5 or less (Kline, 2011). In this study, Mardia's coefficient is 734.113, considerably below the threshold of 2,600 ($50 \times (50 + 2)$); however, the critical ratio surpasses 5 with a value of 114.613. Considering the inherent robustness of maximum likelihood estimation in the face of minor deviations from multivariate normality, the technique remains reliable and valid, even when the data contradict statistical suppositions. Consequently, utilizing maximum likelihood estimation within this study retains its trustworthiness.

4.4.1. Competition model

Because each of the 12 latent variables corresponds to an item, the model can be considered a first-order measurement model with 12 latent variables. However, the correlation between the latent variables should be analyzed, and whether the 12 latent variables can be explained by a higher-order variable should be confirmed. To address the above problem, Noar (2003) suggested adopting a competition model, indicated that the best model can be selected through the comparison of the competition models (null, single factor, multifactor orthogonal, multifactor oblique, and second-order factor models). Because EFA was used to identify the 12 latent variables, this study only compared the multifactor orthogonal, multifactor oblique, and second-order factor models to select the best model for validating the mathematics SDL scale.

TABLE 6 Overview of the mathematics SDL scale: Dimensions, factors, contents, and number of items.

Dimension	Factors	Contents	Items no.
Learning motivation	Self-efficacy	Individual's belief in their own ability to successfully perform a task or achieve a goal.	18
	Identified motivation	Individuals consciously acknowledge the value and relevance of a particular activity to their personal goals or well-being, even if the activity itself is not inherently enjoyable.	
	Intrinsic motivation	Individuals engage in an activity or behavior because it is inherently satisfying, enjoyable, or interesting to the individuals.	
	Extrinsic motivation(Work values)	Individuals perform a task or job because of external rewards or outcomes, such as pay, recognition, or promotions.	
	Motivation (Achievement-oriented)	Individuals are driven by an internal desire to outperform or succeed rather than external rewards.	
Learning strategy	Refined strategies	The process of improving or adapting existing methods, approaches, or plans to better achieve a desired goal or outcome.	8
	Rehearsal strategies	The techniques or methods that individuals use to improve their memory and recall of information.	
	Critical thinking strategies	A set of cognitive skills applied to evaluate and analyze information, arguments, and evidence in order to make informed decisions or form well-reasoned opinions.	
Self-regulation strategy	Self-regulation in problem-solving	Individuals' ability to manage their thoughts and behaviors in order to effectively solve a problem or overcome a challenge.	17
	SRL in mathematics	Individuals' ability to manage their own learning process in mathematics, including setting goals, monitoring progress, and adapting strategies to improve understanding and learning performance.	
Learning resource management	Time management and study environment	The process of organizing, planning, and allocating one's time and study environment effectively	7
	Interpersonal interaction and help-seeking	Engaging with instructors and classmates to collaboratively learn and soliciting assistance or guidance to improve understanding of mathematical concepts.	

This study compared the models using the calibration sample ($N = 507$). The results are shown in Table 7. The multifactor orthogonal model had the worst fit. Indices such as χ^2, χ^2/df, root mean square error of approximation (RMSEA), comparative fit index (CFI), NNFI (non-normed fit index), SRMR (standardized root mean square), and GFI (goodness of fit index) did not satisfy the criteria, while ECVI (Expected Cross-Validation Index), AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion) were relatively large (17.98, 9098.67, and 9521.52). In terms of the multifactor oblique model, χ^2, χ^2/df, RMSEA, CFI, NNFI, SRMR, and GFI did not satisfy the criteria, and ECVI, AIC, and BIC were still relatively large despite the reduction (14.89, 7534.03, and 7985.25).

In terms of the second-order factor model, the results showed that although χ^2 was significant, χ^2/df was only 3.46, which reached the loose standard (the general standard requires less than three and the loose standard requires less than five). Furthermore, the RMSEA, CFI, NNFI, and SRMR reached an acceptable level. The GFI was not ideal but understandable because there were 50 observed variables. Moreover, the ECVI, AIC, and BIC were the smallest (8.40, 4252.17, 4725.76) among the three models, indicating that the second-order factor model had the best fit. In summary, the second-order factor model is a more appropriate validation model for the mathematics SDL scale.

4.4.2. Second-order factor model

The model fit indices of the second-order CFA model with 12 latent variables are listed in Table 8. As mentioned in the previous section, the important fit indices such as RMSEA = 0.070 and χ^2/df = 3.46 were good or acceptable, and the incremental fit indices such as CFI, NFI, and NNFI were acceptable, and the parsimonious fit indices such as PGFI and PNFI were good. Huang (2007) suggested the use of the majority rule to evaluate a model. Because this model

TABLE 7 Evaluation of the fitness of the competition models (calibration sample N=507).

Index model	$\chi^2(df)$	χ^2/df	RMSEA	CFI	NNFI	SRMR	GFI	ECVI	AIC	BIC
Multifactor orthogonal model	8898.67 (1175)	7.57	0.114	0.63	0.61	0.354	0.45	17.98	9098.67	9521.52
Multifactor oblique model	7322.03 (1169)	6.26	0.102	0.70	0.69	0.102	0.45	14.89	7534.03	7985.25
Second-order factor model	4028.17 (1163)	3.46	0.070	0.86	0.85	0.087	0.69	8.40	4252.17	4725.76

has many variables and achieving good results in all indices is difficult, the fitness of the proposed second-order CFA model of mathematics SDL scale is acceptable.

Table 9 shows that the reliability (R^2) of individual observed variables was between 0.33 and 0.85, which satisfied the requirement that the reliability of individual observed variables should be greater than 0.20 (Jöreskog and Sörbom, 1989). This result indicated that the observed variables had good reliability. The composite reliability of the latent variables was between 0.73 and 0.96, which satisfied the requirement that the reliability should be above 0.60 (Fornell and Larcker, 1981). This indicated that the composite reliability is acceptable.

Furthermore, Table 9 shows that the standardized factor loadings of all observed variables and the corresponding latent variables were between 0.57 and 0.92. The standardized loadings of all observed variables were above the threshold of 0.45 (Jöreskog and Sörbom, 1989). This indicated that the observed variables were sufficient to reflect the constructed latent variables. The average variance extracted (AVE) of the latent variables is between 0.47 and 0.78, indicating that the contribution of the observed variables to the latent variables was not inferior to the contribution of the error (Fornell and Larcker, 1981).

TABLE 8 Overall fitness of the modified model of the mathematics SDL scale (measured sample N=507).

Overall fit index	Evaluation standard	Model index	Evaluation result
Absolute fit index			
Likelihood-ratio χ^2	$p \geqq 0.05$	4028.17***	
GFI	≧ 0.90 or 0.80	0.69	Poor
AGFI	≧ 0.90 or 0.80	0.66	Poor
SRMR	≦ 0.08	0.087	Fair
RMSEA	≦ 0.08	0.070	Good
Incremental fit index			
NFI	≧ 0.90	0.82	Fair
NNFI	≧ 0.90	0.85	Fair
RFI	≧ 0.90	0.81	Fair
IFI	≧ 0.90	0.86	Fair
CFI	≧ 0.90	0.86	Fair
Parsimonious fit index			
PGFI	≧ 0.50	0.63	Good
PNFI	≧ 0.50	0.77	Good
PCFI	≧ 0.50	0.82	Good
Likelihood-ratio χ^2/df	≦ 3 or ≦ 5(loose standard)	3.46	Acceptable

The evaluation standard of model fit indices is based on Huang (2007) and Doll et al. (1994). ***$p < 0.001$.

4.5. Cross-validation

In this stage, this study examined the stability of the model, that is, to measure the stability of the calibration sample and validation sample through cross-validation. First, this study used a tight strategy, directly applied the model of the calibration sample ($N_1 = 507$) to the validation sample ($N_2 = 507$) and examined whether the factor loadings and covariance between latent variables were the same. Second, this study examined whether the MMF chi-square values (minimum fit function χ^2) of the loose strategy (factor loading and covariance between latent variables were freely estimated) and the tight strategy were significant.

The results (as shown in Table 10) indicated that for the tight strategy, the χ^2 of the validation sample ($N_2 = 507$) was 4603.49, and its ratio was 53.47%. This indicated that the contribution of the validation sample ($N_2 = 507$) was slightly higher than that of the calibration sample ($N_1 = 507$). The test results showed that the model can be applied to different samples in the same population, which means that the proposed model is validated.

In addition, because of the nested relation between loose and tight strategies, a chi-square difference test was performed. Table 10 shows that under the loose strategy, the MFFχ^2 of the validation sample was 4545.63, which was 57.86 less than the MFFχ^2 of the validation sample under the tight strategy. However, it was not significant when the degree of freedom was 50, indicating that the cross-validation was supported.

4.6. Measurement invariance tests

This study used the chi-square difference test, ΔRMSEA, ΔSRMR, and ΔCFI to test the measurement invariance of the measurement model across gender and grade. If $\Delta\chi 2$ is not significant, it means that the measurement model is invariant. However, the chi-square difference is likely to become significant as the sample size increases.

TABLE 9 Factor loading, reliability, and average variance explained of the mathematics SDL scale (calibration sample N=507).

Latent variable	Observed variable	Standardized factor loading	t-value	Reliability of individual item(R^2)	CR	AVE (%)
SDL in mathematics (Second-order latent variables)	Intrinsic motivation	0.78	18.75***	0.61	0.96	0.67
	Extrinsic motivation 1	0.76	16.81***	0.58		
	Extrinsic motivation 2	0.80	16.27***	0.64		
	Identified motivation	0.90	17.35***	0.82		
	Self-efficacy	0.55	12.29***	0.30		
	Time management and study environment	0.77	15.65***	0.59		
	Interpersonal interaction and help-seeking	0.93	16.54***	0.87		
	SRL in mathematics	0.90	15.10***	0.81		
	Self-regulation in problem-solving	0.87	20.69***	0.76		
	Rehearsal strategies	0.55	11.40***	0.30		
	Refined strategies	0.97	15.81***	0.94		
	Critical thinking strategies	0.89	20.80***	0.79		
Intrinsic motivation	A1	0.92	–	0.85	0.86	0.61
	A2	0.71	19.32***	0.50		
	A4	0.87	27.48***	0.76		
	A6	0.59	14.90***	0.35		
Extrinsic motivation (Work values)	A7	0.85	–	0.73	0.90	0.69
	A8	0.88	24.97***	0.77		
	A9	0.87	24.48***	0.75		
	A10	0.71	18.10***	0.50		
Extrinsic motivation (Achievement-oriented)	A11	0.80	–	0.64	0.83	0.63
	A12	0.82	18.86***	0.67		
	A13	0.76	17.35***	0.57		
Identified motivation	A14	0.75	–	0.57	0.81	0.58
	A17	0.75	16.61***	0.57		
	A18	0.79	17.42***	0.62		
Self-efficacy	A21	0.91	–	0.82	0.93	0.78
	A22	0.90	31.53***	0.82		
	A23	0.88	29.65***	0.77		
	A24	0.85	27.20***	0.72		
Rehearsal strategies	A25	0.87	–	0.75	0.83	0.62
	A26	0.82	18.13***	0.67		
	A27	0.65	14.67***	0.42		
Refined strategies	A28	0.66	–	0.43	0.76	0.51
	A29	0.74	14.67***	0.55		
	A31	0.74	14.53***	0.54		
Critical thinking strategies	A32	0.88	–	0.77	0.85	0.74
	A33	0.85	23.45***	0.72		

(Continued)

TABLE 9 (Continued)

Latent variable	Observed variable	Standardized factor loading	*t*-value	Reliability of individual item(R^2)	CR	AVE (%)
SRL in mathematics	A35	0.66	–	0.43	0.90	0.57
	A 37	0.69	13.78***	0.47		
	A 38	0.61	12.45***	0.37		
	A 40	0.83	16.21***	0.69		
	A 41	0.80	15.63***	0.63		
	A 42	0.84	16.35***	0.71		
	A 43	0.82	16.01***	0.67		
Self-regulation in problem-solving	A 45	0.87	–	0.76	0.96	0.70
	A 46	0.76	21.63***	0.58		
	A 47	0.86	26.94***	0.74		
	A 48	0.87	27.73***	0.76		
	A 49	0.86	26.64***	0.73		
	A 50	0.88	28.39***	0.78		
	A 51	0.79	23.10***	0.63		
	A 52	0.77	21.89***	0.59		
	A 53	0.84	25.82***	0.71		
	A 54	0.85	26.17***	0.72		
Time management and study environment	A55	0.80	–	0.64	0.82	0.53
	A56	0.88	20.51***	0.77		
	A57	0.63	14.19***	0.39		
	A59	0.57	12.89***	0.33		
Interpersonal interaction and help-seeking	A60	0.71	–	0.50	0.73	0.47
	A62	0.63	13.05***	0.39		
	A63	0.72	14.91***	0.52		

CR is composite reliability; AVE is average variance extracted.
***$p < 0.001$.

TABLE 10 Summary of cross validation of the model for the mathematics SDL scale (validation sample N_2=507).

Compared strategy	Model fitness MMFχ^2 (*df*)	Validation sample MMFχ^2	ΔMFFχ^2	χ^2 Ratio
Loose strategy	8573.79 (2326)	4545.63	57.86 (50)	53.02%
Tight strategy	8609.51 (2376)	4603.49		53.47%

The chi-square difference was less than 67.50 (df = 50); thus, it was not significant at the 5% level.

Therefore, according to scholars' recommendations, this study selected the model fit indices such as ΔRMSEA and ΔCFI as the evaluation criteria. When ΔRMSEA <0.01, ΔCFI < 0.01, and ΔSRMR <0.03, it means that factor loading is invariant (Chen, 2007; Wu et al., 2015). When ΔRMSEA <0.01, ΔCFI < 0.01, and ΔSRMR <0.01, it means intercept and residual are invariant (Chen, 2007; Wu et al., 2015).

Table 11 shows that the fitness of the baseline model was acceptable (gender: $\chi^2 = 9031.18$, RMSEA = 0.053, SRMR = 0.120, and CFI = 0.839; grade: $\chi^2 = 9018.61$, RMSEA = 0.053, SRMR = 0.091, and CFI = 0.839).

Second, Model 2 was constructed to determine whether the factor loadings of the two samples were constant. Table 12 shows that the chi-square difference ($\Delta\chi^2 = 87.67$) between Gender Model 2 and Model 1 was significant. However, ΔRMSEA = 0.000 was less than 0.01, ΔSRMR = −0.001 was less than 0.03, and ΔCFI = −0.002 was less than 0.01, indicating that the proposed scale had a constant factor loading across gender groups. In addition, the chi-square difference between Grade Model 2 and Model 1 ($\Delta\chi^2 = 22.78$) was not significant. ΔRMSEA = 0.000 was less than 0.01, ΔSRMR = 0.001 was less than 0.03, and ΔCFI = 0.000 was less than 0.01, indicating that the proposed scale had a constant factor loading across grade groups.

Third, based on the constant factor loading, Model 3 was constructed to test whether the structural weights of the two samples were constant and whether the structural weight was constant across gender and grade groups. The chi-square difference ($\Delta\chi 2 = 73.20$) between Gender Models 3 and 2 was significant, but ΔRMSEA = 0.000 was less than 0.01, ΔSRMR = 0.026 was less than 0.03, and

ΔCFI = −0.001 was less than 0.01, indicating that this scale had constant structural weight across gender groups. In addition, the chi-square difference ($\Delta\chi^2 = 10.36$) between Grade Models 3 and 2 was not significant, ΔRMSEA = 0.000 was less than 0.01, ΔSRMR = 0.001 was less than 0.01, and ΔCFI = 0.001 was less than 0.01, indicating that this scale had constant structural weight across grade groups. Fourth, based on the constant factor loading, Model 4 was constructed to test whether the residuals of the two samples were constant and whether the residuals were constant across gender and grade groups. The chi-square difference ($\Delta\chi 2 = 82.41$) between Gender Models 4 and 3 was significant, but ΔRMSEA = 0.000 was less than 0.01, ΔSRMR = −0.001 was less than 0.01, and ΔCFI = −0.005 was less than 0.01, indicating that this scale had residual invariance across gender groups. In addition, the chi-square difference ($\Delta\chi^2 = 137.98$) between Grade Models 4 and 3 was significant, but ΔRMSEA = −0.001 was less than 0.01, ΔSRMR = 0.000 was less than 0.01, and ΔCFI = - 0.002 was less than 0.01, indicating that this scale had residual invariance across grade groups. In conclusion, in terms of the measurement model, this scale has strong measurement invariance across gender and grade groups.

TABLE 11 Summary of model fit indices of mathematics SDL scale (*N*=1,014).

	χ^2	*df*	*p*-value	RMSEA	SRMR	CFI
Gender						
M1: Baseline model	9031.18	2,326	0.000	0.053	0.120	0.839
M2: Constant factor loading	9118.84	2,364	0.000	0.053	0.119	0.837
M3: Constant structural weight	9192.04	2,376	0.000	0.053	0.145	0.836
M4: Residual invariant	9451.04	2,438	0.000	0.053	0.144	0.831
Grade						
M1: Baseline model	9018.61	2,326	0.000	0.053	0.091	0.839
M2:Constant factor loading	9041.39	2,364	0.000	0.053	0.092	0.839
M3:Constant structural weight	9051.74	2,376	0.000	0.053	0.093	0.840
M4: Residual invariant	9189.72	2,438	0.000	0.052	0.093	0.838

4.7. Analysis of latent mean differences between different gender and grade groups

This study compared the difference between male and female students and the difference among students in grades seven, eight, and nine in their answers to the mathematics SDL scale. As shown in Table 13, the difference in total score between gender groups was significant in the *t*-test ($p < 0.001$), indicating that in terms of mathematics SDL abilities, male students ($M = 135.97$) have higher perception scores than female students ($M = 132.47$).

After examining the performance of male and female students on various factors of SDL in mathematics, we observed that male students had a significantly higher latent means than female students in terms of many factors of SDL in mathematics, including intrinsic motivation, extrinsic motivation (work values), self-efficacy, refined strategies, critical thinking strategies, and time management and study environment. The difference in self-efficacy was the largest (ΔM = 0.41), followed by intrinsic motivation (ΔM = 0.24). The latent mean of identified motivation of female students was slightly higher than that of male students.

In addition, the F-test showed that the mathematics SDL scores of students in different grades were significantly different, as shown in Table 14. The total score of the seventh grade ($M = 137.69$) was significantly higher than that of the eighth grade ($M = 133.33$) and ninth grade ($M = 131.77$). Although the score of the eighth grade was higher than that of the ninth grade, it was still not significant.

After examining the latent mean difference of various factors of SDL in mathematics among students of different grades, we observed that most of the factors of the seventh graders were significantly higher than those of the eighth and ninth graders, such as extrinsic motivation-work values (2.96, 2.78, 2.71), extrinsic motivation (achievement-oriented) (3.15, 3.02, 3.03), identified motivation (2.94,

TABLE 12 Summary of model invariance across gender and grade (*N*=1,014).

	$\Delta\chi^2$	Δ*df*	*p*-value	ΔRMSEA	ΔSRMR	ΔCFI
Gender model comparison						
M2 vs. M1	87.67	38	0.000	0.000	−0.001	−0.002
M3 vs. M2	73.20	12	0.000	0.000	0.026	−0.001
M4 vs. M3	82.41	12	0.000	0.000	−0.001	−0.005
Grade model comparison						
M2 vs. M1	22.78	38	0.976	0.000	0.001	0.000
M3 vs. M2	10.36	12	0.585	0.000	0.001	0.001
M4 vs. M3	137.98	62	0.000	−0.001	0.000	−0.002

TABLE 13 Summary of differences in various factors across genders.

Factor	Gender	Number of students	Mean	Standard deviation	*t*-value	*p*-value
Intrinsic motivation	Male	3,177	2.64	0.79	12.08***	<0.001
	Female	2,960	2.40	0.71		
Extrinsic motivation (Work values)	Male	3,177	2.87	0.76	5.67***	<0.001
	Female	2,960	2.76	0.71		
Extrinsic motivation (Achievement-oriented)	Male	3,177	3.06	0.75	−0.78	0.435
	Female	2,960	3.07	0.70		
Identified motivation	Male	3,177	2.83	0.77	−2.83**	0.005
	Female	2,960	2.88	0.69		
Self-efficacy	Male	3,177	2.42	0.88	19.30***	<0.001
	Female	2,960	2.01	0.78		
Rehearsal strategies	Male	3,177	2.29	0.77	−1.41	0.159
	Female	2,960	2.32	0.69		
Refined strategies	Male	3,177	2.61	0.79	4.23***	<0.001
	Female	2,960	2.53	0.70		
Critical thinking strategies	Male	3,177	2.90	0.80	4.57***	<0.001
	Female	2,960	2.81	0.73		
Self-regulation in problem-solving	Male	3,177	2.95	0.71	−0.01	0.992
	Female	2,960	2.95	0.62		
SRL in mathematics	Male	3,177	2.65	0.74	1.90	0.058
	Female	2,960	2.61	0.64		
Time management and study environment	Male	3,177	2.49	0.77	2.67**	0.008
	Female	2,960	2.44	0.65		
Interpersonal interaction and help-seeking	Male	3,177	2.72	0.77	−1.07	0.283
	Female	2,960	2.74	0.68		
Total score	Male	3,177	135.97	31.43	4.70***	<0.001
	Female	2,960	132.47	26.74		

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

2.83, 2.78), rehearsal strategies (2.42, 2.27, 2.22), critical thinking strategies (2.94, 2.82, 2.82), self-regulation in problem-solving (3.02, 2.92, 2.91), SRL in mathematics (2.72, 2.62, 2.55), and time management and study environment (2.54, 2.46, 2.39). Some factors of the eighth grader are higher than those of the ninth grade, such as extrinsic motivation (work values), SRL in mathematics, and time management and study environment. In summary, mathematics SDL ability does not improve with increasing grade level (see Table 14).

5. Conclusions and recommendations

This study established a mathematics SDL scale to investigate the SDL performance of Taiwan junior high school students 3 years after the implementation of the new curriculum guideline and to understand the differences across genders and grades. First, in terms of reliability analysis, the composite reliabilities of the 12 factors are between 0.73 and 0.96 and are all above 0.60, indicating that the scale has good composite reliability. In terms of validity analysis, after expert review, the items of the scale passed the I-CVI and S-CVI standards; therefore, the scale has content validity. This study conducted an EFA using a pretest sample of 681 students to establish the final items of the mathematics SDL scale. Subsequently, this study used a final sample of 5,456 junior high school students for CFA, cross-validation, and measurement invariance tests across genders and grades. The CFA indicated that the second-order model fit the observed data well. In other words, the 12 factors under the four dimensions, namely learning motivation, learning strategies, mathematical self-regulation, and learning resource management can be explained by the SDL in mathematics.

Second, the cross-validation indicated that the scale's model has strong stability. Third, the measurement invariance tests across genders and grades indicated that the measurement model of this scale had strong measurement invariance across gender and grade groups. The results of this study demonstrated the richness and complexity of SDL in mathematics. Follow-up studies can be conducted in other countries or regions that promote SDL curriculum reform to confirm the reliability and validity of this scale.

The researcher examined the latent mean differences in mathematics SDL abilities by gender and grade based on the

TABLE 14 Summary of differences in various factors across grades.

Factor	Grade	Number of students	Mean	Standard deviation	*F*-score	*p*-value	Scheffe post-hoc comparison
Intrinsic motivation	(1) Seventh grade	2,041	2.55	0.75	1.43	0.238	
	(2) Eighth grade	2,146	2.51	0.77			
	(3) Ninth grade	1,950	2.52	0.76			
Extrinsic motivation (Work values)	(1) Seventh grade	2,041	2.96	0.72	64.54***	<0.001	1 > 2 > 3
	(2) Eighth grade	2,146	2.78	0.74			
	(3) Ninth grade	1,950	2.71	0.74			
Extrinsic motivation (Achievement-oriented)	(1) Seventh grade	2,041	3.15	0.68	21.22***	<0.001	1 > 2,3
	(2) Eighth grade	2,146	3.02	0.73			
	(3) Ninth grade	1,950	3.03	0.76			
Identified motivation	(1) Seventh grade	2,041	2.94	0.71	23.02***	<0.001	1 > 2,3
	(2) Eighth grade	2,146	2.83	0.73			
	(3) Ninth grade	1,950	2.78	0.74			
Self-efficacy	(1) Seventh grade	2,041	2.21	0.85	0.40	0.671	
	(2) Eighth grade	2,146	2.23	0.85			
	(3) Ninth grade	1,950	2.22	0.87			
Rehearsal strategies	(1) Seventh grade	2,041	2.42	0.74	41.26***	<0.001	1 > 2,3
	(2) Eighth grade	2,146	2.27	0.73			
	(3) Ninth grade	1,950	2.22	0.72			
Refined strategies	(1) Seventh grade	2,041	2.62	0.74	7.23**	0.001	1 > 3
	(2) Eighth grade	2,146	2.55	0.74			
	(3) Ninth grade	1,950	2.53	0.76			
Critical thinking strategies	(1) Seventh grade	2,041	2.94	0.75	15.35***	<0.001	1 > 2,3
	(2) Eighth grade	2,146	2.82	0.76			
	(3) Ninth grade	1,950	2.82	0.78			
Self-regulation in problem-solving	(1) Seventh grade	2,041	3.02	0.65	15.51***	<0.001	1 > 2,3
	(2) Eighth grade	2,146	2.92	0.67			
	(3) Ninth grade	1,950	2.91	0.68			
SRL in mathematics	(1) Seventh grade	2,041	2.72	0.69	28.10***	<0.001	1 > 2 > 3
	(2) Eighth grade	2,146	2.62	0.70			
	(3) Ninth grade	1,950	2.55	0.69			

(Continued)

10.3389/fpsyg.2023.1145442

10.3389/fpsyg.2023.1145442

TABLE 14 (Continued)

Factor	Grade	Number of students	Mean	Standard deviation	*F*-score	*p*-value	Scheffe post-hoc comparison
Time management and study environment	(1) Seventh grade	2,041	2.54	0.73	21.32***	<0.001	1 > 2 > 3
	(2) Eighth grade	2,146	2.46	0.70			
	(3) Ninth grade	1,950	2.39	0.71			
Interpersonal interaction and help-seeking	(1) Seventh grade	2,041	2.75	0.74	1.71	0.180	
	(2) Eighth grade	2,146	2.73	0.72			
	(3) Ninth grade	1,950	2.70	0.73			
Total score	(1) Seventh grade	2,041	137.69	28.41	22.24***	<0.001	1 > 2,3
	(2) Eighth grade	2,146	133.33	29.47			
	(3) Ninth grade	1,950	131.77	29.75			

*$p<0.05$; **$p<0.01$; ***$p<0.001$.

answers of 6,137 students in the pretest and formal test. The analysis results indicated that there is a gender difference in mathematics SDL. Male students are higher than female students in many factors, with the greatest difference being in intrinsic motivation and self-efficacy. This finding is similar to PISA's (OECD, 2014) study on Taiwanese students' motivation and beliefs in mathematics learning. Because the PISA survey was conducted before the implementation of the new curriculum guideline, the findings implied that gender differences in mathematics learning are not effectively addressed by the new curriculum guideline. It is worth noting that male and female students have no differences in innate mathematical abilities (Lindberg et al., 2010), they have similar mathematical attitudes (Ghasemi and Burley, 2019), and the difference is primarily owing to sociocultural factors (Kane and Mertz, 2012). A myth remains in Taiwan's culture that men are suitable for studying science and technology and women are suitable for studying humanities, and women are not encouraged to study STEM-related subjects and work in related industries. Although Taiwan ranks first in Asia in terms of the gender equity index (Executive Yuan, 2022) and the mathematics curriculum encourages the inclusion of gender equality, the findings of PISA and this study suggest that there is still scope for improvement in mathematics education in Taiwan in terms of gender equity. In addition, the SDL in mathematics does not increase with grade. This result is similar to that of Mok et al. (2007) and different from that of Chen et al. (2021). This shows that although the new wave of curriculum reform in Taiwan has contributed to improving the SDL of the next generation, it is not reflected in the subject learning.

These findings implied the importance of teacher education for developing students' SDL. Although much research has been conducted on SDL from the student perspective, more research on how teachers can guide students in SDL is needed (Dignath and Büttner, 2018). Subsequent studies could explore how mathematics teachers facilitate girls or higher graders to develop self-efficacy in mathematics and intrigue their intrinsic motivation. Take the item, I have always believed that mathematics is one of my best subjects, as an example. This assertion underscores the necessity to foster a positive self-perception in girls and elder students, empowering them to excel in mathematical disciplines. In addition, future studies could consider the dual role (Kramarski, 2018) that teachers play in SDL professional development and SDL enhancement. These studies may contribute to a more diverse and enriched field of study.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Committee of National Taiwan Normal University (NTNUREC-202005ES024). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

C-HL and C-HC contributed to the conceptualization and design of the study. C-HC collected the data and performed the statistical analysis. C-HL wrote the first draft of the manuscript. C-HC reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Taiwan's National Science and Technology Council under Grant numbers (MOST 109-2511-H-656 -001 -MY3 and MOST 111-2410-H-142-003). The funding agencies had no involvement in this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amey, B. E. (2008). *An exploration of the relationship between experiential learning and self-directed learning readiness*. (Doctoral dissertation). Boca Raton, FL: Florida Atlantic University.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037/0033-295X.84.2.191

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606. doi: 10.1037/0033-2909.88.3.588

Bentler, P. M., and Chou, C. P. (1987). Practical issues in structural modeling. *Sociol. Methodol. Res.* 16, 78–117. doi: 10.1177/0049124187016001004

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.

Brandt, W. C. (2020). *Measuring student success skills: a review of the literature on self-directed learning. 21st century success skills*. Dover, NH: National Center for the Improvement of Educational Assessment.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Chen, C. C., and Chang, C. H. (2007). *Quantitative research and statistical analysis*. Taipei, Taiwan: New Sharing.

Chen, C.-H., Chen, K.-Z., and Tsai, H.-F. (2021). Did self-directed learning curriculum guidelines change Taiwanese high-school students' self-directed learning readiness? *Asia-Pac. Educ. Res.* 31, 409–426. doi: 10.1007/s40299-021-00582-w

Cheng, S.-F., Kuo, C.-L., Lin, K.-C., and Hsieh, J.-L. (2010). Development and preliminary testing of a self-rating instrument to measure self-directed learning ability of nursing students. *Int. J. Nurs. Stud.* 47, 1152–1158. doi: 10.1016/j.ijnurstu.2010.02.002

Corno, L. (2001). "Volitional aspects of self-regulated learning" in *Self-regulated learning and academic achievement: Theoretical perspectives*. eds. B. J. Zimmerman and D. H. Schunk. *2nd* ed (Hillsdale, NJ: Erlbaum), 191–225.

Crocker, L., and Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.

Dalkey, N. C. (1969). *The Delphi method: An experimental study of group opinion*. Santa Monica, CA: Rand Corp.

de Bruin, A. B. H., and van Merriënboer, J. J. G. (2017). Bridging cognitive load and self-regulated learning research: a complementary approach to contemporary issues in educational research. *Learn. Instr.* 51, 1–9. doi: 10.1016/j.learninstruc.2017.06.001

Delbecq, A. L., Van de Ven, A. H., and Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Glenview, IL: Scott, Foresman.

DeVellis, R. F. (1998). *Scale development: Theory and applications*. CA: Sage.

Dignath, C., Buettner, G., and Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educ. Res. Rev.* 3, 101–129. doi: 10.1016/j.edurev.2008.02.003

Dignath, C., and Büttner, G. (2018). Teachers' direct and indirect promotion of self-regulated learning in primary and Secondary School Mathematics Classes – insights from video-based classroom observations and teacher interviews. *Metacognition Learn.* 13, 127–157. doi: 10.1007/s11409-018-9181-x

Doll, W. J., Xia, W., and Torkzadeh, G. (1994). A confirmatory factor analysis of the end-user computing satisfaction instrument. *MIS Q.* 12, 259–274. doi: 10.2307/249524

Fornell, C., and Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.2307/3151312

Gabrielle, D. M., Guglielmino, L. M., and Guglielmino, P. J. (2006). Developing self-directed learning readiness of future leaders in a military college through instructional innovation. *Int. J. Self-Directed Learn.* 3, 24–35.

Ghasemi, E., and Burley, H. (2019). Gender, affect, and math: a cross-national meta-analysis of trends in international mathematics and science study 2015 outcomes. *Large Scale Assess. Educ.* 7:10. doi: 10.1186/s40536-019-0078-1

Guglielmino, L. M. (1977). *Development of the self-directed learning readiness scale (doctoral dissertation)*. Athens, Georgia: University of Georgia. Available at: ProQuest Dissertations and Theses Global database UMI No. 7806004.

H.K. Curriculum Development Council (2017). *Secondary education curriculum guide draft booklet 3 effective learning and teaching: developing lifelong and self-directed learners*. Available at: http://www.edb.gov.hk/attachment/en/curriculum-development/renewal/Guides/SECG%20booklet%203_20170531.pdf

Huang, F. M. (2007). *Theory and application of structural equation models* (*5th*). Taipei, Taiwan: Wu-Nan Press.

Jöreskog, K. G., and Sörbom, D. (1989). *LISREL 7: a guide to the program and applications*. Chicago, IL: SPSS.

Kane, J. M., and Mertz, J. E. (2012). Debunking myths about gender and mathematics performance. *Not. AMS* 59, 10–21. doi: 10.1090/noti790

Kirschner, P. A., and Hendrick, C. (2020). *How learning happens: Seminal works in educational psychology and what they mean in practice*. New York, NY: Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (*3rd*). New York, NY: Guilford Press.

Knowles, M. (1975). *Self-directed learning: A guide for learners and teachers*. New York, NY: Cambridge Books.

Kramarski, B. (2018). "Teachers as agents in promoting students' SRL and performance" in *Handbook of self-regulation of learning and performance*. eds. D. H. Schunk and J. A. Greene (New York, NY: Routledge), 223–239.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychol. Bull.* 136, 1123–1135. doi: 10.1037/a0021276

Marsh, H. W., Balla, J. R., and McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. *Psychol. Bull.* 103, 391–410. doi: 10.1037/0033-2909.103.3.391

Marsh, H. W., and Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: first- and higher order factor models and their invariance across groups. *Psychol. Bull.* 97, 562–582. doi: 10.1037/0033-2909.97.3.562

Mevarech, Z. R., and Kramarski, B. (2014). *Critical maths for innovative societies: The role of metacognitive pedagogies on mathematical reasoning*. Paris, France: OECD.

Mevarech, Z. R., Verschaffel, L., and De Corte, E. (2018). "Metacognitive pedagogies in mathematics classrooms: from kindergarten to college and beyond" in *Handbook of self-regulation of learning and performance*. eds. D. H. Schunk and J. A. Greene (New York, NY: Routledge/Taylor and Francis Group), 109–123.

Ministry of Education (2018). *Mathematics curriculum guidelines of 12-year basic education for elementary and secondary schools and general senior high schools*. Taipei, Taiwan.

Mok, M. M. C., Cheng, Y. C., Leung, S. O., Shan, P. W. J., Moore, P., and Kennedy, K. (2007). "Self-directed learning as a key approach to effectiveness of education: a comparison among mainland China, Hong Kong, Macau, and Taiwan" in *International handbook of school effectiveness and improvement*. ed. T. Townsend (Dordrecht, Netherlands: Springer), 839–858.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: learners, contexts, and cultures*. Washington, DC: The National Academies Press.

National Academy for Educational Research. (2018). *Curriculum guidelines for 12-year basic education: General guidelines*. Taipei, Taiwan: Ministry of Education. Available at: https://cirn.moe.edu.tw/Upload/file/32077/83646.pdf

New Zealand Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, NZ: New Zealand.

Noar, S. M. (2003). The role of structural equation modeling in scale development. *Struct. Equ. Modeling* 10, 622–647. doi: 10.1207/S15328007SEM1004_8

OECD (2014). PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014), Paris: OECD Publishing. doi: 10.1787/9789264201118-en

OECD (2018). *PISA 2021Mathematics framework (draft)*. Available at: https://pisa2021-maths.oecd.org/files/PISA%202021%20Mathematics%20Framework%20Draft.pdf

Panadero, E. (2017). A review of self-regulated learning: six models and four directions for research. *Front. Psychol.* 8, 1–28. doi: 10.3389/fpsyg.2017.00422

Pintrich, P. R., Smith, D. A. F., Garcia, T., and Mckeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor, MI: University of Michigan.

Polit, D. F., and Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res. Nurs. Health* 29, 489–497. doi: 10.1002/nur.20147

Polya, G. (1957). *How to solve it?* (*2nd*). Princeton: Princeton University Press.

Saks, K., and Leijen, Ä. (2014). Distinguishing self-directed and self-regulated learning and measuring them in the e-learning context. *Procedia Soc. Behav. Sci.* 112, 190–198. doi: 10.1016/j.sbspro.2014.01.1155

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic Press.

Schoenfeld, A. H. (1992). "Learning to think mathematically: problem solving, metacognition, and sense making in mathematics" in *Handbook of research on mathematics teaching and learning: A project of the National Council of teachers of mathematics*. ed. D. A. Grouws (Reston, VA: National Council of Teachers of Mathematics), 334–370.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educ. Res.* 43, 404–412. doi: 10.3102/0013189X14554450

Schunk, D. H. (2012). *Learning theories: an educational perspective* (*6th*). Boston: Allyn & Bacon.

Schunk, D. H. (2020). *Learning theories: an educational perspective* (*8th*). Boston: Allyn & Bacon.

Schunk, D. H., and Mullen, C. A. (2013). "Motivation" in *International guide to student achievement*. eds. H. John and M. A. Eric (New York, NY: Routledge), 67–69.

Schunk, D. H., and Rice, J. M. (1987). Enhancing comprehension skill and self-efficacy with strategy value information. *J. Read. Behav.* 19, 285–302. doi: 10.1080/10862968709547605

Schunk, D. H., and Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *J. Read. Behav.* 23, 351–364. doi: 10.1080/10862969109547746

Senge, P., Cambron-McCabe, N., Lucas, T., Smith, B., Dutton, J., and Kleiner, A. (2000). *Schools that learn: a fifth discipline fieldbook for educators, parents, and everyone who cares about education*. New York, NY: Doubleday/Currency.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Thousand Oaks, CA: Sage.

Steffens, K. (2015). Competences, learning theories and MOOCs: recent developments in lifelong learning. *Eur. J. Educ.* 50, 41–59. doi: 10.1111/ejed.12102

Stillman, G., and Mevarech, Z. R. (2010). Metacognitive research in mathematics education: from hot topic to mature field. *ZDM Int. J. Math. Educ.* 42, 145–148. doi: 10.1007/s11858-010-0245-x

Vermeer, H. J., Boekaerts, M., and Seegers, G. (2000). Motivational and gender differences: sixth-grade students' mathematical problem-solving behavior. *J. Educ. Psychol.* 92, 308–315. doi: 10.1037/0022-0663.92.2.308

Wiliam, D. (2018). *Embedded formative assessment* (*2nd*). Bloomington, IN: Solution Tree Press.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learn. Individ. Differ.* 8, 327–353. doi: 10.1016/S1041-6080(96)90022-9

Wu, M. L. (2012). *SPSS operation and application for statistical analysis of questionnaire*. Taipei, Taiwan: Wu-Nan Press.

Wu, T. H., Chang, C. C., Chen, C. Y., Wang, J. D., and Lin, C.-Y. (2015). Further psychometric evaluation of the self-stigma scale-short: measurement invariance across mental illness and gender. *PLoS One* 10, 1–12. doi: 10.1371/journal.pone.0117592

Yuan, E.. (2022). *Gender at a glance in R.O.C.* Taipei, Taiwan.

Zimmerman, B. J. (2000). "Attaining self-regulation: a social cognitive perspective" in *Handbook of self-regulation*. eds. M. Boekaerts, P. R. Pintrich and M. Zeidner (San Diego: Academic Press), 13–39.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *Am. Educ. Res. J.* 45, 166–183. doi: 10.3102/0002831207312909

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 29 June 2023
DOI 10.3389/fpsyg.2023.1193675

OPEN ACCESS

EDITED BY
Shuhua An,
California State University,
Long Beach, United States

REVIEWED BY
Muhd Khaizer Omar,
Center for Academic Development Universiti Putra Malaysia, Malaysia
Watcharee Ketpichainarong,
Mahidol University, Thailand

*CORRESPONDENCE
Weilan Mo
✉ 675071915@qq.com

RECEIVED 25 March 2023
ACCEPTED 05 June 2023
PUBLISHED 29 June 2023

CITATION
Cheng X, Mo W and Duan Y (2023) Factors contributing to learning satisfaction with blended learning teaching mode among higher education students in China.
Front. Psychol. 14:1193675.
doi: 10.3389/fpsyg.2023.1193675

Factors contributing to learning satisfaction with blended learning teaching mode among higher education students in China

Xiaoliang Cheng[1], Weilan Mo[2]* and Yujing Duan[1]

[1]Education Evaluation and Supervision Division, Chengdu University of Technology, Chengdu, China,
[2]Student Office, Chengdu University of Traditional Chinese Medicine, Chengdu, China

Blended learning has increasingly grown in importance as a method of classroom instruction in Chinese higher education classrooms in the context of fast-evolving network information technology, higher standards of educational informatization, and growing attention to the reform of teaching modes in higher education. The efficiency of blended learning can be increased by better understanding the students' learning satisfaction and its key influencing factors. Based on the theories of constructivism and phenomenology, the study constructs an index system of student satisfaction with blended learning in higher education, and conducts a questionnaire survey on 650 students with blended learning experience in 6 universities in Sichuan Province, China, obtaining 598 valid questionnaires after reviewing the collected questionnaires for missing values. This study uses descriptive statistical analysis (DSA), one-way ANOVA, Pearson correlation analysis, and multiple linear regression (MLR) to analyze the effects of each factor in the index system on satisfaction. Results indicate that the overall level of student satisfaction with blended learning in universities is moderately high, with students' self-satisfaction being the lowest, and that substantial disparities exist in the evaluation of satisfaction with blended learning on various online resources, online teaching forms, and offline teaching methods. This study applies multiple linear regression (MLR) to conclude that students' learning attitudes, curriculum design, and teachers' teaching methods are the most important factors influencing satisfaction with blended learning in universities. Results indicate that a blended learning system be built from the three dimensions of students, teachers, and curriculum, offering a theoretical foundation and point of reference for the ongoing reform of blended learning in higher education. The study is of great significance in optimizing teaching quality and deepening the reform of blended learning.

KEYWORDS

blended teaching, student satisfaction, affecting elements, Chinese university, variance analysis

Introduction

With the rapid expansion of information, technology, and the Internet in the 21st century, the relationship between higher education and information technology has grown closer (Mitchell and Forer, 2010). The development and reform of higher education have gradually moved in the new direction of curriculum and IT integration (Gerbica, 2011; Xu and Shi, 2018). Blended learning is part of the ongoing convergence of two archetypal learning environments (Ates, 2009). The "Implementation Opinions of the Ministry of Education on the Construction of First-class Undergraduate Curriculum" issued by the Ministry of Education of the People's Republic of China (2019), clarified that online and offline blended courses should be customized to local, school, and course-specific circumstances, placing special emphasis on the reform and innovation of curriculum content and teaching methods. Blended learning has made a historic leap in recent years in China. Taking Chinese MOOCs (Massive Open Online Courses) as an example, universities in western China implemented blended learning for 1.26 million courses in 2022 alone, with 210 million students participating in it, significantly raising the teaching quality of teachers and the learning outcomes of students in this region (People's Daily, 2023). "The new integration of technology and education will unleash a new revolution, and blended learning will become the new normal of higher education in the future" suggested Wu Yan (Xinhuanet, 2021), director of the Department of Higher Education of the Ministry of Education. Due to its dynamic and modern nature, blended learning brings higher education a fresh lease on life.

Blended learning undoubtedly demonstrates a new route for future teaching development in the era of ever-changing network information technology (Gerbica, 2011), high standards of educational informatization (Picciano and Dziuban, 2007), and growing attention to teaching mode reform (Garrison and Kanuka, 2004). China's blended learning combines the benefits of traditional and online education to achieve the dual mode of "online + offline," and the introduction of this teaching mode will lead to significant advancements in higher education (Zhu and Hu, 2021). Blended learning is currently being popularized and developed in both domestic and international higher education (Garrison and Kanuka, 2004). Increasing attention is also being given to modeling the effectiveness of e-learning mechanisms, including factors such as the use of online assessments, students' attitudes toward e-learning mechanisms, and the application of different teaching methods (Carver et al., 2004; Chang and Tung, 2008; Sivo et al., 2010; Sadeghi et al., 2014). In-depth investigation and research on students' satisfaction with blended learning, as well as further evaluation of various influencing factors involved in blended learning, play a vital role in enhancing blended learning theory and practice.

This study starts from the research results on student satisfaction with blended learning in higher education, both domestically and internationally. Firstly, it comprises the basic concepts and related theories of blended learning, student satisfaction, constructivist learning theory, etc. Secondly, through two modes of online and offline questionnaires, a survey was conducted on 598 students of different majors and grades in six universities in China to examine the implementation of blended learning. Thirdly, descriptive statistical analysis (DSA), one-way ANOVA, Pearson's correlation analysis, and multiple linear regression (MLR) were used to analyze the influence of each factor in the index system on satisfaction. It analyzes the current situation and problems of university students' satisfaction with blended learning from students' perspectives and explores the factors that affect satisfaction with blended learning in higher education. The study further explores the focus and direction of blended learning improvement, which provides a reference for the reform and innovation of blended learning and is of great significance in optimizing teaching quality and deepening the reform of blended learning.

Research review

Research on blended learning

Domestic and international scholars have undergone a metamorphosis with the continuous development of blended learning, moving from basic conceptual research to technology application research, as well as evaluation and management research. Studies on the effects of blended learning and its influencing factors are becoming more in-depth.

Garrison and Vaughan (2008) indicated that blended learning was developed from the strengths of face-to-face and distance learning. According to Linder (2017), the blended learning method is a way of using technology to give students access to a variety of learning environments, where face-to-face activities are often combined with technology-mediated activities to allow students to have more active learning opportunities and more targeted instruction when learning outside of the classroom. Pape (2010) proposed that the advantages of blended learning include providing students with multiple ways to demonstrate their knowledge while engaging various learning styles and helping students build independent and self-directed learning skills, which are essential for lifelong learners. Tayebinik and Puteh (2013) pointed out that blended learning promotes a stronger sense of engagement and community than traditional face-to-face or fully online teaching blended learning is a flexible, scalable, and meaningful way of teaching and learning.

Blended learning was initially brought to China in 2003 by Zhu and Meng (2003), who suggested that it can be defined as instruction applying multiple delivery techniques in order to optimize learning outcomes and the cost of learning program delivery. Based on the level of teaching space, Chen et al. (2019) proposed that blended learning expands students' learning beyond the typical classroom's dozens of square meters and out into the online and offline worlds, and is a mixture of learning environments as well as learning spaces. Wei and Tan (2021) believes that blended learning is a teaching mode guided by behaviorist and constructivist learning theories and with the aid of modern educational technology, Internet technology, and other technical means. It deeply integrates traditional face-to-face instruction, practical instruction, and online instruction to achieve the highest teaching efficiency and effect. In general, blended learning involves the use of different teaching resources, "online + offline" multiple learning platforms as the carrier, more thorough learning objectives as the guide, flexible use of a variety of teaching methods and organizational forms, fully combining the benefits of online and offline classrooms.

Research on student satisfaction

The concept of student satisfaction, based on customer satisfaction, was formally introduced into the field of education at the end of the last century by local and international scholars. The higher education sector has become increasingly aware that students' expectations and needs are similar to those of a service business, and has paid further attention to fulfilling their expectations and needs (Cheng and Tam, 1997). The learning experience and satisfaction of students, who are the subjects of education, are important criteria for evaluating teaching quality (Guolla, and Michael., 1999).

Several academics have various definitions of the concept of learning satisfaction based on different study viewpoints. One is the idea of discrepancy based on a comparison of students' learning before and after. It is the psychological satisfaction or dissatisfaction of students' perception of the service level of teaching during the entire teaching service, compared with their self-expectations or demands before learning. The second is the learning experience and learning outcomes perceived by the students themselves (Zhu et al., 2020), and the concept of student satisfaction is proposed in terms of the students' own holistic nature, which is the student's subjective evaluation of various outcomes and experiences related to education (Oliver, 1997). The third is to propose the concept of student satisfaction with particular aspects during school, also referred to as the elemental concept. The idea places a focus on satisfaction at specific levels, including educational methods, living services, hardware facilities, and so on. As a development and improvement of the holistic idea, it also incorporates other elements such as perception of interaction, relationship, support for learning and development, and perception of gain (Liu and Fan, 1995).

With regard to student satisfaction, there is an overwhelming body of research that demonstrates that students have greater satisfaction with blended courses, compared with both traditional face-to-face and fully online modes of education (Owston et al., 2006; Collopy and Arnold, 2009; Castle and McGuire, 2010; Farley et al., 2011). Research on the influencing factors of blended learning satisfaction shows that having continuous access to the instructor is perceived as an important factor in students' satisfaction with blended learning (Martinez-Caro and Campuzano-Bolarin, 2011). Some students report that they receive instructor feedback and their grades faster than in traditional courses (Korr et al., 2012). Taghizadeh and Hajhosseini (2021) investigated 140 TEFL (Teaching English as a Foreign Language) students' attitudes, interactions, and relationships with blended learning quality. They found that students were most satisfied with teacher-student interaction in blended learning, and had the largest impact coefficient on teaching quality and satisfaction. Among them, teachers' teaching ability contributed significantly more to satisfaction than teachers' teaching attitude. Wu Jen-Her through confirmatory factor analysis (CFA), Tennyson Robert D, and Hsia Tzyh-Lih (Wu et al., 2010) concluded that self-efficacy, performance expectations, system functions, content characteristics, interaction, and learning atmosphere are the main factors that affect students' learning satisfaction with BELS (Blended E-Learning System) in blended learning. While learning atmosphere and performance expectations significantly affect learning satisfaction, teacher-student interaction significantly affects learning atmosphere, and self-efficacy, system function, content characteristics, and interaction significantly affect performance expectations.

Research on constructivist learning theory

Constructivist learning theory first originated from the Swiss educationalist Piaget's theory of children's cognitive development and is now commonly considered to comprise Piaget's individual constructivism and Vygotsky's social constructivism. Gordon (2008), Neo and Neo (2009), and Felder (2012) hold the opinion that constructivism has emerged as a powerful theory for explaining how humans learn about the world around them and how new knowledge is formed. According to this theory, the formation of the learner's own knowledge and the development of cognitive structures do not result from the educator's face-to-face instruction, but rather from the interaction of the learner with the surrounding environment, learners constructing their own cognitive structures through both assimilation and conformity modes of behavior, and therefore it is fundamentally personal and depends on the learning environment, knowledgeable and experienced teachers, and experiences in social interactions (Grabinger and Dunlap, 1995). The process of constructing cognitive structures is a form of active absorption of knowledge by the learner, instead of being passive, emphasizing the subjectivity of the learner (Gordon, 2008). Constructivist learning theory indicates that the learning environment plays a crucial role for learners as they continually build and update their own knowledge systems. Individuals who want to better build a learning system in line with their own are to a certain extent influenced by the learning environment, and "scenario," "cooperation," "conversation," and "meaning construction" are the four major factors of the learning environment.

The teacher's unilateral transfer of knowledge to students in traditional instruction is transformed in blended learning, as students shift from passive receivers to active constructors through the continual interplay of their existing and new knowledge. By providing rich and diversified resources, blended learning creates attractive learning scenarios that students find appealing, encourages their participation and initiative in communication and interaction, piques their interest in learning, and makes them active participants in their education.

Research on the complex adaptive blended learning system

The Complex Adaptive Blended Learning System first arose in the disciplines of physics, mathematics, and chemistry to construct a framework for complex adaptive systems, which was intended to be used to enhance knowledge of dynamic and complex topics and nonlinear systems such as neural, ecological, galactic, and social systems (Hadzikadic et al., 2010). Complex adaptive systems are described as living open systems that "exchange matter, energy, or information across boundaries and use this energy exchange to maintain their structure" (Cleveland, 1994).

To fill a gap in blended learning research and further enhance the systematic understanding of blended learning research and practice, Wang et al. (2015) call for a systematic analysis of blended learning (BL) programs by adopting a framework called the Complex Adaptive Blended Learning System (CABLS), which, they argue, "helps to reveal the untapped potential and key issues … such as the provision of learning support, the promotion of institutional engagement, and the non-linear relationships among subsystems in blended learning." Six learning elements are therefore proposed through the Complex Adaptive Blended Learning System (CABLS): Learner, Teacher, Technology, Content, Learning Support, Institution, each with its own characteristics and subsystems (Cleveland-Innes and Wilton, 2018).

These six subsystems will interact in a nonlinear and dynamic manner, consistent with other complex adaptive systems. "Its own characteristics and internal driving forces, depending on surrounding subsystems, to maintain its vitality" (Wang et al., 2015). At the same time, each of these subsystems has its own characteristics or features that are self-motivating while interdependent to maintain competitiveness, and they will interact with each other to form a blended learning system. The CABLS framework aims to "promote a deeper and more accurate understanding of the dynamics and adaptability of hybrid learning" (Wang et al., 2015).

Using this framework, Wang et al. (2015) reviewed a 20-month period (January 2013 to August 2014) and found that research on BL was scattered over that time period, with 95% of the review articles focused on learners, followed by content (79%) and technology (54%). This percentage drops sharply when it comes to focusing on teachers (32%), institutions (17%), and learning support (15%), and there is no systematic study of the links between these elements since most studies only examine one element. Therefore, there is a need to focus on the interconnections between elements and to conduct a more in-depth study of the BL situation.

In summary, the Complex Adaptive Blended Learning System proposes that a set of logically interrelated and hierarchically different systems can be formed through effective classification of six learning elements. The resulting classification system generally has three characteristics: first, the different categories are logical; second, the research results are consistent with previous research results; and third, the classification results are as complete as possible and can cover different categories. According to the topic and specificity of this study, therefore, this study analyzes and organizes the relevant influencing factors in three dimensions: learning dimension, teaching dimension, and curriculum dimension. The learning dimension covers learning attitudes and learning behaviors. The teaching dimension covers teaching attitudes, teaching methods, and teaching abilities. The curriculum dimension covers curriculum design and platform design. These three dimensions cover the six subsystems of the Complex Adaptive Blended Learning System that are more likely to facilitate "deep learning," and the research findings can assist us in making better decisions when implementing blended learning, and effectively support our learners in achieving deeper and more meaningful learning.

Methods and investigations

Data collection and sample description

A questionnaire survey was conducted between September 2021 and January 2022 among 650 college students with blended learning experiences in six universities in Sichuan Province, China, through online and paper questionnaires to ensure the accuracy of the data sample. A total of 650 questionnaires were distributed, and 628 questionnaires were collected, with a return rate of 97%. After checking the missing values of the questionnaires, it was found that a total of 30 questionnaires were missing incomplete. Therefore, 598 valid questionnaires were obtained, with an effective rate of 95%. Among the 598 samples, in terms of gender ratio, 325 female students accounted for 54.3% of the total and 273 male students accounted for 45.7%, with a larger proportion of female students. In terms of grade ratio, 206 freshmen accounted for 34.4%, 147 sophomores accounted for 24.6%, 125 juniors accounted for 20.9%, and 120 seniors accounted for 20.1%, with a large proportion of freshmen, but the overall distribution was balanced. In terms of major categories, humanities and social sciences, science and technology, agriculture and medicine, and arts and sports accounted for 33.9%, 33.3%, and 32.8%, respectively. The distribution of students' major categories in the survey is more balanced. In terms of politic countenance, the total number of students who are members of the Communist Youth League is 472, accounting for 78.9%, which is significantly higher than that of members of the Communist Party of China (including probationary Party members) and the masses.

Questionnaire design

This study is based on a questionnaire survey to quantify the indicators and obtain relevant data. The questionnaire was designed in a targeted and quantitative way by consulting relevant books and materials, and combining them with the actual situation of blended learning in universities. The scale index design of this study is divided into the following two stages: first, after discussing with experts and teachers responsible for the construction of blended learning courses in the university, as well as combining the opinions and suggestions of students who have blended learning experiences, continuous modifications are made in terms of suitability, classification, and semantics, and so on; Secondly, a pre-survey was conducted to distribute questionnaires in a certain range, and the content of the questionnaires was analyzed using the high-low grouping method to ensure that the significance level of each indicator was less than 0.05. Finally, a total of 31 questions were used to construct the "College Blended Learning Student Satisfaction Scale" from 12 indicators, which comprise learning platform, online teaching resources, online teaching forms, offline teaching methods, students' learning attitudes and learning behaviors during learning, teachers' teaching attitudes, teaching methods, and teaching abilities, curriculum design and platform design, and teaching satisfaction, and so on. The scale is divided into three parts: the first part is personal information, which includes gender, grade, major category, and politic countenance; the second part is a survey on the current situation of blended courses, which aims to investigate whether there is a significant distinction between various teaching processes on students' satisfaction; the third and fourth parts are the main questionnaire (see Table 1), which includes learning dimension, teaching dimension, curriculum dimension and satisfaction of teaching, and the variables are all scored on a five-point Likert scale, corresponding to scores from one to five, and respondents were able to choose the most appropriate answer for their actual situation.

Exploratory factor analysis and validity testing

The scales in the third and fourth parts were not designed as classical scales, wherein the principal component analysis of exploratory factors was used to verify the correspondence between the items and the variables. When KMO (Kaiser-Meyer-Olkin) > 0.9, the effect of factor analysis is ideal (Wu, 2003). According to the test results of different question items, the total KMO value of the questionnaire is $0.974 > 0.9$, the Sig value is $0.00 < 0.05$, the cumulative explained variance is 79.73%, and the variance explained after rotation of each factor was above 10%, indicating that the exploratory factor

TABLE 1 Questionnaire structure of student satisfaction and influencing factors of blended learning in universities.

Dimension		Indicators		Corresponding question number
Part I	Personal information	Gender, grade, major category, and politic countenance		A1–A4
Part II	Current situation of blended courses	Learning platform		B1
		Online teaching resources		B2
		Online teaching format		B3
		Offline teaching method		B4
Part III	Learning dimension	Learning attitude	Learning interest	C1
			Learning awareness	C2
			Learning self-efficacy	C3
		Learning behavior	Learning concentration	C4
			Learning reflection	C5
			Level of communication and interaction	C6
	Teaching dimension	Teaching attitude	Careful preparation before and after class	C7
			Timely feedback	C8
		Teaching method	Active classroom interaction	C9
			Rich teaching methods	C10
		Teaching ability	Information technology capability	C11
			Teaching and research capabilities	C12
	Curriculum dimension	Course design	Richness of course content	C13
			Reasonableness of course objectives	C14
			Appropriateness of online and offline distribution	C15
			The scientific nature of the assessment	C16
		Platform design	Ease of platform operation	C17
			Completeness of platform functions	C18
Part IV	Satisfaction of teaching			D1–D5

analysis results are good. The factor loading coefficients of each factor are higher than 0.6, which indicates good overall scale validity (Table 2).

This study adopts Cronbach's alpha reliability analysis. The Cronbach's alpha of the three dimensions of students, teachers, and curriculum in the third part are 0.945, 0.976, and 0.969 respectively, and the Cronbach's alpha of teaching satisfaction in the fourth part is 0.975, the alpha coefficient of each dimension was above 0.9. The α coefficients of each dimension are above 0.9, which shows that the questionnaire has good reliability (Table 3).

TABLE 2 Questionnaire validity and cumulative explained variance analysis.

Cumulative Explained Variance		79.730%
Kaiser-Meyer-Olkin measure of sampling adequacy		0.974
Bartlett's Sphericity test	Chi-square approximation	20410.570
	df	325
	Sig.	0.000

Results and date analysis

Analysis of students' satisfaction with blended learning in universities

The survey data is shown in Table 4, from which the average value of satisfaction with blended learning is 3.49, which is at the upper-middle level. Among them, more than 60% of the students agreed or strongly agreed with three indicators, which shows that satisfaction with blended learning in universities is good. The students who agree and strongly agree with the indicator "In general, I am satisfied with the blended learning." are the most, reaching 61.7%, and the students who agree and strongly agree with the indicator "By studying this course, I have completed the expected learning objectives or tasks." are the least, only 50.8%, indicating that nearly half of the students are dissatisfied with the learning objectives and completion of learning tasks after the course.

Satisfaction analysis based on three dimensions of teaching, learning, and curriculum

Table 5 displays the results of the satisfaction survey for the three dimensions of teaching, learning, and curriculum, and college students are the least satisfied with their own learning status.

TABLE 3 Questionnaire validity analysis.

Indicators		Cronbach's alpha	Number of projects
Part III	The credibility of learning dimension	0.945	6
	The credibility of teaching dimension	0.976	6
	The credibility of curriculum dimension	0.969	6
Part IV	The credibility of satisfaction of teaching	0.975	5
Overall Credibility of the Questionnaire		0.985	23

TABLE 4 Satisfaction rate of blended learning in universities (%).

Indicator	Strongly disagreed	Disagree	Neutral	Agree	Strongly agree	Mean value	Population mean	Standard deviation
In general, I am satisfied with the blended learning.	2.0	13.5	22.7	43.3	18.4	3.63	3.49	0.855
If possible, I will choose blended learning for learning.	1.8	13.4	24.6	44.1	16.1	3.59		
I would recommend this blended course and its instructor to others.	1.8	16.6	20.2	40.5	20.9	3.61		
Compared with traditional learning and online learning alone, blended learning g has more advantages.	2.0	26.1	21.6	40.0	10.4	3.35		
By studying this course, I have completed the expected learning objectives or tasks.	1.7	22.6	25.9	43.3	7.5	3.31		

TABLE 5 Satisfaction statistics of the three dimensions of teaching, learning, and curriculum in blended learning in universities.

	Index mean	Index standard deviation	Population mean	Standard deviation
Learning attitude	3.38	0.956	3.36	0.819
Learning behavior	3.35	0.866		
Teaching attitude	3.66	0.989	3.62	0.843
Teaching method	3.61	0.889		
Teaching ability	3.61	0.927		
Course design	3.56	0.861	3.59	0.852
Platform design	3.62	0.927		

Tables 6–8 display the satisfaction rates for each indicator of the three dimensions. There are obvious differences in their satisfaction with self-attitude and self-behavior, which reflect the disunity of knowing and doing. The mean value of students' satisfaction with teachers' teaching attitude is greater than the mean value of satisfaction with teaching methods and teaching ability, which indicates that teachers' teaching techniques and abilities need to be further optimized. In terms of curriculum, the design of teaching contents, the hours and the frequency of instruction fall short of properly fulfilling the needs of the students.

Differential analysis of students' satisfaction with blended learning in universities

The analysis of the variability of satisfaction based on individual characteristics of college students by independent sample t-test and one-way ANOVA test results is shown in Table 9. The significant p-value is greater than 0.05, while the results show that no significant difference exists between the satisfaction of different personality characteristics.

One-way ANOVA were conducted on the satisfaction of teaching on different learning platforms, different online resources, different online teaching forms, and different offline teaching methods, and the details are shown in Tables 10–13. The results show that there is no significant difference in students' satisfaction with blended learning depending on the learning platform. However, a significant difference exists in students' satisfaction with blended learning due to different online resources, online teaching forms, and offline teaching methods. Teacher-built fragmented courseware and teacher-built systematic courses, online teaching forms with fixed locations and time, and mutual discussions and student-led lectures are more popular among students. Adding technology to in-person teaching and learning may foster engagement and improve learning outcomes. According to the SAMR (Substitution, Augmentation, Modification, Redefinition)

TABLE 6 Satisfaction rate of each indicator in learning dimensions of blended learning in universities (%).

Indicator		Strongly disagreed	Disagree	Neutral	Agree	Strongly agree	Mean value
Learning attitude	I have some interest in blended learning courses.	7.7	16.9	34.8	24.6	16.1	3.24
	This course is very important for my future development.	2.7	13.4	28.6	35.1	20.2	3.57
	Before the course, I will set learning goals and be confident that I will be able to complete this course.	2.8	23.2	22.9	39.1	11.9	3.34
Learning behavior	I am more focused in online or offline courses alone.	4.2	13.7	36.5	28.9	16.7	3.40
	I often reflect on my studies in my spare time, judging my mastery of the course.	3.3	16.9	38.5	31.8	9.5	3.27
	During blended learning, I will have active communication and interactions with teachers and classmates in various ways.	2.7	15.6	34.4	35.1	12.2	3.38

TABLE 7 Satisfaction rate of each indicator in teaching dimensions of blended learning in universities (%).

Indicator		Strongly disagreed	Disagree	Neutral	Agree	Strongly agree	Mean value
Teaching attitude	In blended learning, teachers carefully prepare lessons before class and are responsible after class.	1.8	16.6	17.9	41.0	22.7	3.66
	Timely feedback on questions, discussions and assignments are provided by the teacher during blended learning.	2.7	13.7	19.6	42.6	21.4	3.66
Teaching method	Teachers frequently develop interactive environments and interactive scenarios during blended learning.	3.0	13.0	22.4	41.5	20.1	3.63
	In blended learning, teachers adopt various teaching methods such as lecture, case study, discussion and practice.	2.0	14.9	22.4	41.5	19.2	3.61
Teaching ability	Teachers integrate information technology deeply with the subjects in blended learning, and have good IT skills.	2.2	18.1	22.7	39.8	17.2	3.52
	In blended learning, the teachers are knowledgeable and have good teaching ability.	1.5	10.5	20.2	41.5	21.3	3.70

TABLE 8 Satisfaction rate of each indicator in curriculum dimensions of blended learning in universities (%).

Indicator		Strongly disagreed	Disagree	Neutral	Agree	Strongly agree	Mean value
Course design	The content and resources of both online and offline learning are rich and interesting, and can satisfy my desire to learn.	2.2	12.2	31.6	40.6	13.4	3.51
	The objectives of both online and offline learning are well designed	1.5	12.0	25.8	44.0	16.7	3.62
	Appropriate design of both online and offline learning (including the hours, frequency, interaction and connection of online and offline learning, etc.).	1.3	10.4	33.3	41.0	14.0	3.55
	The evaluation and assessment of the course is designed in a scientific way.	1.0	15.6	26.9	40.3	16.2	3.56
Platform design	Great convenience and ease of learning with the learning platform.	1.2	12.2	23.9	41.0	21.7	3.70
	The various functions of the learning platform (such as notes, tests and interactions, etc.) are complete, and I can make good use of them	1.8	11.9	24.1	41.8	17.4	3.64

model, blended learning adds technology to the online and offline classroom and is technology-supported learning. In general, blended learning in universities basically covers substitution, augmentation, and modification, and less for redefinition. In the comparative analysis, significant differences were found in students' satisfaction with blended instruction based on different offline teaching methods. Mutual discussion and student-led lectures were more popular among students. These two teaching methods have brought each other closer,

TABLE 9 Satisfaction statistics of college students with different characteristics on blended learning.

Characteristics	Category	*N*	Mean value	Standard deviation	T-value	F-value
Gender	Male	273	3.44	0.891	−1.238	
	Female	325	3.53	0.824		
Grade	Freshman	206	3.35	0.921		1.841
	Sophomore	147	3.36	0.789		
	Junior	125	3.52	0.708		
	Senior	120	3.79	0.885		
Major	Humanities and Social Science	203	3.52	0.839		1.464
	Science, Engineering, Agriculture and Medicine	199	3.50	0.860		
	Arts and Sports	196	3.44	0.869		
Politic countenance	Chinese Communist Party members (including probationary Party member)	54	3.62	0.806		1.384
	Communist Youth League member	472	3.48	0.880		
	The masses	72	3.38	0.709		

$*p < 0.05$, $***p \leq 0.001$.

TABLE 10 Statistics of students' satisfaction with blended learning on different learning platforms.

Platform Type		*N*	Mean value	Standard deviation	F-value
Super star learning APP	Not selected	83	3.32	0.946	1.633
	Select	515	3.51	0.838	
MOOC and client ware	Not selected	280	3.40	0.894	1.627
	Selected	318	3.64	0.754	
QQ group and WeChat group	Not selected	394	3.41	0.871	1.727
	Selected	204	3.55	0.838	
Ding talk	Not selected	481	3.47	0.865	1.092
	Selected	117	3.56	0.812	
Other	Not selected	534	3.48	0.834	1.822
	Selected	64	3.55	1.004	

$*p < 0.05$, $***p \leq 0.001$.

TABLE 11 Statistics of students' satisfaction with blended learning of different online resources.

Online resource type		*N*	Mean value	Standard deviation	F-value
Teacher-built systematic courses	Not selected	150	3.34	0.745	5.734*
	Selected	448	3.72	0.846	
National-level open course	Not selected	346	3.50	0.886	5.423*
	Selected	252	3.62	0.735	
Other courses	Not selected	397	3.45	0.813	5.645*
	Selected	201	3.57	0.886	
Teacher-built fragmented courseware	Not selected	401	3.56	0.834	6.547*
	Selected	197	3.87	0.847	

$*p < 0.05$, $***p \leq 0.001$.

improved students' interaction, communication, and engagement, given students more initiative, and made it easier for teachers to monitor students' learning progress and status. There were significant differences in student satisfaction with blended learning based on different online teaching forms. The mean value of student satisfaction with blended learning was highest among fixed-location and fixed-time modes of online teaching forms. The main distinction between synchronous and asynchronous learning is whether the time and location of learning are consistent, with synchronous learning guaranteeing students' learning pace and efficiency and asynchronous

TABLE 12 Statistics of students' satisfaction with blende learning in different online teaching forms.

Online teaching format		*N*	Mean value	Standard deviation	F-value
Unfixed locations, unfixed time	Not selected	145	3.67	0.953	6.334*
	Selected	453	3.44	0.868	
Unfixed locations, fixed time	Not selected	361	3.45	0.845	6.867*
	Selected	237	3.52	0.795	
Fixed locations fixed time	Not selected	556	3.45	0.865	5.235*
	Selected	42	3.78	0.878	

$^*p < 0.05$, $^{***}p \le 0.001$.

TABLE 13 Statistics of students' satisfaction with blended learning of different offline teaching methods.

Offline teaching method		*N*	Mean value	Standard deviation	F-value
Teacher-led lectures	Not selected	120	3.79	0.657	13.987***
	Selected	478	3.23	0.889	
Mutual discussions	Not selected	278	3.53	0.788	9.6044***
	Selected	320	3.67	0.722	
Student-led lectures	Not selected	302	3.49	0.546	10.4554***
	Selected	296	3.25	0.912	

$^*p < 0.05$, $^{***}p \le 0.001$.

learning testing students' self-motivation and autonomy. There is a problem with online learning integrity due to the irresponsible behavior of some students toward learning. Students simply play learning videos without carefully watching them, affecting their learning effectiveness, which goes against the original purpose of online learning.

Analysis and discussion

Correlation analysis of influencing factors

To explore the correlation between student satisfaction with the three factors of the learning dimension, teaching dimension, and curriculum dimension of blended learning in universities, this study processed the data of the three-dimensional scales accordingly through Pearson correlation analysis and multiple linear regression (MLR) to ensure that the factors affecting satisfaction with blended learning in universities could be reflected. Pearson correlation analysis was conducted between the three influencing dimensions (learning dimension, teaching dimension, and curriculum dimension) and students' satisfaction, and the results are shown in Table 14. According to the results of correlation analysis, the *p*-values corresponding to the satisfaction of blended learning and the three factors of learning dimension, teaching dimension, and curriculum dimension are all close to 0, indicating that the learning dimension, teaching dimension, and curriculum dimension are significantly positively correlated with students' satisfaction with blended learning, i.e., the increase in student satisfaction with themselves, teachers, and curriculum would be followed by an increase in satisfaction with blended learning. The correlation coefficients are less than 0.70, so there is no multicollinearity phenomenon. The correlation coefficients of learning, teaching, and curriculum with students' satisfaction were 0.457, 0.331, and 0.414, respectively, which shows that the correlation between the learning dimension and students' satisfaction is the strongest, followed by the curriculum dimension, and the teaching dimension is the weakest.

Regression analysis of influencing factors

The correlation analysis yields significant correlations among the variables. The next step of regression analysis is carried out. The multiple linear regression (MLR) analysis method is adopted to verify whether a significant correlation exists between the learning dimension, the teaching dimension, the curriculum dimension at all levels, and the satisfaction of teaching, and then to understand the impact of each dimension on the satisfaction with blended learning. The learning dimension includes learning attitude and learning behavior, the teaching dimension includes teaching attitude, teaching method, and teaching ability, and the curriculum dimension includes course design and platform design, so as to establish a multiple linear regression equation. Taking the satisfaction with blended learning as the dependent variable Y, a regression analysis is conducted by including X1 (satisfaction with learning attitude), X2 (satisfaction with learning behavior), X3 (satisfaction with teaching attitude), X4 (satisfaction with teaching method), X5 (satisfaction with teaching ability), X6 (satisfaction with course design), and X7 (satisfaction with platform design) in turn. Results are shown in Table 15. R2 is 0.605, which is significantly greater than 0, representing that the regression model is scientific. It also shows that the independent variables formulated in this study, including the learning dimension (learning attitude, learning behavior), teaching dimension (teaching attitude, teaching method, teaching ability), and curriculum dimension (course design, platform design), can reflect 62.5% of the degree of change in the satisfaction of teaching. The *F*-value of the regression model is

TABLE 14 Correlation Matrix between influencing factors and satisfaction of blended learning.

	Satisfaction of blended learning	Learning dimension	Teaching dimension	Curriculum dimension
Satisfaction of blended learning	1			
Learning dimension	0.457**	1		
Teaching dimension	0.331**	0.458**	1	
Curriculum dimension	0.414**	0.411**	0.320**	1

**Significant at the 0.01 level.

TABLE 15 Regression analysis of factors on the satisfaction with blended learning (beta value).

	Satisfaction with blended learning
Learning dimensions	
Learning attitude	0.221***
Learning behavior	0.125**
Teaching dimensions	
Teaching attitude	0.165***
Teaching method	0.181***
Teaching ability	0.116*
Curriculum dimensions	
Course design	0.192***
Platform design	0.096*
B	2.450
F	140.206***
Adjusted R2	0.605

*0.01 < sig < 0.05; **0.001 < sig ≤ 0.01; ***sig ≤ 0.001.

140.206, and the corresponding significance level is 0.000 < 0.001, indicating a significant linear relationship between the predictor variables (learning attitudes, learning behaviors, teaching attitudes, teaching methods, teaching abilities, course design, and platform design) and the dependent variable satisfaction with blended learning. From the data in the table, the sig value corresponding to the t-statistic of any constant X is less than 0.05, in which all passed the t-test at the significance level of 0.05, indicating that this linear regression model is meaningful.

From the Beta values, the regression coefficient of X1 (satisfaction with learning attitude) is 0.221; the regression coefficient of X2 (satisfaction with learning behavior) is 0.125; the regression coefficient of X3 (satisfaction with teaching attitude) is 0.165; the regression coefficient of X4 (satisfaction with teaching method) is 0.181; the regression coefficient of X5 (satisfaction with teaching ability) is 0.116; the regression coefficient of X6 (satisfaction with course design) is 0.192; the regression coefficient of X7 (satisfaction with platform design) is 0.096. That is, if the students' satisfaction index of learning attitude, learning behavior, teaching attitude, teaching method, teaching ability, course design, and platform design all increase by 1 standard unit, the satisfaction index of blended learning will increase by 0.221, 0.125, 0.165, 0.181, 0.116, 0.192, and 0.096 basic units, respectively. Thus, learning attitudes, course design, and teaching methods are the factors that have the largest influence on the satisfaction of blended learning, whereas platform design has the least influence.

According to the results of the regression analysis in Table 15, the following regression equation was derived between the variables:

Satisfaction with Blended Learning = 2.450 + 0.221 * Learning Attitude + 0.125 * Learning Behavior + 0.165 * Teaching Attitude + 0.181 * Teaching Method + 0.116 * Teaching Ability + 0.192 * Course Design + 0.096 * Platform Design.

Conclusion

College students are relatively satisfied with blended learning, and the satisfaction level of each dimension in learning is likewise in the upper-middle range, indicating that the majority of students are in favor of blended learning. According to the data, students are most satisfied with the teacher, followed by the course, and least satisfied with themselves. Specifically, students' satisfaction with teachers' teaching attitude, online learning platform, and self-learning awareness is high and at a good level. However, there are issues like students' low interest in learning, lack of learning action, and failure to meet students' needs in terms of teaching methods, teaching content resources, and learning platform technology. To support the development of blended learning and boost teaching effectiveness, further exploration and problem-solving are necessary. In some key statistics, there are notable differences in the satisfaction with blended learning. Specifically, there are significant differences in satisfaction ratings among different online resources, different online teaching forms, and different offline teaching methods. In terms of individual characteristics like grades, majors, and learning platforms, there is no significant difference in the satisfaction rating. For the influencing factors of college students' satisfaction with blended learning, the results of correlation analysis show that the learning dimension has the strongest correlation with student's satisfaction, followed by the curriculum dimension, and the teaching dimension has the weakest correlation. Thus, further regression analysis reveals significant positive correlations between student's satisfaction with blended learning and learning attitude, learning behavior, teaching attitude, teaching method, teaching ability, curriculum design, and platform design. Among them, learning attitudes, course design, and teaching methods have a greater influence on students' satisfaction with blended learning, and platform design has the least influence.

In the comparative analysis, there is a significant difference in students' satisfaction with the blended learning of the Civics and Political Science class based on different offline teaching methods. The mutual discussion and communication style, as well as the student-led teaching style, are more popular among students because these two teaching methods emphasize the students' initiative and allow students to fully interact with their classmates and teachers, and it also enable the teacher to better tailor their teaching to the students' actual situation. Therefore, it is advocated to put interactive communication at the core of classroom teaching. The online teaching process can attract students' attention and increase their interest in learning through questions, answers, discussions, and game-based activities. Offline teaching is problem-oriented, integrating communication and discussion into classroom teaching as a necessary part, increasing students' participation, and focusing on cultivating students' ability to discern, think, and express. Based on different online teaching forms, there were significant differences in students' satisfaction with blended learning in the Civics courses. The mean value of student satisfaction with blended learning is highest in the online teaching form with a fixed location and fixed time. Due to integrity issues with students' online learning and their irresponsible behavior toward learning, the original purpose of online learning is being compromised. Therefore, online learning should strengthen its supervision of students' learning. This can be achieved by giving students some degree of "freedom" while continually monitoring their selection of the online learning mode, time, and location, and guiding them to make organized arrangements for their own learning that fall within a reasonable range. Additionally, we should teach students self-awareness of online learning, independent learning behaviors, and the ability to promptly adjust their learning progress in response to the course dynamics.

The results of Pearson correlation analysis show that the learning dimension has the strongest correlation with satisfaction and is the primary factor influencing students' satisfaction with blended learning. The findings of the regression analysis further support the idea that students' satisfaction with blended learning is significantly influenced by their learning behavior and learning attitude in the learning dimensions. Learning behaviors and learning attitudes mainly comprise important factors such as learning interest, learning awareness, learning self-efficacy, learning concentration, learning reflection, and the level of communication and interaction, which together affect students' satisfaction with blended learning in universities. Our results are in alignment with the literature reporting that achievement in blended courses is influenced to a greater extent by students' conceptions of learning, their ability to accept responsibility for their learning, and the degree of interactivity outside of the classroom (Mitchell and Honore, 2007; Moore and Gilmartin, 2010; Bliuc et al., 2011; Chou and Chou, 2011; Smyth et al., 2012). In comparison with the average satisfaction of teachers, the average satisfaction of teachers' teaching attitudes is greater than the average satisfaction of teaching methods and teaching abilities, and the impact of teachers' teaching methods is greater, these factors include teachers' positive or negative attitudes toward technology use, learners' proficiency levels, teachers' training, teachers' and students' accessibility to technology, and cost. Each one of these factors plays a vital role in decisions regarding implementing a blended learning approach in language classrooms (Sharma and Barrett, 2008). Students have the lowest mean value of satisfaction with themselves when compared to the mean value of satisfaction with teachers and the mean value of satisfaction with courses. The student's actual performance in class, which serves as the focus of educational activities, has an important influence on their satisfaction. Only when the student's interest in learning is completely mobilized can a positive teaching effect be realized. In response to the issues of students' low self-satisfaction, lack of learning interest and learning motivation in blended learning, it is necessary not only to impart relevant knowledge during actual teaching but also to foster the students' enjoyment in learning knowledge on the basis of it.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XC wrote the original manuscript. XC and WM performed the data collection and treatment. YD reviewed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Chengdu University of Technology "Double First-Class" initiative Construction Philosophy and Social Sciences Key Construction Project, No: ZDJS202306, Special Project of "Research on Informatization of Higher Education" of the Chinese Society of Higher Education, and Sichuan Province Chengdu University of Technology 2022 Talent Cultivation Quality and Teaching reform project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ates, R. (2009). The handbook of blended learning: global perspectives, local designs. *Turk. Online J. Dist. Educ.* 10, 66–76. doi: 10.1002/9781118269381

Bliuc, A. M., Ellis, R. A., Goodyear, P., and Piggott, L. (2011). A blended learning approach to teaching foreign policy: student experiences of learning through face-to-face and online discussion and their relationship to academic performance. *Comput. Educ.* 56, 856–864. doi: 10.1016/j.compedu.2010.10.027

Carver, S., Evans, A., and Kingston, R. (2004). Developing and testing an online tool for teaching GIS concepts applied to spatial decision-making. *J. Geogr. High. Educ.* 28, 425–438. doi: 10.1080/0309826042000286983

Castle, S. R., and Mcguire, C. (2010). An analysis of student self-assessment of online, blended, and face-to-face learning environments: implications for sustainable education delivery. *Int. Educ. Stud.* 3, 36–40. doi: 10.5539/ies.v3n3p36

Chang, S. C., and Tung, F. C. (2008). An empirical investigation of students' behavioral intentions to use the online learning course websites. *Br. J. Educ. Technol.* 39, 71–83. doi: 10.1111/j.1467-8535.2007.00742.x

Chen, M. Y., Yang, Y. P., Chen, M., and Chen, Z. H. (2019). Research on the construction and practice of blended learning space under the vision of smart education. *China Dist Educ* 11:7.

Cheng, Y. C., and Tam, W. M. (1997). Multi-models of quality in education. *Qual. Assur. Educ.* doi: 10.1108/09684889710156558

Chou, A. Y., and Chou, D. C. (2011). Course management systems and blended learning: an innovative learning approach. *Decis. Sci. J. Innov. Educ.* 9, 463–484. doi: 10.1111/j.1540-4609.2011.00325.x

Cleveland, J. (1994). Complexity theory: basic concepts and application to systems thinking. Available at: http://www.slideshare.net/johncleveland/complexity-theory-basic-concepts

Cleveland-Innes, M., and Wilton, D. (2018). *Guide to blended learning*. Toronto: AlphaPlus centre.

Collopy, R., and Arnold, J. M. (2009). To blend or not to blend: online and blended learning environments in undergraduate teacher education. *Iss Teach Educ* 18, 85–101.

Farley, A., Jain, A., and Thomson, D. (2011). Blended learning in finance: comparing student perceptions of lectures, tutorials and online learning environments across different year levels. *Econ Papers* 30, 99–108. doi: 10.1111/j.1759-3441.2010.00094.x

Felder, R. M. (2012). Engineering education: A tale of two paradigms. *Shaking the Foundations of Geo-Engineering Education*, 9–14.

Garrison, D. R., and Kanuka, H. (2004). Blended learning: uncovering its transformative potential in higher education. *Internet High. Educ.* 7, 95–105. doi: 10.1016/J.IHEDUC.2004.02.001

Garrison, D. R., and Vaughan, N. D. (2008). Blended learning in higher education: framework, principles, and guidelines. *J. Phys. Ther. Educ.* 13–30. doi: 10.1002/9781118269558

Gerbica, P. (2011). Teaching using a blended approach – what does the literature tell us? *Educ. Media Int.* 48, 221–234. doi: 10.1080/09523987.2011.615159

Gordon, M. (2008). Between constructivism and connectedness. *J. Teach. Educ.* 59, 322–331. doi: 10.1177/0022487108321379

Grabinger, R. S., and Dunlap, J. C. (1995). Rich environments for active learning: a definition. *Res. Learn. Technol.* 3:9606. doi: 10.3402/rlt.v3i2.9606

Guolla, A., and Michael, S. (1999). Assessing the teaching quality to student satisfaction relationship: applied customer satisfaction research in the classroom. *J. Mark. Theory Pract.* 7, 87–97. doi: 10.1080/10696679.1999.11501843

Hadzikadic, M., Carmichael, T., and Curtin, C. (2010). Complex adaptive systems and game theory: an unlikely union. *Complexity* 16, 34–42. doi: 10.1002/cplx.20316

Korr, J., Derwin, E. B., Greene, K., and Sokoloff, W. (2012). Transitioning an adult-serving university to a blended learning model. *J Contin High Educ* 60, 2–11. doi: 10.1080/07377363.2012.649123

Linder, K. E. (2017). Fundamentals of Hybrid Teaching and Learning. *New Dir. Teach. Learn.* 2017, 11–18. doi: 10.1002/tl.20222

Liu, J., and Fan, G. R. (1995, 2019). The mechanism of teacher-student interaction on students' self-realization and satisfaction in the context of the reform of "Curriculum Civics" in higher education--an empirical analysis based on structural equation modeling. *Modern Educ Manage* 350, 117–123. doi: 10.16697/j.cnki.xdjygl.2019.05.021

Martinez-Caro, E., and Campuzano-Bolarin, F. (2011). Factors affecting students' satisfaction in engineering disciplines: Traditional vs. blended approaches. *Eur. J. Eng. Educ.* 36, 473–483. doi: 10.1080/03043797.2011.619647

Ministry of Education of the People's Republic of China. (2019). *Implementation Opinions of the Ministry of Education on the Construction of First-class Undergraduate Curriculum*. Beijing: State Council Bulletin. 612.

Mitchell, P., and Forer, P. (2010). Blended learning: The perceptions of first-year geography students. *J. Geogr. High. Educ.* 34, 77–89. doi: 10.1080/03098260902982484

Mitchell, A., and Honore, S. (2007). Criteria for successful blended learning. *Industr Commerc Train Industr Commerc Train* 39, 143–149. doi: 10.1108/00197850710742243

Moore, N., and Gilmartin, M. (2010). Teaching for better learning: A blended learning pilot project with first-year geography undergraduates. *J. Geogr. High. Educ.* 34, 327–344. doi: 10.1080/03098265.2010.501552

Neo, M., and Neo, T. K. (2009). Engaging students in multimedia-mediated constructivist learning- students' perceptions. *Educ. Technol. Soc.* 12, 254–266.

Oliver, R. L. (1997). *A Behavioral Perspective on the Consumer*. Routledge.

Owston, R. D., Garrison, D. R., and Cook, K. (2006). "Blended learning at Canadian universities: issues and practices" in *The Handbook of Blended Learning: Global Perspectives, Local Designs*. eds. C. J. Bonk and C. R. Graham (San Francisco, CA: Pfeiffer), 338–350.

Pape, L. (2010). Blended teaching and learning. *Educ. Digest.* 67, 16–21. doi: 10.1007/978-0-387-34731-8_24

People's Daily. (2023). MOOCs Open Up A New Space for Digital Education. Available at: http://www.moe.gov.cn/jyb_xwfb/s5148/202301/t20230119_1039731.html?authkey=boxdr3.

Picciano, A. G., and Dziuban, C. D. (2007). *Blended Learning: Research Perspectives*. Springer: Netherlands.

Sadeghi, R., Sedaghat, M. M., and Ahmadi, F. S. (2014). Comparison of the effect of lecture and blended teaching methods on students' learning and satisfaction. *J. Adv. Med. Educ. Prof.* 2:146.

Sharma, P., and Barrett, B. (2008). *Blended learning: Using technology in and beyond the language classroom*. Oxford, UK: Macmillan education.

Sivo, S. A., Pan, C. C., and Hash-Vaughn, D. L. (2010). Combined longitudinal effects of attitude and subjective norms on student outcomes in a web-enhanced course: A structural equation modelling approach. *Br. J. Educ. Technol.* 38, 861–875. doi: 10.1111/j.1467-8535.2006.00672.x

Smyth, S., Houghton, C., Cooney, A., and Casey, D. (2012). Students' experiences of blended learning across a range of postgraduate programmers. *Nurse Educ. Today* 32, 464–468. doi: 10.1016/j.nedt.2011.05.014

Taghizadeh, M., and Hajhosseini, F. (2021). Investigating a blended learning environment: Contribution of attitude, interaction, and quality of teaching to satisfaction of graduate students of TEFL. *Asia Pac. Educ. Res.* 30, 459–469. doi: 10.1007/s40299-020-00531-z

Tayebinik, M., and Puteh, M. (2013). Blended Learning or E-learning? In: *International Magazine on Advances in Computer Science and Telecommunications (IMACST)*, 3, 103–110.

Wang, Y., Han, X., and Yang, J. (2015). Revisiting the blended learning literature: using a Complex Adaptive Systems Framework. *Educ. Technol. Soc.* 18, 380–393.

Wei, L. S., and Tan, Y. P. (2021). Tang Yifa. Research on the construction of blended curriculum based on the characteristics of higher vocational teaching. *Educ Career* 3:17. doi: 10.13615/j.cnki.1004-3985.2021.03.017

Wu, M. L. (2003). *SPSS statistical applications in practice: questionnaire analysis and applied statistics*. Beijing: Science Press.

Wu, J. H., Tennyson, R. D., and Hsia, T. L. (2010). A study of student satisfaction in a blended e-learning system environment. *Comput. Educ.* 55, 155–164. doi: 10.1016/j.compedu.2009.12.012

Xinhuanet. (2021). Wu Yan: Do well in the "new infrastructure" of education and take the road of independent talent nurturing. Available at: http://education.news.cn/2021-12/02/c_1211471188.htm.

Xu, Z., and Shi, Y. (2018). Application of constructivist theory in flipped classroom-take college English teaching as a case study. *Theory Pract Lang Stud* 8, 880–887. doi: 10.17507/tpls.0807.21

Zhu, Z. T., and Hu, J. (2021). Innovation in epidemic situation education after technology empowerment: a new pattern of online and offline integrated teaching. *Open Educ Res* 1, 13–23. doi: 10.13966/j.cnki.kfjyyj.2021.01.002

Zhu, Z. T., and Meng, Q. (2003). Blended learning in distance education. *China Dist. Educ.* 19:7.

Zhu, L. C., Wang, N., and Du, Y. T. (2020). Research on College Students' Online Learning Satisfaction and Its Influencing Factors and Improving Strategies. *J. Natl. Acad. Educ. Adm.* 269, 82–88. doi: 10.3969/j.issn.1672-4038.2020.05.013

TYPE Original Research
PUBLISHED 10 July 2023
DOI 10.3389/fpsyg.2023.1146458

OPEN ACCESS

EDITED BY
Shuhua An,
California State University, Long Beach,
United States

REVIEWED BY
Chang-Hua Chen,
National Changhua University of Education,
Taiwan
Mehmet Yüksel,
Gazi University, Türkiye
Antonio Luque,
University of Almeria, Spain
Miriam Marleen Gebauer,
University of Bamberg, Germany

*CORRESPONDENCE
Elisabeth Rathgeb-Schnierer
rathgeb-schnierer@mathematik.uni-kassel.de

RECEIVED 17 January 2023
ACCEPTED 23 June 2023
PUBLISHED 10 July 2023

CITATION
Brumm L and Rathgeb-Schnierer E (2023) The relationship between accuracy in numerosity estimation, math achievement, and math interest in primary school students.
Front. Psychol. 14:1146458.
doi: 10.3389/fpsyg.2023.1146458

The relationship between accuracy in numerosity estimation, math achievement, and math interest in primary school students

Leonie Brumm and Elisabeth Rathgeb-Schnierer*

Department of Mathematics and Natural Science, University of Kassel, Kassel, Germany

Estimation is a primary activity in everyday life, so getting it "right" in primary school provides a foundational basis in mathematical reasoning. This study focuses on numerosity estimation in primary mathematics, which is one of four types of estimation reported in literature. In numerosity estimation, a non-numerical quantitative representation is typically translated into a number. While it is assumed that fostering numerosity estimation has a great impact on the development of mathematical skills, research indicates that math achievement is influenced by non-cognitive aspects such as students` math interests. So, math interest could also influence the accuracy in numerosity estimation. In this study, we investigate the relationship between accuracy in numerosity estimation, math achievement, and math interest in third-grade students. For capturing accuracy in numerosity estimation in a standardized way, we developed an online numerosity estimation test. For assessing the construct of math interest, we used an existing questionnaire. Math achievement was assessed by a standardized math test that includes two subtests focusing on arithmetic and application tasks. The sample was comprised of 185 third-grade students. We analyzed the data using correlation and multiple linear regression analysis. The results showed a significant positive correlation between math interest and math achievement. However, no relationship was found between accuracy in numerosity estimation and math interest nor between accuracy in numerosity estimation and math achievement. These partly unexpected findings suggest further studies dedicated to numerosity estimation and its relationship to other constructs.

KEYWORDS

estimation, numerosity estimation, math achievement, math interest, primary education, mathematics

1. Introduction

Estimation is a field of interest for educational and cognitive psychology researchers (Dowker, 2003; Verschaffel et al., 2007). On the one hand, it is an essential part of mathematical cognition, and on the other hand, it bears a solid connection to mathematical procedures (Siegler and Booth, 2005). In addition, estimation is relevant for everyday activities in the lives of children and adults and is an essential core activity in everybody's life (Siegler and Booth, 2005; Andrews et al., 2021). Recent literature distinguishes four types of estimation: measurement estimation, computational estimation, numerosity estimation, and number line estimation (Sayers et al., 2020). The present study focuses on numerosity estimation and aims to analyze the connection between numerosity estimation and math achievement as well as the relationship between math interest as a non-cognitive process and accuracy in numerosity estimation.

1.1. Numerosity estimation

Numerosity estimation is considered one of four types of estimation (Sayers et al., 2020), each of which shares fundamental characteristics. For this reason, we will first define estimation in general and subsequently focus on characteristics that are particularly relevant for numerosity estimation. Overall, estimation can be defined as mental comparison and measurement (Sowder and Wheeler, 1989; Schipper, 2009; Ruwisch, 2014) characterized by "a process of translating between alternative quantitative representations, at least one of which is inexact" (Siegler and Booth, 2005, p. 204). Within all situations and tasks that call for producing an estimate, no exact answer is required, and an approximate answer is sufficient (Sowder and Wheeler, 1989). Estimation may be the most efficient way to solve a given problem if a precise value requires too much time or means (Booth and Siegler, 2006; Albarracín and Gorgorió, 2019).

Numerosity estimation, also known as quantity estimation, refers to estimating discrete quantities (Crites, 1992; Andrews et al., 2021). It "requires translating a nonnumerical quantitative representation (e.g., a visual representation of the approximate volume and density of candies in a jar) into a number" (Siegler and Booth, 2005, p. 198) without resorting to complete counting.

Previous research has shown that while children basically do not estimate adequately, performance improves with age (e.g., Siegler and Booth, 2005; Luwel and Verschaffel, 2008). For example, 6th graders estimate significantly better than 2nd and 4th graders, but no difference occurs between second and fourth graders (Luwel and Verschaffel, 2008). Research has also reported various strategies for estimating a quantity, such as simple arithmetic operations like addition or multiplication while de- and recomposing the quantity to estimate (Siegel et al., 1982; Crites, 1992; Luwel and Verschaffel, 2008). In this sense, a quantity structure needs to be recognized and utilized to estimate, for example by decomposing the quantity to estimate a subset that is followed by recomposing the quantity by using arithmetic operations. Thus, the skill to execute simple arithmetic operations can be essential to performing specific strategies in numerosity estimation. To date, there has been relatively little effort to link numerosity estimation with other constructs such as math achievement and math interest.

1.2. Math achievement and estimation

In general, estimation is one essential aspect of learning and understanding mathematics. Sriraman and Knott (2009) consider it a fundamental mathematical skill since estimation activities support the benefit and "development of mathematical concepts (and procedures) and cultivate critical thinking" (p. 210). Similarly, Siegler and Booth (2005) emphasize that estimation requires "going beyond rote application of procedures and applying mathematical knowledge in flexible ways. This type of adaptive problem-solving is a fundamental goal of contemporary mathematics instruction" (p. 197). Accordingly, primary school mathematics should aim to develop students who are flexible problem solvers possessing independent thinking strategies (Siegler and Booth, 2005; Schütte, 2008).

Research literature suggests that estimation accuracy is related to many aspects of mathematical skills (Siegler and Booth, 2005). For example, students who are gifted estimators show better arithmetic skills (Booth and Siegler, 2006) in counting (Barth et al., 2009; Bartelet et al., 2014), number sense (Crites, 1992; Sowder, 1992), and strategy flexibility (Siegler and Booth, 2005; Luwel and Verschaffel, 2008). Furthermore, fostering estimation abilities can greatly impact the development of these skills (e.g., Luwel et al., 2005; Siegler and Booth, 2005). Similarly, Dowker (2003) argues that the effect of estimation can be beneficial to arithmetic achievement because the process of estimation stimulates the development of awareness of number relations as well as resourcefulness with them.

Estimation is a fundamental part of learning mathematics as it plays a decisive role in the acquiring basic arithmetic skills (e.g., Siegler and Booth, 2005; Sayers et al., 2016). Moreover, there is consensus that while estimation is connected to number sense (Verschaffel et al., 2007), the latter is a construct that is difficult to define (Griffin, 2004). One attempt put forward by Sayers and Andrews (2015) conceives foundational number sense as a bundle of number-related essential competencies that require instruction to develop. In this context, estimation is one of eight identified components of foundational number sense (Sayers et al., 2016). Concurrently, foundational number sense is assumed to be fundamental for estimation processes (Siegler and Booth, 2004), as well as for understanding mathematics. The relation between number sense and math achievement is emphasized in literature and empirically supported (e.g., Aubrey and Godfrey, 2003; Aunio and Niemivirta, 2010). For instance, several research projects confirm simple arithmetic competence, another component of foundational number sense (Sayers et al., 2016), to be a strong predictor of later mathematical success (e.g., Malofeeva et al., 2004; Berch, 2005; Jordan et al., 2007; Ivrendi, 2011). As mentioned above, number sense can also be fundamental in performing numerosity estimation. Generally, it is widely believed that estimation is a determinant of later arithmetical achievement (Siegler and Booth, 2004; Sayers et al., 2016; Andrews et al., 2021).

The association between estimation and math achievement has been addressed in several studies (e.g., Siegler and Booth, 2004; Booth and Siegler, 2006; Schneider et al., 2009; Sasanguie et al., 2013; Wong et al., 2016) that have predominantly concentrated on number line estimation (Schneider et al., 2009; Sasanguie et al., 2013). Numerosity estimation has also been addressed by various studies that either address its relation to a single aspect of math achievement or examine a specific type of task related to math achievement. Results of such studies suggest that counting is fundamental to a successful process of solving numerical estimation tasks (Lipton and Spelke, 2005; Barth et al., 2009). Barth et al. (2009) report that the accuracy of children's estimates depends on their counting ability (Barth et al., 2009). Kindergarten students' efficiency in numerosity estimation explains, besides counting and comparing, a unique part of the variance in arithmetic achievement in first grade but no significant association between estimation and arithmetic achievement was found (Bartelet et al., 2014). Besides, according to their research, Wong et al. (2016) suggest that sole contributions of diverse estimation abilities affect arithmetic achievement for six-year-old children. Measures of numerosity estimation predict arithmetic achievement (Wong et al., 2016). Booth and Siegler (2006) report a positive correlation between judging numerosities and math achievement in second and third graders.

A substantial body of research focuses on the relation between numerosity estimation and single aspects of math achievement or between one specific type of task in numerosity estimation and math achievement. Although the literature often outlines a connection between numerosity estimation and math achievement, there is only limited empirical evidence of whether math achievement and accuracy in numerosity estimation regarding different types of perception tasks are related. Consequently, the relation between numerosity estimation and its impact on math as well as arithmetic achievement is not well understood.

1.3. Math interest and estimation

In recent decades, educational and psychological research has increasingly studied the influence of interest on learning and development in various educational settings (Krapp, 2002). In this vein, most researchers distinguish between individual/personal interest and situational interest (e.g., Krapp, 2002; Hidi and Renninger, 2006). Individual interest is a relation between a student and a specific content (e.g., Krapp, 2002; Hidi and Renninger, 2006) that is relatively stable and well-developed (e.g., Krapp, 2000; Hidi and Renninger, 2006; Jansen et al., 2016). Situational interest can be understood as a short-term affective engagement caused by external factors (e.g., Hidi, 1990; Krapp, 2002; Hidi and Renninger, 2006). Furthermore, interest is an affective, motivational construct linked to intrinsic motivation (e.g., Köller et al., 2001; Krapp, 2002, 2007; Renninger et al., 2002; Aunola et al., 2006).

Research findings and reviews suggest that interest contributes significantly to performance (e.g., Hidi and Harackiewicz, 2000; Krapp, 2002; Lee et al., 2014; Trautwein et al., 2015). For example, higher interest supports engagement and persistence in completing school tasks (e.g., Eccles and Wigfield, 2002; Renninger et al., 2002). In addition, an activity based on interest can include enjoyment and involvement, which can enhance attention, concentration, and positive affect, which in turn influence any learning process (Eccles and Wigfield, 2002; Hidi, 2006). This reported research also implicates an effect of math interest on numerosity estimation accuracy. It may be assumed that for instance higher attention and concentration during an estimation can also have an influence on estimation accuracy by capturing and observing the quantity to be estimated in a concentrated way. For example, structures within the quantity could be identified that can be used for the estimation process. Consequently, the estimation process influences the estimation accuracy. However, previous studies have not investigated the influence of non-cognitive processes like math interest on accuracy in estimation in general or numerosity estimation in particular. Although a broad body of studies concludes that non-cognitive processes have a differential impact on math achievement (e.g., Hattie, 2009; Lee and Shute, 2010; Jansen et al., 2016), no clear link has been established between numerosity estimation and math interest.

While the early school years are essential for learning mathematics and developing math interest, only a few studies of estimation have focused on preschool and primary school students (Aunola et al., 2006). In such studies, findings on the connection between math interest and math achievement are inconsistent. For example, Fisher et al. (2012) reported a positive relationship between math achievement and math interest in preschool students, and subsequently students' initial achievement levels predicted their later level of interest in math tasks. In the same way, early math interest predicted later math achievement. They suggest that this "reciprocal relationship […] has already begun by preschool" (Fisher et al., 2012, p. 679). In contrast, Gottfried (1990) found a significant relation between young children's intrinsic motivation and their math achievement, but their motivation did not predict standardized achievement test scores. One study of students from kindergarten through 12th grade reported that among non-cognitive constructs, motivation and interest are related to students' academic achievement (Lee and Shute, 2010). In contrast, however, Lee and Chen (2019) report that their cross-country study found in only 20% of the tested countries, interest had "a moderately strong predictive power for math achievement ($r \geq 0.224$) […] at the fourth grade" (p. 8). Across all countries, the average correlation between interest and math achievement was weak in the fourth grade (Lee and Chen, 2019), but it was consistently stronger in secondary school (Lee and Chen, 2019). Indeed, for that matter, the overall results from Aunola et al. (2006) and Viljaranta et al. (2009) showed a cumulative developmental cycle between children's mathematical and arithmetic performance and math-related task motivation in primary school (Aunola et al., 2006). A cumulative developmental cycle means the higher the performance, the more math-related task motivation later, which predicted further math performance (Aunola et al., 2006; Viljaranta et al., 2009). They "use the term task motivation to refer to children's interest value or intrinsic motivation" (Aunola et al., 2006, p. 23).

In summary, research focusing on the relationship between math interest and math achievement is ambivalent, but multiple findings suggest that interest contributes significantly to various aspects of a learning process that lead to better performance. If the assumption that numerosity estimation and math achievement are related is correct, then the relationship between achievement and interest raises the question of whether there is also a relationship between math interest and accuracy in numerosity estimation on the condition that students consider estimation as part of mathematics. Even though estimation is considered an essential aspect of learning and understanding mathematics, there is currently no research literature about the relationship between math interest or other non-cognitive processes and accuracy in numerosity estimation.

1.4. Aims and research questions

The current study examined the relations between accuracy in numerosity estimation, math achievement, and math interest. Particularly, this project investigated the relationship between math achievement and accuracy in numerosity estimation as well as the influence of accuracy in numerosity estimation on math achievement. In addition, the relationship between math interest and accuracy in numerosity estimation was analyzed. Accordingly, two main research questions and associated hypotheses were examined:

1. To what extent is there a relationship between math achievement and estimation accuracy for two- and three-dimensional numerosity estimation tasks in third grade?

In the literature, numerosity estimation and math achievement are assumed to be related. Previous research has shown that numerosity estimation influences arithmetic achievement (Wong et al., 2016). Booth and Siegler (2006) found that there is a relationship between one specific type of numerosity estimation task and math achievement in second and third grade. Therefore we expect:

> H_1: Math achievement and estimation accuracy for two- and three-dimensional numerosity estimation tasks are positively related.

Concerning our second focus:

2. To what extent is there a relationship between math interest in third-grade students and their accuracy in numerosity estimation tasks?

As already addressed in the first research question, a relation between numerosity estimation accuracy and math achievement is assumed and has been tentatively confirmed in terms of specific tasks or mathematical partial aspects (Booth and Siegler, 2006; Wong et al., 2016). A reciprocal relationship between children's math achievement and math interest is already supposed to start in preschool and a positive link has already been established (e.g., Fisher et al., 2012; Jansen et al., 2016). Since numerosity estimation is a mathematical activity, and provided that (1) math interest and math achievement as well as (2) math achievement and accuracy in numerosity estimation are positively related, we expect:

> H_2: Math interest and numerosity estimation accuracy for two- and three-dimensional numerosity estimation tasks are positively related.

2. Materials and methods

2.1. Participants and procedure

This project sampled 185 third-grade student volunteers who returned a parent-signed consent form. Additionally, we obtained teachers' and head teachers' consent. Student ages ranged from 8.7 years to 11.1 years ($M=9.5$, $SD=0.4$). Ninety-four students were girls (approximately 51%), and 91 were boys (approximately 49%). Students came from 13 classrooms in five German public schools, composed of two city schools, one suburban school, and two village schools. About 88% of the students were born in Germany, and about 74% used German as their first language.

Since we wanted to avoid counting and did not focus on quasi-simultaneous acquisition, we chose a number range larger than 30. In the second grade in Germany, the number range is extended from 20 to 100. By choosing the third grade, we could be sure that the number range had already been extended to 100 and that a student's lack of number range development could not be the main reason for an inappropriate estimate. Tasks in the number range 100–150 were included to create challenging tasks and to be able to better differentiate estimation accuracy.

To answer the research questions, we designed a quantitative study and collected data using a digital standardized test to determine estimation accuracy, a digital questionnaire to measure non-cognitive interest in mathematics, and a standardized test of math achievement. The data collection took place during school hours from May 2022 to July 2022. The digital numerosity estimation test and digital questionnaires were administered in one lesson (45 min) per class on the same day. We conducted the standardized paper-pencil-test (45 min) to measure math achievement on another day.

2.2. Measures

2.2.1. Numerosity estimation accuracy

To our knowledge, no published instrument reliably measures accuracy in numerosity estimation in a standardized way. For this reason, we developed an online numerosity estimation test for primary school students.

Numerosity estimation means estimating the number of structured or unstructured elements in a bounded collection without exact counting. Therefore, 21 estimation tasks were "presented in such a way as to preclude exact counting of the items" (Hogan and Brenzinski, 2003, p. 260). To avoid counting as an entire strategy, each task (i.e., picture of a number of elements) was presented for 20 s (Luwel and Verschaffel, 2008). The quantities in the number range of 31–144 also hindered counting all elements (Albarracín and Gorgorió, 2019). Such a range of quantities is realistic to estimate for third-grade students because the number ranges have already been introduced. In addition, this test did not include tasks requiring knowledge external to estimation, so this would not affect estimation accuracy. After the picture disappeared, students had 40 s more to adjust their results with a slider. Independently from the items, they could adjust their estimates between zero and 500 on the slider scale. We chose this uniform response format for all items so that the estimates would not be affected by different number ranges of the slider scale and to counteract errors that could occur when using a keyboard to type in the result (e.g., numbers are not typed in according to place value). To adjust the slider, only the computer mouse was needed. The handling was comparatively intuitive. Under the slider, the selected number was visible during the adjustment of the slider and finally, the result could be read. Altogether, students had 1 min to estimate each of 21 items.

To ensure content validity, the test items were constructed to represent a broad range of numerosity estimation tasks. Items featured a range of task characteristics: dimension (two- or three-dimensional), the arrangement of the elements (structured or unstructured), and the nature of the elements (equal or unequal regarding size, shape, and color). Some items were created based on the items Luwel and Verschaffel (2008) used in their study. The images of two-dimensional numbers have shown, for example, iconic cars, coins or animals. Figure 1 shows an example item of a two-dimensional quantity with unstructured arrangement and unequal items (the cars have different sizes, shapes, and colors).

Finally, it is essential to regard that on a screen, it was only possible to see a representation of a three-dimensional quantity, which means that students must perceive the number of elements as three-dimensional in the first step of the estimation. That is, they must realize that some elements were behind those in front and could not be seen. The pictures of the three-dimensional numbers have shown, for example, real building bricks, plug cubes, or

FIGURE 1
Example item: two-dimensional, unstructured, unequal items.

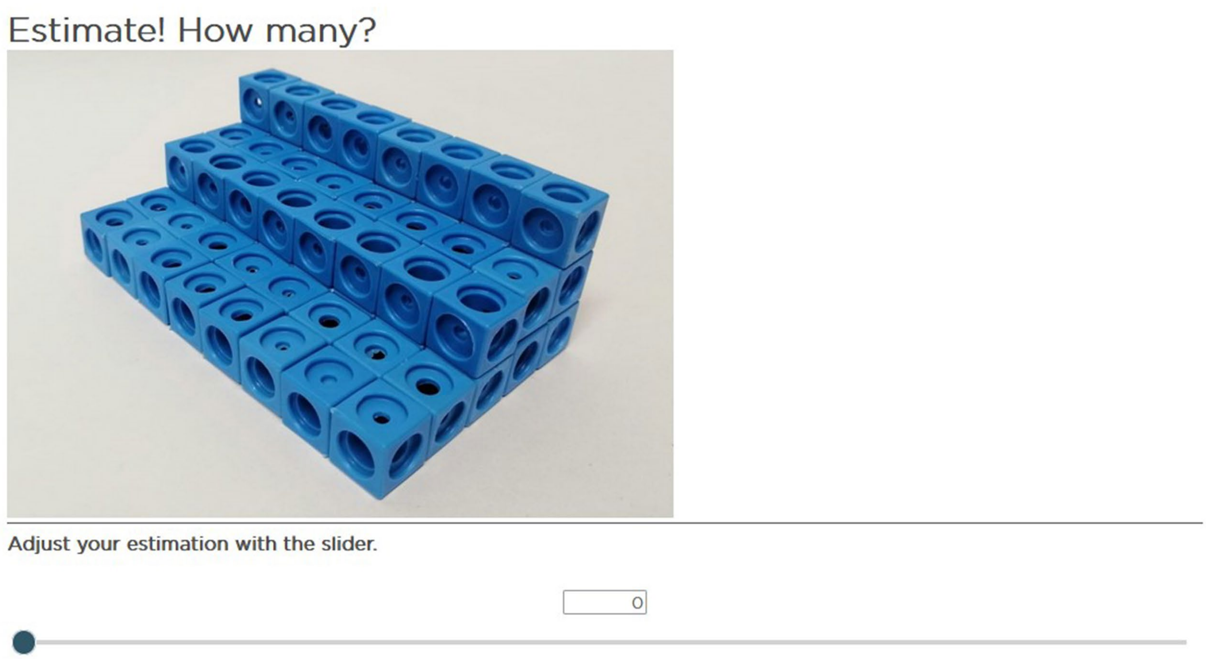


FIGURE 2
Example item: three-dimensional, structured, equal elements.

wooden cubes. Figure 2 displays an example item of a three-dimensional quantity with structured arrangement and equal elements.

Before the test started, we introduced the term "estimation" and gave instructions on the test procedure and the response format of the items. In a pretest, students worked on two test items and had the opportunity to ask questions about the test procedure. After this introduction phase, the actual test started.

The value for any item in the numerosity estimation test reflected a standardized absolute deviation, calculated as follows:

$$\frac{Actual\ value - Estimated\ value}{Standard\ deviation\begin{pmatrix} in\ relation\ to\ acutal\ value \\ instead\ of\ \ mean\ value \end{pmatrix}}.$$

We used this standardization because as the number of elements in an item increases, so too does the standard deviation of estimates. Since the number of elements to be estimated influences the mean value, we calculated the standard deviation concerning the actual value. As relative errors occur more frequently with small numbers than large ones, the raw values or the absolute deviation would not be comparable due to the different number ranges of the numbers to be estimated. Standardizing all values in this way made response estimates comparable for data analysis.

In relation to testing evaluation, no values were excluded except 0 and 500, which correspond to the minimum and maximum values on the slider. Hence, we assumed these values are not valid. Using the study sample, we measured the instrument's construct validity with exploratory factor analysis (principal axes factor analysis; Brandt, 2020; Moosbrugger and Kelava, 2020). Four factors resulted. We eliminated three of the 21 items due to similar double loading or content fit. Thus, at least four items form a construct and the items were not strongly reduced with respect to content validity (Brandt, 2020). All items in one factor were tested again within a principal component analysis to ensure only one factor was within. Consequently, the reliability of one scale of items of one factor was measured by Cronbach's Alpha. In addition, we examined the corrected item-total-correlation in each subscale of the numerosity estimation test. Item-total-correlations between 0.4 and 0.7 are considered good (Kelava and Moosbrugger, 2020). Within the scales *SmallN*, *Mix* and *3DlargeN* the corrected item-total-correlation of each item was sufficiently large (≥0.4). For the scale *2DlargeN*, the corrected item-total-correlation of three items was between 0.47 and 0.55. One item had a corrected item-total-correlation of 0.37. Since this value was close to a good value, the item is theoretically significant for the characteristic of interest, and the reliability does not change significantly by removing this item, the item was not removed. In summary, the exploratory factor analysis for construct validation revealed four dimensions and in line with the item-total-correlations one-dimensionality within these four dimensions (Moosbrugger and Kelava, 2020). Good item-total-correlations provide indications that the items measure the same characteristic in terms of content (Kelava and Moosbrugger, 2020).

Subsequently, we computed factor scores as means over all values of items measuring one similar factor because the factor loadings per factor did not differ too much. The means of the four scales represent a standardized mean of deviation from the numerosity to be estimated: the smaller the deviation, the higher the estimation accuracy. Table 1 shows the resulting four scales incorporating the factor scores as measures for accuracy in numerosity estimation.

The first scale *SmallN*, with six items, represents items in the number range from 31 to 45 and had a reliability of α = 0.78. Scale *2DlargeN* shows a reliability of α = 0.69 and includes four items in the number range from 82 to 132. All items were two-dimensional in this scale. *2DlargeN* was the only scale that has questionable reliability. However, the value is very close to a satisfactory value, and the items in this scale were meaningful in terms of content, which is why this scale was included in further analyses. The third scale, *Mix*, also comprised four items with a reliability of α = 0.75. Two items in *Mix* were in the numerical range of around 70, one contained 123 elements to estimate, and another showed 42 elements that were arranged three-dimensionally and unstructured. In this scale, two items were arranged in an unstructured way, and two were arranged in a structured way. Finally, scale *3DlargeN*, with four items and reliability of α = 0.71, represents items belonging to the number range from 72 to 144. All were three-dimensional items. In summary, the estimation accuracy measure was comprised of four scales, each with alpha 0.69–0.78, and consisting of four to six items.

2.2.2. Math achievement

For assessing math achievement, we administered a timed, standardized test for third graders (May and Bennöhr, 2021). The test was presented in a booklet and consisted of 74 items organized into two subtests. The first subtest, with 56 items, focused on basic knowledge of arithmetic and computational abilities. For this reason, we refer to this subtest as arithmetic achievement. The following contents were in the arithmetic subtest:

– Subtraction and addition [e.g., the following tasks are to be solved: "48 − 34 + 35" (May and Bennöhr, 2021, p. 4) or "Which number is 450 greater than 240?" (May and Bennöhr, 2021, p. 12)].
– Understanding of place value in combination with subtraction and addition [e.g., "You have four hundreds and take away 12 tens. How many whole hundreds do you have left?" (May and Bennöhr, 2021, p. 6)].
– Completing number series [e.g., this series of "50, 46, 42, 38" is to be supplemented by two further numbers (May and Bennöhr, 2021, p. 8)].

TABLE 1 Estimation scales.

Scale	Items	Number range <50/50–100/>100	Dimension 2D/3D	Arrangement structured/unstructured	Cronbach's alpha (α)
SmallN	6	6/0/0	4/2	3/3	0.78
2DlargeN	4	0/1/3	4/0	3/1	0.69
Mix	4	1/2/1	2/2	1/3	0.75
3DlargeN	4	0/1/3	0/4	2/2	0.71

– Acquisition of the number of small cubes in a three-dimensional figure (that exists of small cubes).
– Solving an arithmetic square and arithmetic triangle.
– Inserting appropriate arithmetic symbols.
– Numberline estimation.

The second subtest comprises 18 application-oriented tasks (application task achievement), such as:

– Solving word problems [e.g., students solve: "24 children come to the sport. Six children are in one group. How many groups are there?" (May and Bennöhr, 2021, p. 18)].
– Extracting information from a table.
– Reading a temperature scale.
– Number registration of small cubes needed to make at the end a big cube with the structure of 3 × 3 × 3 small cubes (12 small cubes are already rudimentarily represented in the structure of the bigger cube).
– Reading the weight displayed on a scale.

Including the introduction and instruction, the test lasted about 45 min, corresponding to a standard lesson in German classrooms. The test was conducted in a classroom with all students in one class.

The test was evaluated in such a way that one point was given for each correctly answered item. It was therefore possible to achieve a total of 74 points if all items of the test were answered correctly. Consequently, a sum score was calculated for each student from the correctly answered items. The reliability of the whole test was $\alpha = 0.93$ (May and Bennöhr, 2018), with $\alpha = 0.91$ for the first subtest (arithmetic achievement) and $\alpha = 0.81$ for the second subtest (application task achievement). Therefore, the reliability of the test can be considered excellent (Döring and Bortz, 2016).

2.2.3. Math interest

For measuring math interest, we used two published questionnaires from large-scale comparison studies. First, the scale "Interesse an Mathematik" (interest in mathematics) was derived from the IGLU 2001 questionnaire survey (Bos et al., 2005), and the second came from the national questionnaire of the PISA 2012 "INTMAT—Freude und Interesse an Mathematik" (interest in and enjoyment of mathematics; Mang et al., 2018). The scale "interest in mathematics" with five items, slated initially for students at the end of fourth grade, showed a reliability of $\alpha = 0.75$ (Bos et al., 2005). In our sample, the reliability of this instrument was $\alpha = 0.76$. The five items test math interest in different areas. For example, the children must answer the item "I find it exciting to discover rules or tricks in mathematics myself." The other scale "interest in and enjoyment of mathematics" (INTMAT) with four items showed a reliability of $\alpha = 0.89$ for 15-year-old students (Mang et al., 2018). These items were linguistically appropriate for third-grade students and we obtained a reliability of $\alpha = 0.85$ with our sample. "I look forward to my math lessons" is, for example, one of the four items.

We adapted the response options from the PISA 2012 study to the options of IGLU 2001 and used a four-point Likert format for students' responses within both scales. The response options were "not correct" (coded as 1), "partly correct" (coded as 2), "almost correct" (coded as 3), and "correct" (coded as 4). Generally, we read every sentence out loud to the class before students answered the item.

Since both scales showed an almost identical mean value and median (Interest in mathematics: $M = 3.07$, $Mdn = 3.20$; INTMAT: $M = 3.04$, $Mdn = 3.25$) as well as an identical standard deviation ($SD = 0.79$), we decided to merge the two scales together for further analyses. The reliability of all items of both scales is $\alpha = 0.88$ (9 items). Further analyses were calculated using the mean of the responses of the nine items and refer to this derived scale as "math interest."

2.3. Statistical analysis

For statistical analysis, we used the sum scores of the obtained points in the math achievement test, the mean of the two interest scales, and the mean of each of the four scales of the numerosity estimation test.

The analyses were calculated using SPSS Statistics. To examine the relationship between accuracy in numerosity estimation, math achievement, and math interest in third grade, descriptive and correlational analyses were used. In addition, we performed a multiple linear regression analysis to examine the predictive capacity of math interest, the four scales of accuracy in numerosity estimation, and gender (male coded 0, female coded 1) on math achievement as the criterion variable. Since research has repeatedly shown that gender can also have an impact on mathematical competencies (Else-Quest et al., 2010; Keller et al., 2022), gender was included as independent variable. In German primary schools, boys often show slightly higher competencies than girls (Stanat et al., 2022).

3. Results

3.1. Descriptive results

Descriptive statistics for each of the four numerosity estimation scales are presented in Table 2, which shows the mean and standard deviation for each scale.

Recall that the mean for estimation accuracy derived from the standardized absolute deviation from the actual value (see section 2.2.1). Because of the standardization, these values are comparable and represent numerical/value proximity; the smaller the mean, the better the estimation accuracy. Thus, in scales *2DlargeN* and *3DlargeN*, students' estimates deviated most from the actual value in comparison to the other two scales. On the one hand, *2DlargeN* contains items that

TABLE 2 Descriptive statistics of estimation scales.

	M	*SD*
SmallN	0.59	0.56
2DlargeN	0.87	0.37
Mix	0.71	0.53
3DlargeN	0.85	0.40
SmallN_Pd	36.83%	41.47%
2DlargeN_Pd	40.13%	17.05%
Mix_Pd	38.67%	30.29%
3DlargeN_Pd	41.88%	19.40%

Pd, Percentage deviation from the actual value.

represent two-dimensional quantities in the number range of 82–132, and on the other hand, *3DlargeN* includes items that represent three-dimensional quantities in a slightly larger number range of 72–144. The mean was almost identical for these two scales. According to Table 2, the children's estimates deviated least from the actual values in *SmallN*, which reflects the number range 31–45. Since the standardized values of estimation accuracy are difficult to interpret, the mean and standard deviation of the percentage deviation from the actual value are also displayed in Table 2.

Table 3 displays students' mean sum scores in the math achievement test, the means of both subscales, and the mean for math interest. The mean sum score corresponds to the average number of correct items. For the sample, about 63% of the tasks of the math achievement test were solved correctly. In the arithmetic achievement subtest, students solved an average of 67% of the tasks correctly. In comparison, application tasks including word problems proved more difficult, with only 52% of these tasks solved correctly. Table 3 also presents the mean and standard deviation for math interest. The maximum value for math interest would be 4, and 3.05 is the mean value in our sample.

3.2. Relationship between numerosity estimation accuracy and math achievement

Table 4 displays results from our correlation analysis. Although there were a few negative weak correlations, they were not statistically significant. A negative relationship means that the higher the math achievement, the lower the standardized absolute deviation (i.e., close proximity). Regarding math achievement in total, there was only a weak negative correlation with scale *SmallN* ($r = -0.13$, $p > 0.05$). Furthermore, this scale also correlated weakly and negatively with both subscales of the math test. Focusing on the subscales of the math achievement test, there was also a weak correlation between scale *Mix* and achievement in application tasks ($r = -0.11$, $p > 0.05$), but no correlation with arithmetic achievement or math achievement in general. It is noticeable that all Pearson coefficients were negative except between scale *2DlargeN* and arithmetic achievement ($r = 0.02$, $p > 0.05$) as well as *3DlargeN* and arithmetic achievement ($r = 0.03$, $p > 0.05$).

We conducted a multiple linear regression analysis, with math achievement as the dependent variable, to investigate how the residuals of estimation accuracy, math interest and gender predict math achievement. To test the model's assumptions, high multicollinearity had to be ruled out (Urban and Mayerl, 2011). For this reason, the correlations between the independent variables and the variance inflation factor were calculated. The results show no high correlation between the scales of estimation accuracy and no high correlation between these scales and math interest (Table 4). In addition to the results of the correlation analysis, the resulting variance inflation factors between 1.01 and 1.40 confirm that there is no multicollinearity for the independent variables mathematics, gender, *SmallN*, *2DlargeN*, *Mix, and 3DlargeN*. Accordingly, this requirement for regression analysis was fulfilled (Urban and Mayerl, 2011). We also tested whether the residuals are normally distributed and checked for homoscedasticity using the Breusch-Pagan test. The normal distribution of the residuals and homoscedasticity were confirmed. Table 5 shows the results of the multiple linear regression analysis.

The results showed that the overall regression model explains 6% of total variances in math achievement, $F(6, 178) = 2.87$, $p < 0.05$, and revealed that accuracy in numerosity estimation does not significantly predict math achievement.

TABLE 3 Descriptive statistics of math achievement and math interest.

	Items	*M*	*SD*
Math achievement	74	46.75	11.34
Arithmetic achievement	56	37.62	8.62
Application task achievement	18	9.32	3.62
Math interest	9	3.05	0.69

3.3. Relationship between numerosity estimation accuracy and math interest

To test our second hypothesis, we firstly analyzed the relationship between math achievement and math interest of third-grade students using Pearson's *r*. We examined math interest and math achievement with both subscales, arithmetic achievement, and application task achievement (see Table 4).

TABLE 4 Correlations between numerosity estimation accuracy, math achievement and math interest.

	1	2	3	4	5	6	7	8
1. SmallN	--							
2. 2DlargeN	0.08	--						
3. Mix	0.49**	0.10	--					
4. 3DlargeN	0.33**	0.41**	0.27**	--				
5. Math achievement	−0.13	−0.01	−0.09	−0.00	--			
6. Arithmetic achievement	−0.11	0.02	−0.09	0.03	0.94**	--		
7. Application task achievement	−0.14	−0.03	−0.11	−0.08	0.78**	0.64**	--	
8. Math interest	−0.05	−0.04	−0.17*	0.03	0.21**	0.19**	0.21**	--

*$p < 0.05$. **$p < 0.01$.

TABLE 5 Regression model: contribution to math achievement.

Variable	Unstandardized *b*	Standardized *β*	Standard error	*t*
Constant	39.82***		4.62	8.62
Math interest	3.20**	0.20**	1.20	2.67
Gender	−3.85*	−0.17*	1.63	−2.37
SmallN	−2.80	−0.14	1.72	−1.63
2DlargeN	−0.231	−0.01	2.40	−0.10
Mix	−0.12	−0.01	1.79	−0.07
3DlargeN	1.22	0.04	2.39	−1.63
R^2	0.09			
Adjusted R^2	0.06			
$F(df=6, 178)$	2.87*			

*$p<0.05$; **$p<0.01$; ***$p<0.001$.

The results show a significant positive correlation between math achievement and math interest ($r=0.21$, $p<0.01$). There was a similar significant correlation between math interest and both subscales of math achievement. The regression analysis with math achievement as the criterion variable showed that math interest [$\beta=0.20$, $t(178)=2.67$, $p<0.01$] was a significant predictor of math achievement for third graders.

For testing our second hypothesis (H_2), we analyzed the relationship between estimation accuracy and math interest. The results of this Pearson correlation analysis are presented in Table 4. There was no correlation between math interest and the scales *SmallN*, *2DlargeN,* and *3DlargeN*. However, there was a weak negative, significant correlation between the estimation scale *Mix* that represents different two- and three-dimensional estimation tasks and math interest ($r=-0.17$, $p<0.05$). A negative relationship, in this case, means that the higher the math interest, the lower the standardized absolute deviation, thus the better the estimation accuracy.

4. Discussion

This paper analyzed the relationship between math achievement, math interest, and accuracy in numerosity estimation within different types of tasks for third graders. We targeted the field of numerosity estimation, which has been shown to be important for the development of math skills (Crites, 1992; Sowder, 1992; Booth and Siegler, 2006; Luwel and Verschaffel, 2008). Our discussion relates our findings to the current research literature on developing estimating competence in students.

4.1. Accuracy in numerosity estimation, math achievement, and math interest

We analyzed accuracy in numerosity estimation, operationalized as the standardized deviation from the actual value or quantity to be estimated. As we expected, students' estimation accuracy was highest in *SmallN* in relation to the other scales of estimation. This scale contained the smallest numbers to be estimated in comparison to the other estimation scales. The standard deviation within this scale was comparatively high. This may be because no estimates were excluded. In comparison, estimates of *2DlargeN* and *3DlargeN* were less accurate. One possible explanation for this pattern could be the increased number range covered by these two scales, because they are very similar and relatively large, representing three quantities in the number range > 100 and one quantity in the number range between 50 and 100. It is interesting to note that accuracy in numerosity estimation was relatively similar for both scales, although the quantities in scale *2DlargeN* were displayed in two-dimensional representations, whereas scale *3DlargeN* included three-dimensional representations. It would be reasonable to assume that third-grade students were more familiar with two-dimensional mappings (e.g., a hundred field) and thus may have developed more benchmarks in two-dimensional mappings. In addition, it could be argued that third graders have already encountered two-dimensional "bundles" of elements, which enabled fast coding or spontaneous structuring of two-dimensional quantities. However, while such structuring could have led to a closer proximity to (lower deviation from) the actual value, and contrary to our expectations, that was not the case.

Many qualitative studies of numerosity estimation have analyzed estimation accuracy as the percentage deviation from the real value (Siegel et al., 1982; Crites, 1992; Luwel and Verschaffel, 2008). To compare our results on accuracy with other studies, we reported the mean values of the percentage deviations. However, we assumed that the number of elements to be estimated has a crucial influence on the estimate and, accordingly, the adequacy of an estimate should not be measured only by the percentage deviation. To correct for this tendency, we used the standardized absolute deviation. In the relevant literature, an adequate estimate has often been assessed using different limits or ranges for the percentage deviation from the actual value. For example, Luwel and Verschaffel (2008) defined a limit of up to 10% as a very good estimation and the range between 10% and 25% deviation as a good estimate. Older studies have defined a limit of up to 50% as an adequate estimate (Siegel et al., 1982; Crites, 1992). Regarding the mean of the percentage deviation from the actual value for the four estimation scales in our study, we reasoned that most students did not estimate adequately, according to the less liberal 25% guideline of Luwel and Verschaffel (2008). At the same time, our data agreed with Luwell and Verschaffel's conclusion that primary school children have trouble making adequate estimations. This is also in line with Siegler and Booth (2004) who have argued that low estimation skills in second and fourth graders suggests "a lack of number sense and conceptual structures" (p. 429). Altogether, our students' low estimation accuracy was not an unexpected result, perhaps because estimation is not a very prominent topic in instruction (Luwel and Verschaffel, 2008; Sayers et al., 2020; Andrews et al., 2021). In our opinion, estimation should be considered a learning object, with the ability to adequately estimate addressed through instruction, so that students learn to estimate reflectively, according to estimation task characteristics (Wessolowski, 2014).

Regarding the results of the math achievement test, our third graders performed better on the arithmetic achievement subtest than on the subtest involving application tasks. This is not surprising, because arithmetic is a highly focused subject of instruction (National Council of Teachers of Mathematics, 2000). In the descriptive results, the mean value of math interest was 3.05 on a scale with maximum value of 4.0. The results of the IGLU 2001 study, from which the math interest questionnaire originated, also showed a high mean level of mathematics interest for the end of fourth grade (Walther et al., 2003). In general, primary school students often exhibit relatively high interest (Helmke, 1993), which we can confirm with our data.

4.2. Relationship between accuracy in numerosity estimation and math achievement

Siegler and Booth (2005) have described a solid connection between competence in estimation and mathematical procedures. In this vein, researchers often assume that estimation ability is a determinant of later math performance, particularly arithmetic achievement (Siegler and Booth, 2005; Sayers et al., 2016; Andrews et al., 2021). Few quantitative studies exist that have investigated the relationship between math achievement and numerosity estimation (Booth and Siegler, 2006; Bartelet et al., 2014; Wong et al., 2016). In contrast with those studies, we investigated this relationship by testing numerosity estimation accuracy of third-grade students across different types of perception tasks: either two- or three-dimensional, utilizing different number ranges, with structured or unstructured elements.

In relation to the research questions, our results showed no significant relationship between math achievement of third-grade students and their accuracy in numerosity estimation. It is noteworthy that all but two of the correlation coefficients were negative, since a negative correlation means that the lower the deviation from the actual value, the higher the estimation accuracy. From these results, we conclude that estimation skills do not develop automatically, even in students with high math achievement. Accordingly, those who performed very well on the math achievement test did not automatically achieve a low deviation from the actual value. Thus, we cannot confirm the results from Booth and Siegler (2006) who report a positive correlation between judging numerosities and math achievement in third graders. In their study, they invite the students to estimate which number, out of two, corresponds to the number of candies shown in a container. The different results could therefore be due to the different measures for numerosity estimation and math achievement. Finally, the multiple linear regression analysis revealed that none of the four scales measuring accuracy in numerosity estimation significantly predicted math achievement. Based on these results, we cannot confirm the theoretical assumption that estimation accuracy determines math achievement. Such a conclusion is tentative, however, because we examined the relationship with only third-grade students, using a very particular methodology.

In general, our pattern of results is surprising. After all, a relationship between estimation accuracy when estimating smaller quantities (<50) and math achievement was expected. Structuring the elements, decomposing them, and then recomposing them using simple arithmetic competencies would have been one possible strategy for estimating these numbers (Siegel et al., 1982; Crites, 1992; Luwel and Verschaffel, 2008). We would also have expected a connection with two-dimensional presentation of numbers since those representations are certainly used in math lessons to visualize quantities (e.g., a field of hundreds as a square field of points), and structuring would have been conceivable. Because both estimation and simple arithmetic competence are components of the foundational number sense (Sayers et al., 2016), which in turn is linked to math achievement (e.g., Aubrey and Godfrey, 2003; Aunio and Niemivirta, 2010), we had not expected these results and cannot verify our hypothesis (H_1).

The characteristics of the math achievement test may be one explanation for the fact that the estimation accuracy was not correlated to math achievement in our study. In the math achievement test, flexible approaches and the understanding of the non-symbolic magnitude was hardly addressed. There was only one task in which a non-symbolic magnitude was to capture. But, for solving an estimation task successfully, students "need to have an approximate understanding of the non-symbolic magnitude expressed by the symbolic magnitude" (Bartelet et al., 2014, p. 15). The math achievement tasks focused on precisely applying specific mathematical procedures. In this regard, it is important to consider that estimation is not linked to any specific mathematical procedure but requires flexible approaches and flexible use of already acquired knowledge (Siegler and Booth, 2005; Schütte, 2008). Furthermore, counting is fundamental for numerosity estimation (Barth et al., 2009) and an important mathematical skill (Luwel et al., 2005; Siegler and Booth, 2005; Luwel and Verschaffel, 2008). However, due to the structure of the math achievement test for the third grade, counting skills were not queried. Apart from this, the analyses should be performed again with a larger sample to verify these results. Finally, high ceiling could have influenced student performance on the *2DlargeN* and *3DlargeN* scales. These could have been caused by the items being too difficult for the students, which should be considered before the test is administered again.

4.3. Relationship between accuracy in numerosity estimation and math interest

Previous findings suggest that interest as a non-cognitive aspect contributes significantly to various aspects of a learning process that lead to better learning results (e.g., Hidi and Harackiewicz, 2000; Krapp, 2002; Lee et al., 2014; Trautwein et al., 2015). But previous research and resulting conclusions about the connection between math interest and math achievement have been inconsistent. Overall, most studies have focused on students in higher grades. In that context, there exists a possible connection between estimation and general math ability (Siegler and Booth, 2005) that can be influenced by math interest (e.g., Aunola et al., 2006; Fisher et al., 2012). For this reason, we investigated the relationship between math interest and math achievement first and then subsequently the relationship between math interest and accuracy in numerosity estimation.

We found a weak, positive, significant relationship between math achievement and math interest for students in third grade. The results of our correlational analysis accord well with Lee and Chen (2019) who found a weak correlation between math achievement and math interest in fourth grade. It seems to be important to promote the interest of students, because there can be a stronger relationship later between interest and math achievement in secondary school (Lee and Chen, 2019). Furthermore, the multiple linear regression analysis showed that math interest predicts math achievement on a significant level.

Previous research findings implicate that math interest can affect accuracy in numerosity estimation on the condition that students consider estimation as part of mathematics. To date, there has been no study published that examines the connection between math interest and accuracy in numerosity estimation. In this context, we have attempted to contribute some knowledge to this field by specifically examining such a relationship.

We cannot give an unqualified answer to the extent of a relationship between math interest in third-grade students and accuracy in numerosity estimation tasks because our results were inconsistent. For example, we found a significant but weak, negative correlation between the scale *Mix* and math interest. However, the other three scales of accuracy in numerosity estimation were not related to math interest. We had expected either that all scales would correlate with math interest (H_2) or else that none would if the students did not consider numerosity estimation as mathematical content. So, the pattern of correlations did not support our expectations. In this context, it would be interesting to analyze correlation patterns with a larger sample across more grades, because this result could also be spurious.

In general, we assume that children do not necessarily understand estimation as classical mathematics and, consequently, we cannot conclude a necessary relationship between math interest and accuracy in numerosity estimation. When children were asked about their math interests, they spontaneously appealed to their own experiences in mathematics class.

> "Our view of studies on estimation has reviewed how difficult it is for students who received traditional instruction to understand that besides counting precisely and calculating exact answers, there is also something like estimating and developing appropriate procedures and strategies for making appropriate estimate" (Verschaffel et al., 2007, p. 581).

This quote from Verschaffel et al. (2007) supports our stance and illustrates how children associate mathematics with the world of exact numbers until they are shown otherwise, for example in math lessons. Estimation can be seen as a contrast to the world of exact numbers (Schipper, 2009). Therefore, children do not typically associate estimation with mathematics, because estimating requires recalling activities that may never have been experienced in the first place. In this context, the finding that math interest and estimation accuracy are rarely related is not astonishing.

4.4. Conclusion

Our results provide new insights into numerosity estimation and its relationship with math achievement and math interest. Numerosity estimation is a neglected area in mathematics education (Andrews et al., 2021), but one that holds high potential for mathematics education in primary school.

The results of this study may have been affected by the specific content of the math achievement test. Specifically, the test's emphasis on algorithmic calculation may be seen as one limitation of the study. As we mentioned before, estimation requires independent thinking strategies and the skill to solve a problem in a flexible way. These aspects were not explicitly part of the achievement test we used. From a theoretical perspective, we assumed that numerosity estimation influences the development of number knowledge, which is essential for diverse mathematical proficiency (Wessolowski, 2014). Number knowledge includes having a basic idea of numbers that was also not directly addressed in the achievement test. Therefore, it would be interesting to analyze more specific relations between numerosity estimation accuracy and less procedure-based aspects of math achievement. In addition to the analytical methods used, structural equation models are conceivable for future studies examining such relationships.

We did not find a significant relationship between math achievement and estimation accuracy at this point in third grade, but, in our opinion, it would be worthwhile to further investigate this relationship. This study did not aim to answer the question about the influence of fostering numerosity estimation on math achievement over time or if there is a cumulative developmental cycle. So, future research could focus on this influence in longitudinal or intervention studies. Furthermore, we recommend the consideration of two- and three-dimensional numerosity estimation tasks as well other characteristics to focus on commonalities and differences that may influence (components of) math achievement. Since previous research has found that age has an impact on estimation accuracy, it would be equally reasonable to test this relationship at a higher-grade level (e.g., Siegler and Booth, 2005; Luwel and Verschaffel, 2008). The numerosity estimation test we developed could be very useful for longitudinal studies, since it is possible to use the test multiple times. This test is still in the early stages of development and will be further developed and evaluated. One reasonable focus would be the continuing validation of the test and its components.

Our study assessed general math interest. We could not capture situational interest regarding the estimation tasks or interest in estimation. Consequently, modifying or adapting the items on a scale to specify estimation interest would seem to be an important line of further research. It would be worthwhile to investigate, for example, whether estimation interest or situational interest in solving estimation tasks affects the estimation process and thus estimation accuracy and whether such an affect is stable across age groups.

Due to the relevance of estimation for number sense (Sayers et al., 2016), arithmetic achievement (Booth and Siegler, 2006; Wong et al., 2016), and number knowledge (Wessolowski, 2014), it is important to understand exactly how numerosity estimation is connected to math achievement in general, to specific competencies in detail, and to various non-cognitive components. This study provided new insights into this field, and at the same time, it underscores the importance of paying further attention to numerosity estimation in research and fostering estimation abilities in mathematics education.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Hessian Ministry of Education. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

LB and ER-S contributed to the conception and design of the study. LB organized the database, performed the statistical analysis, and wrote the first draft of the manuscript. ER-S wrote sections of the manuscript and supervised the writing process. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Albarracín, L., and Gorgorió, N. (2019). Using large number estimation problems in primary education classrooms to introduce mathematical modelling. *Int J Innov Sci Mathematics Educ.* 27, 45–57. doi: 10.30722/IJISME.27.02.004

Andrews, P., Xenofontos, C., and Sayers, J. (2021). Estimation in the primary mathematics curricula of the United Kingdom: ambivalent expectations of an essential competence. *Int. J. Math. Educ. Sci. Technol.* 53, 2199–2225. doi: 10.1080/0020739X.2020.1868591

Aubrey, C., and Godfrey, R. (2003). The development of children's early numeracy through key stage 1. *Br. Educ. Res. J.* 29, 821–840. doi: 10.1080/0141192032000137321

Aunio, P., and Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learn. Individ. Differ.* 20, 427–435. doi: 10.1016/j.lindif.2010.06.003

Aunola, K., Leskinen, E., and Nurmi, J. (2006). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *Br. J. Educ. Psychol.* 76, 21–40. doi: 10.1348/000709905X51608

Bartelet, D., Vaessen, A., Blomert, L., and Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J. Exp. Child Psychol.* 117, 12–28. doi: 10.1016/j.jecp.2013.08.010

Barth, H., Starr, A., and Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cogn. Dev.* 24, 248–264. doi: 10.1016/j.cogdev.2009.04.001

Berch, D. B. (2005). Making sense of number sense: implications for children with mathematical disabilities. *J. Learn. Disabil.* 38, 333–339. doi: 10.1177/00222194050380040901

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, A., et al. (2005). IGLU. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente. Münster: Waxmann.

Brandt, H. (2020). "Exploratorische Faktorenanalyse" in *Testtheorie und Fragebogenkonstruktion*. eds. H. Moosbrugger and A. Kelava. *3rd* ed (Berlin: Springer), 13–38.

Crites, T. (1992). Skilled and less skilled estimators' strategies for estimating discrete quantities. *Elem. Sch. J.* 92, 601–619. doi: 10.1086/461709

Döring, N., and Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Dowker, A. (2003). "Young children's estimates for addition: the zone of partial knowledge and understanding" in *The development of arithmetic concepts and skills: Constructing adaptive expertise*. eds. A. Baroody and A. Dowker (Mahwah, NJ: Erlbaum), 243–265.

Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annu. Rev. Psychol.* 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053

Fisher, P. H., Dobbs-Oates, J., Doctoroff, G. L., and Arnold, D. H. (2012). Early math interest and the development of math skills. *J. Educ. Psychol.* 104, 673–681. doi: 10.1037/a0027756

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *J. Educ. Psychol.* 82, 525–538. doi: 10.1037/0022-0663.82.3.525

Griffin, S. (2004). Building number sense with number worlds: a mathematics program for young children. *Early Child Res. Q.* 19, 173–180. doi: 10.1016/j.ecresq.2004.01.012

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.

Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. *Zeitschrift für Pädagogische Psychologie.* 7, 77–86.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Rev. Educ. Res.* 60, 549–571. doi: 10.2307/1170506

Hidi, S. (2006). Interest: A unique motivational variable. *Educ. Res. Rev.* 1, 69–82. doi: 10.1016/j.edurev.2006.09.001

Hidi, S., and Harackiewicz, J. M. (2000). Motivating the academically unmotivated: a critical issue for the 21st century. *Rev. Educ. Res.* 70, 151–179. doi: 10.3102/00346543070002151

Hidi, S., and Renninger, K. A. (2006). The four-phase model of interest development. *Educ. Psychol.* 41, 111–127. doi: 10.1207/s15326985ep4102_4

Hogan, T. P., and Brezinski, K. L. (2003). Quantitative estimation: one, two, or three abilities? *Math. Think. Learn.* 5, 259–280. doi: 10.1207/S15327833MTL0504_02

Ivrendi, A. (2011). Influence of self-regulation on the development of children's number sense. *Early Childhood Educ J.* 39, 239–247. doi: 10.1007/s10643-011-0462-0

Jansen, M., Lüdtke, O., and Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: examining between-person and within-person variation in five domains. *Contemp. Educ. Psychol.* 46, 116–127. doi: 10.1016/j.cedpsych.2016.05.004

Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x

Kelava, A., and Moosbrugger, H. (2020). "Deskriptivstatistische Itemanalyse und Testwertbestimmung" in *Testtheorie und Fragebogenkonstruktion*. eds. H. Moosbrugger and A. Kelava. *3rd* ed (Berlin: Springer), 13–38.

Keller, L., Preckel, F., Eccles, J. S., and Brunner, M. (2022). Top-performing math students in 82 countries: an integrative data analysis of gender differences in achievement, achievement profiles, and achievement motivation. *J. Educ. Psychol.* 114, 966–991. doi: 10.31234/osf.io/73wap

Köller, O., Baumert, J., and Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *J. Res. Math. Educ.* 32, 448–470. doi: 10.2307/749801

Krapp, A. (2000). "Interest and human development during adolescence: an educational-psychological approach" in *Motivational psychology of human development*. ed. J. Heckhausen (London: Elsevier), 109–128.

Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learn. Instr.* 12, 383–409. doi: 10.1016/S0959-4752(01)00011-1

Krapp, A. (2007). An educational–psychological conceptualisation of interest. *Int. J. Educ. Vocat. Guid.* 7, 5–21. doi: 10.1007/s10775-007-9113-9

Lee, J., and Chen, M. (2019). Cross-country predictive validities of non-cognitive variables for mathematics achievement: evidence based on TIMSS 2015. EURASIA journal of mathematics, science and technology. *Education* 15:8. doi: 10.29333/ejmste/106230

Lee, W., Lee, M., and Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemp. Educ. Psychol.* 39, 86–99. doi: 10.1016/j.cedpsych.2014.02.002

Lee, J., and Shute, V. J. (2010). Personal and social-contextual factors in K–12 academic performance: an integrative perspective on student learning. *Educ. Psychol.* 45, 185–202. doi: 10.1080/00461520.2010.493471

Lipton, J. S., and Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Dev.* 76, 978–988. doi: 10.1111/j.1467-8624.2005.00891.x

Luwel, K., Lemaire, P., and Verschaffel, L. (2005). Children's strategies in numerosity judgment. *Cogn. Dev.* 20, 448–471. doi: 10.1016/j.cogdev.2005.05.007

Luwel, K., and Verschaffel, L. (2008). Estimation of 'real' numerosities in elementary school children. *Eur. J. Psychol. Educ.* 23, 319–338. doi: 10.1007/BF03173002

Malofeeva, E., Day, J., Saco, X., Young, L., and Ciancio, D. (2004). Construction and evaluation of a number sense test with head start children. *J. Educ. Psychol.* 96, 648–659. doi: 10.1037/0022-0663.96.4.648

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., et al. (2018). *PISA 2012 Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Münster, New York: Waxmann.

May, P., and Bennöhr, J. (2018). 'Angaben zur Testentwicklung', in *Handbuch. Konzept, theoretische Grundlagen und Normierung. Kompetenzerfassung in Kindergarten und Schule von Vorschule/Kita bis zum 4. Schuljahr*. eds. P. May and J. Bennöhr. (Berlin: Cornelsen). 112–129.

May, P., and Bennöhr, J. (2021). *KEKS 3 Mathematik. Kompetenzerfassung in Kindergarten und Schule*. Berlin: Cornelsen.

Moosbrugger, H., and Kelava, A. (2020). "Qualitätsanforderungen an Tests und Fragebogen ("Gütekriterien")" in *Testtheorie und Fragebogenkonstruktion*. eds. H. Moosbrugger and A. Kelava. *3rd* ed (Berlin: Springer), 13–38.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: The National Council of Teachers of Mathematics.

Renninger, K. A., Ewen, L., and Lasher, A. K. (2002). Individual interest as context in expository text and mathematical word problems. *Learn. Instr.* 12, 467–490. doi: 10.1016/S0959-4752(01)00012-3

Ruwisch, S. (2014). Das Ungefähre zu schätzen wissen. Die Bedeutung des Ungefähren und wie man sich ihm nähert. *Grundschule Mathematik.* 42, 40–43.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012

Sayers, J., and Andrews, P. (2015). 'Foundational number sense: Summarising the development of an analytical framework', in *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education*, eds. K. Krainer, and N. Vondrová (Prague), 361–367.

Sayers, J., Andrews, P., and Björklund Boistrup, L. (2016). "The role of conceptual subitising in the development of foundational number sense" in *Mathematics education in the early years*. eds. T. Meany, O. Helenius, M. L. Johansson, T. Lange and A. Wernberg (Heidelberg, New York, Dordrecht, London: Springer), 371–394.

Sayers, J., Petersson, R., Rosenqvist, E., and Andrews, P. (2020). 'Estimation: an inadequately operationalised national curriculum competence', in *Proceedings of the British Society for Research into Learning Mathematics*, ed. R. Marks (BSRLM), 40:, 14.

Schipper, W. (2009). *Handbuch für den Mathematikunterricht an Grundschulen*. Braunschweig: Schroedel.

Schneider, M., Grabner, R. H., and Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *J. Educ. Psychol.* 101, 359–372. doi: 10.1037/a0013840

Schütte, S. (2008). *Qualität im Mathematikunterricht der Grundschule sichern. Für eine zeitgemäße Unterrichts-und Aufgabenkultur*. München: Oldenbourg.

Siegel, A. W., Goldsmith, L. T., and Madson, C. R. (1982). Skill in estimation problems of extent and numerosity. *J. Res. Math. Educ.* 13, 211–232. doi: 10.2307/748557

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., and Booth, J. L. (2005). "Development of numerical estimation: a review" in *Handbook of mathematical cognition*. ed. J. I. D. Campbell (Boca Ratan, FL: CRC Press), 197–212.

Sowder, J. (1992). "Estimation and number sense" in *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics*. ed. D. Grouws (Reston, VA: NCTM), 371–389.

Sowder, J., and Wheeler, M. (1989). The development of concepts and strategies used in computational estimation. *J. Res. Math. Educ.* 20, 130–146. doi: 10.5951/jresematheduc.20.2.0130

Sriraman, B., and Knott, L. (2009). The mathematics of estimation: possibilities for interdisciplinary pedagogy and social consciousness. *Interchange* 40, 205–223. doi: 10.1007/s10780-009-9090-7

Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., and Henschel, S. (2022). *IQB-Bildungstrend. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Münster: Waxmann Verlag.

Trautwein, U., Lüdtke, O., Nagy, N., Lenski, A., Niggli, A., and Schnyder, I. (2015). Using individual interest and conscientiousness to predict academic effort: additive, synergistic, or compensatory effects? *J. Pers. Soc. Psychol.* 109, 142–162. doi: 10.1037/pspp0000034

Urban, D., and Mayerl, J. (2011). *Regressionsanalyse: Theorie, Technik und Anwendung,4th*Edn. Wiesbaden: VS Verlag für Sozialwissenschaften.

Verschaffel, L., Greer, B., and De Corte, E. (2007). "Whole number concepts and operations" in *Second handbook of research on mathematics teaching and learning*. ed. F. Lester (Charlotte: Information Age Publishing), 557–628.

Viljaranta, J., Lerkkanen, M., Poikkeus, A., Aunola, K., and Nurmi, J. (2009). Cross-lagged relations between task motivation and performance in arithmetic and literacy in kindergarten. *Learn. Instr.* 19, 335–344. doi: 10.1016/j.learninstruc.2008.06.011

Walther, G., Geiser, H., Langeheine, R., and Lobemeier, K. (2003). "Mathematische Kompetenzen am Ende der vierten Jahrgangsstufe" in *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. eds. W. Bos, M. P. Lankes, K. Schwippert, G. Walther and R. Valtin (Münster: Waxmann Verlag), 189–226.

Wessolowski, S. (2014). Schätz mal, wie viele das sind! Schätzaufgaben als langfristiger Lerngegenstand. *Mathematik differenziert.* 1, 16–20.

Wong, T. T., Ho, C. S., and Tang, J. (2016). Consistency of response patterns in different estimation tasks. *J. Cogn. Dev.* 17, 526–547. doi: 10.1080/15248372.2015.1072091

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 14 July 2023
DOI 10.3389/fpsyg.2023.1185677

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Sum Kwing Cheung,
The Education University of Hong Kong,
Hong Kong SAR, China
Robert Reeve,
The University of Melbourne, Australia

*CORRESPONDENCE
Massimo Piccirilli
✉ massimo.piccirilli@unipg.it;
✉ piccirilli.massimo@gmail.com

RECEIVED 13 March 2023
ACCEPTED 03 July 2023
PUBLISHED 14 July 2023

CITATION
Piccirilli M, Lanfaloni GA, Buratta L, Ciotti B, Lepri A, Azzarelli C, Ilicini S, D'Alessandro P and Elisei S (2023) Assessment of math anxiety as a potential tool to identify students at risk of poor acquisition of new math skills: longitudinal study of grade 9 Italian students. *Front. Psychol.* 14:1185677.
doi: 10.3389/fpsyg.2023.1185677

Assessment of math anxiety as a potential tool to identify students at risk of poor acquisition of new math skills: longitudinal study of grade 9 Italian students

Massimo Piccirilli[1]*, Gianni Alberto Lanfaloni[2], Livia Buratta[3], Beatrice Ciotti[2], Alessandro Lepri[2], Cristina Azzarelli[2], Silvia Ilicini[2], Patrizia D'Alessandro[2] and Sandro Elisei[2]

[1]Department of Medicine, University of Perugia, Perugia, Italy, [2]Serafico Institute, Assisi, Perugia, Italy, [3]Department of Philosophy, Social Sciences and Education, University of Perugia, Perugia, Italy

Introduction: Numerous international educational institutions have sounded the alarm about the gradual increase in the number of students failing to achieve a sufficient level of proficiency in mathematical abilities. Thus, the growing interest in identifying possible solutions and factors interfering with learning seems justified. In recent years, special attention has accrued to the possible role played by emotional factors.

Methods: In the present investigation, students in the first grade of a technical vocational secondary school are followed to assess the influence of math anxiety (MA) on the development of skill acquisition in calculus. A math skills assessment test is administered on two occasions, at the beginning and end of the school year.

Results: Results highlighted that the score on the anxiety scale, administered at the beginning of the year, negatively correlated with the score obtained on the mathematics test, administered at the end of the school year: the higher the level of anxiety, the worse the performance. Furthermore, the score obtained in the second administration makes it possible to divide the students tested into two groups: students who improved their performance and students who did not benefit at all from repeating the test. In these two groups, an analysis of the relationships between the outcome of the end-of-year mathematics test and the level of MA at the beginning of the year showed that MA correlates negatively with performance only in students who will fail to acquire new expertise in mathematics over the course of the school year.

Discussion: The results suggest that MA may interfere with the smooth development of math skills. Assessing the level of MA at the beginning of the school year could prove to be a useful tool in identifying which and how many students are at risk of failing to achieve the skills expected from the usual course of instruction. A consideration of anxiety as one of the variables at play in the genesis of learning difficulties may prompt educators to modify teaching methodology and strategies by increasing focus on the impact of the emotional dimension on learning.

KEYWORDS

math anxiety, AMAS, math achievement, adolescence, educational psychology, mathematics performance, secondary school students

Introduction

The invention of numbers represents a milestone in human history that ultimately led, through progressive developments that took place over the past 4,000 years, to the number system currently in use (Boyer and Merzbach, 2010). At an individual level, the acquisition of mathematical skills turns out to be a highly complex process (Butterworth, 1999). Indeed, a newborn infant possesses the ability to distinguish between numbers with different values (e.g., one, two, and many), but genetic heritage is not sufficient to allow the manipulation of numbers (Dehaene, 1996). Achieving this manipulation requires a long period of learning, which in our society is mediated by teaching through compulsory schooling (Dehaene, 2021). That is why mathematics is one of the main domains of the school curriculum. A substantial proportion of subjects, however, struggle when confronted with contexts in which it is necessary to show some knowledge of mathematics (Fennema and Sherman, 1976). Difficulties in learning mathematics may indicate a clinical issue—a specific disorder called dyscalculia. In dyscalculia, the origin of the deficit is cognitive (Lucangeli, 1999; Passolunghi, 2011). However, even those who do not suffer from a cognitive disorder may demonstrate a negative attitude when faced with the need to study mathematics. In fact, the number of students who at the end of their schooling show that they have not achieved the minimum skills needed for everyday life appears to be constantly rising (Invalsiopen, 2022; Save the Children, 2022).

Numerous attempts have been made to understand the causes of these widespread difficulties in mathematics and, in recent years, great prominence has been given to math anxiety (MA), a concept introduced by Dreger and Aiken (1957) and later defined by Richardson and Suinn (1972) as "feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations." This negative reaction may manifest itself in various ways (Hembree, 1990): at an emotional level (as feelings of discomfort, apprehension, aversion, worry, frustration, or fear), at a physical level (as malaise, tachycardia, muscle tension, shortness of breath or sweating, and neuro vegetative reactivity) or at a behavioral level (as refusal to go to school or avoidance of homework and study). MA's severity may be such that it can result in a true phobia, namely, an unreasonable fear when faced with real or even imaginary number-related situations (Gough, 1954).

Implicit in the definition of MA is that it is a specific form of anxiety, distinct from state anxiety and trait anxiety as well as the anxiety that may occur when one has to perform activities other than mathematics, such as public speaking in the case of social phobia. The subsequent literature, with some exceptions, seems to have confirmed the specificity of MA (Hembree, 1990; Devine et al., 2012; Carey et al., 2017; Choe et al., 2019; Pizzie and Kraemer, 2019), and its theoretical structure as an autonomous entity has been further corroborated by the identification of its neurobiological basis (Lyons and Beilock, 2012; Young et al., 2012; Artemenko et al., 2015; Supekar et al., 2015; Suárez-Pellicioni et al., 2016).

Also implicit in the definition is that MA may adversely affect the ability to solve mathematical problems. Indeed, it is sufficiently well documented that subjects with high levels of MA tend to have poor performance in mathematics tasks although they are able to learn other school subjects in an adequate manner (Hembree, 1990; Ma, 1999; Ashcraft and Krause, 2007; Maloney et al., 2011; Barroso et al., 2021; Dowker and Sheridan, 2022).

In our previous study of first-year technical vocational secondary school students, we examined the relationships between performance on a standardized test to assess different mathematics skills and anxiety, assessed as trait anxiety, state anxiety, and math-specific anxiety (Buratta et al., 2019). The math test is aimed at evaluating the functional architecture of numerical processing mechanisms, according to the model of numerical cognition proposed by McCloskey et al. (1985): various subtests present problems of increasing complexity useful for measuring the level of ability achieved by the student on the basis of the school grade. Data, derived by simultaneous multivariate linear regression analysis exploring the role of the single anxiety scales (maths, trait, and state) on the mathematics performance scores, highlighted that MA was the only variable that was significantly linked to the arithmetical performance. Specifically, MA was not linked with basic math performance but with task assessing more advanced competence. On the contrary, math performance was not affected by the presence of trait or state anxiety.

In order to better understand this relevant issue, we have planned to reevaluate the same students at the end of their school year. To facilitate the analysis of the effects of anxiety on performance, we preferred to keep the difficulty of the mathematical test unchanged. As a matter of fact, having students face a more complex task in the second test than in the first would have made it difficult to differentiate the effects of anxiety from the consequences of increasing the difficulty of the task (Geary et al., 2019). For this purpose, at the end of the school year, the same students examined in the first stage of this investigation performed the same test that they had done at the beginning of the school year. Theoretically, at the end of their school year, students, having carried out the formal teaching activities foreseen by the institutional curriculum, should have acquired new knowledge and skills in mathematics. Consequentially, by running the same task for the second time at the end of the school year, all students should show an improvement in their performance, i.e., they should obtain a higher score on the second administration of the test. Contrariwise, a lower score in the second evaluation than in the first would indicate that the student has not benefited from the teaching offered during the school year. In our hypothesis, the presence of MA at the beginning of the school year could negatively interfere with the mathematical performance tested at the end of the course of study, despite the specific teaching activity received during the school year. Furthermore, we hypothesize that a high level of MA may hinder the ability to acquire new math skills and that the link to MA could be different for students who did not benefit at all from repeating the test than for students who improved their performance.

Subjects and method

Subjects

73 Italian-speaking students (59 male and 14 female) attending the first grade of a technical vocational secondary school (corresponding to grade nine of English-speaking schools and of the Italian INVALSI evaluation system) participated in this phase of the survey. The mean age at the time of the beginning of the school year, when students first participated in our study, was 14 years and 7

months (SD = 4 months, range = 14–15.1 years). Diagnosis of dyscalculia or a specific learning disorder (as documented by the certification issued by the health authority to the educational institution to organize a personalized treatment in accordance with the Italian guidelines on specific learning disabilities) was an exclusion criterion. The participants and their parents were informed about the purposes of the research and gave signed informed consent. Participation in the study did not include any type of reward. The study was approved by CEAS, the Local Ethics Committee of the Umbria Region of Italy.

Procedure

Students were tested on two separate occasions, approximately 6 months apart, once in January at the end of the first term (T0) and a second time in June at the end of the second term (T1). This period corresponds in Italian schools to the moments in which formal tests are carried out to measure the level of learning achieved by the students. All participants first filled in the three questionnaires that assessed, in order, trait, math, and state anxiety; later they completed the mathematical tests. They did all measures individually in class during school hours and were aware that the data obtained would remain anonymous and would not be shared with their teachers.

Measures

Abbreviated math anxiety scale (AMAS)

AMAS is a questionnaire consisting of nine items scored from 1 to 5 (on a five-point Likert scale); the score can thus range from a minimum value of 9 to a maximum value of 45, which is indicative of the highest possible level of math anxiety (Hopko et al., 2003). The items assess two distinct subscales: "Learning Math Anxiety" (five items that explore anxiety related to math study, such as "Watching the teacher break down a complex problem on the blackboard") and "Math Evaluation Anxiety" (four items that explore anxiety related to assessment situations, such as "Doing a written math examination/test"). In this study, we used the Italian version which was validated by Primi et al. (2014). When administered to secondary school students, the Italian adaptation of AMAS exhibits psychometric properties similar to those of the original test with respect to internal consistency (Cronbach's alpha ≥0.80) and test–retest reliability (Cronbach's alpha ≥0.81); furthermore, the two dimensions established in the original AMAS (Learning Math Anxiety and Math Evaluation Anxiety) were evident also in the Italian version as well as the invariance across genders. In addition, transcultural validity of math anxiety assessment with the AMAS has been documented in numerous studies (Vahedi and Farrokhi, 2011; Cipora et al., 2015) and has allowed collaborative studies between researchers of different languages, for example, Italian and English (Hill et al., 2016; Mammarella et al., 2018; Carey et al., 2019; Wang et al., 2020).

State and trait anxiety inventory (STAI-Y)

STAI-Y is a 40-item questionnaire scored from 1 to 4 (on a four-point Likert scale) and consists of two subscales of 20 items that assess state anxiety, which is anxiety related to a specific moment or event (such as might be represented by anxiety in relation to the test), and trait anxiety, which is a condition related mainly to personality characteristics (Spielberger, 1989). In the first phase of this investigation (Buratta et al., 2019), the use of STAI-Y allowed us to differentiate the role of MA from that of other forms of anxiety and to establish that the levels of trait anxiety and state anxiety do not correlate with mathematical performance and do not influence it negatively. The Italian version of the questionnaire, which was validated by Pedrabissi and Santinello (1989), was used in this study: the internal consistency values is 0.91 for the State anxiety scale and 0.85 for Trait anxiety scale; the test–retest reliability is 0.49 for state subscale and 0.82 for trait anxiety. Given the good psychometric properties and its availability in different languages that make it useful for cross-cultural investigations, the STAI-Y is the best-known and most widely used self-report questionnaire to assess anxiety in research and clinical practice.

Battery for the assessment of calculation ability (ABCA 14-16)

ABCA 14-16 is a battery of paper-pencil tests for the assessment of mathematical skills in 14- to 16-year-old subjects (Baccaglini-Frank et al., 2013). This battery of tests has been constructed with reference to the modular model of McCloskey et al. (1985) which hypothesizes that the mental representation of numerical knowledge is independent of other cognitive systems and is structured in three modules which are in turn functionally distinct. The different tests in the battery provide a specific profile that identifies which components are deficient within an individual student's mathematical skills comparing the percentile scores to the cut-off criteria of the normative sample. The items assess different levels of mathematical proficiency by requiring the solution of tasks of different complexity (for example, "145.28–23.39 =" or "57.8 × 2.94 ="). In the present study, we considered subtests that in the first phase of our investigation were affected by the level of MA. These subtests consist of advanced math skills that investigate how students have stored combinations of numbers within the calculation system and whether they are able to access them automatically (for example, "$2 + 3 \times 4 =$" or "$100/10^2 =$"). Scoring, which is based on the number of correct answers, ranges from a minimum of zero to a maximum of 28, and high scores correspond to superior performance. The ABCA test is widely used in Italy to assess the achievement of the educational objectives set by the school curriculum.

Statistical analysis

The Java Structural Program was used for data analysis (JASP Team, 2022, Version 0.16.3). Descriptive statistics in terms of mean and standard deviation were employed to describe the scores obtained in the AMAS questionnaire and ABCA 14-16 test. The relationship between MA levels at T0 and math performance at T1 was examined using the Pearson correlation coefficient.

To further investigate this relationship, we divided the group of students into more homogeneous subgroups. As regards the AMAS questionnaire, the students who obtained scores above the mean by one standard deviation constituted the subgroup with high levels of anxiety (HMA), while the students with scores below the mean by one standard deviation constituted the subgroup characterized by low levels of anxiety (LMA). Student t-test was used to compare the scores obtained by the two subgroups in mathematics performance.

As regards math performance at T1, the score obtained in the math test at T1 was used as an indicator of the student's ability to make use of the study period between the two assessments to acquire better math competence; a higher score on the second than on the first assessment identified the subgroup of students who improved performance, showing that they have learned new calculus skills (subgroup with improved math performance, IMP), while other students constituted the subgroup whose performance did not improve, showing that they have failed to acquire new calculus skill (subgroup with worsened math performance, WMP). A paired t-test was utilized to compare the mathematical performance in the two assessments at T0 and T1 and a linear regression, based on the AMAS questionnaire score at T0 as an independent variable and the math performance score at T1 as a dependent variable, was used to evaluate the possible influence exerted by the MA on math performance of the two subgroups of students.

Results

The AMAS questionnaire indicated an average anxiety level of 21.79 ± 6.08. Subscale "Learning Math" showed an average score of 8.45 (SD = 2.77) and subscale "Math Evaluation" an average score of 13.42 (SD = 4.15). There were 15 (20, 27%) students belonging to the HMA subgroup characterized by high levels of anxiety (M = 30.33, SD = 1.88), while 14 (19, 18%) belong to the LMA subgroup characterized by low levels of anxiety (M = 14, SD = 1.07). HMA students showed significantly ($p = 0.001$) worse math performance (M = 12.66, SD = 7.28) than LMA students (M = 20.71, SD = 3.73).

As regards the mathematical test performed at T1, the average score obtained was 16.37 (SD = 6.3). The correlation coefficient showed an inverse relationship between the MA and the score obtained in the math test ($r = -0.386$, $p = 0.002$): the higher the MA level recorded at the beginning of the school year, the lower the score on the math test at the end of the year. This relationship persisted even considering separately the two "Learning Math" ($r = -0.388$, $p = 0.002$) and "Math Evaluation" ($r = -0.306$, $p = 0.02$) components of the AMAS questionnaire.

In the interpretation of this inverse correlation, it must be kept in mind that an increase in the score in the second administration of the mathematical test compared to the first can be interpreted as an index of the acquisition of new math skills resulting from the teaching received during the school year. Conversely, the deterioration of performance, i.e., the decrease in the score obtained when the test was administered for the second time, can be interpreted as the consequence of the inability to acquire better math skills. Actually, there were 42 (57.53%) belonging to the IMP subgroup characterized by an improvement in math performance from an average score of 12.4 ± 7.1 to 17.47 ± 6.18 ($t = 8.93$; $p < 0.001$), while the remaining 31 students (42.47%) belonged to the WMP subgroup characterized by a math performance deterioration from an average score of 17.03 ± 5.52 to 14.87 ± 6.24 ($t = -4.75$; $p < 0.001$). A linear regression, conducted separately for the two subgroups of students, showed that the relationship between MA level and mathematics performance was significant for the 31 students whose performance did not improve with time [$F(29,1) = 19.8$; $r = -0.637$; $R^2 = 0.406$; $p < 0.001$] but not for the 42 students whose performance improved [$F(40,1) = 1.564$; $r = 0.194$; $R^2 = 0.038$, $p = 0.22$]. For the students who received no advantage from a repetition of the test, the model accounted for 40% of the variance (Figure 1).

Discussion

In the present study, the development of acquisition of math skills in a group of 73 students in the first grade of secondary school was followed. For this purpose, a math skills assessment test (the Battery for the assessment of calculation ability ABCA 14-16) was administered for the first time at the beginning and repeated at the end of the year. This battery of math tests is commonly used in Italian schools to identify the specific difficulties encountered by the student in solving mathematical problems as well as to follow the evolution of mathematical skills in relation to age and education. The test score is considered a good measure of the degree of mathematical proficiency.

The investigation aimed to examine whether students' math anxiety can affect their acquisition of new mathematical abilities. Results highlighted that the score on the anxiety scale, administered at the beginning of the year, negatively correlated with the score obtained on the mathematics test, administered at the end of the school year: the higher the level of anxiety, the worse the performance.

In recent years, special attention has accrued to the possible role played by emotional factors and several studies have documented that MA negatively affects math performance. First, MA induces avoidance behavior towards mathematics, thus consequently reducing learning opportunities. Second, it is documented that MA can interfere at a cognitive function level with both working memory and inhibitory processes (Hopko et al., 2002; Caviola et al., 2012; Mammarella et al., 2018; Skagerlund et al., 2019; Soltanlou et al., 2019; Pelegrina et al., 2020; Van den Bussche et al., 2020; Dowker and Sheridan, 2022). Moreover, studies in which MA has purposefully increased yield poor performance (Kellogg et al., 1999; Galdi et al., 2013), while those in which MA is decreased have improved performance (Furner and Duffy, 2002; Park et al., 2014). However, it should be noted that these data are mainly derived from cross-sectional studies, while the results of studies that have used a longitudinal design are not completely consistent. In our investigation, the level of perceived MA at the beginning of the school year correlated with the score obtained in the end-of-year test, showing that it has a negative effect over time. Interestingly, this negative effect seems to be exerted exclusively on those students who are unable to improve their performance, despite the teaching received and the time elapsed between the two administrations of the same test. Indeed, in the students who did not gain from a retest, the performance had a negative relationship to the MA level. Instead, this relationship was not present in students who improved their skills over time. Given that the improvement in the score obtained on the test is attributable to the achievement of greater competence, the finding of a deterioration in performance, when the task is presented for the second time, can be interpreted as the consequence of an inability to learn how to solve the problems proposed by the math test. Put differently, MA correlated negatively with performance only in students who failed to acquire new knowledge in mathematics over the course of a school year. Ultimately, MA seems to act as a factor that can interfere with the acquisition of new knowledge and skills in mathematics. Furthermore, data allow suggesting the possibility of using an early MA-level assessment to

10.3389/fpsyg.2023.1185677

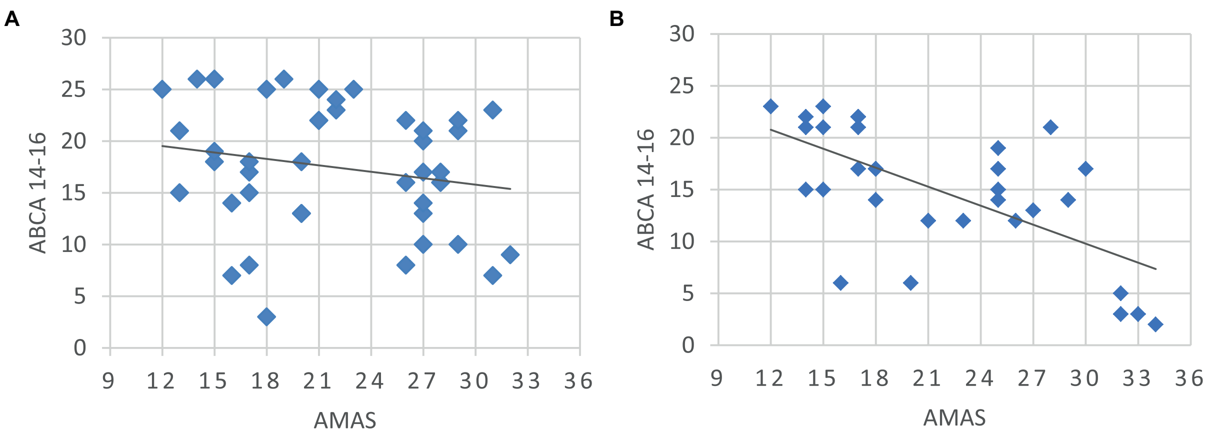


FIGURE 1
Relationship between perceived anxiety about mathematics and performance on the mathematics test. **(A)** Students whose performance improves over time ($r = 0.194$; $R^2 = 0.038$; $p = 0.22$). **(B)** Students whose performance does not improve over time ($r = -0.637$; $R^2 = 0.406$; $p = 0.001$). AMAS, score obtained at the beginning of the school year on the on the Math Anxiety Assessment Questionnaire; ABCA 14-16, score obtained at the end of the school year on the Italian battery for the assessment of calculation ability.

identify students at high risk of poor acquisition of new calculus skills and, consequently, of experiencing difficulties in studying mathematics.

Clearly, our data do not imply that anxiety is the only factor involved in negatively affecting math performance. Certainly, the role of teachers and the family context cannot be overlooked; for example, it has been suggested that MA does not stem directly from the study of mathematics but occurs mainly due to the way mathematics is taught and presented (Turner et al., 2002; Geist, 2010). Moreover, the role of cultural stereotypes should not be underestimated. Cultural stereotypes, which are sometimes transmitted at the family level as well as at the school level, include the idea that if one is not born with a “math gene,” it is useless to strive to study math (Rattan et al., 2012; Goetz et al., 2013; Vukovic et al., 2013). In this regard, it seems fair to emphasize that the characteristics of our sample ensured an acceptable uniformity of contextual aspects: math teachers as well as the environmental context remained unchanged throughout the school year; therefore, these factors should not have significantly affected the results.

Another factor called into question was the age at which the math test was administered. According to some surveys (Wu et al., 2012; Hill et al., 2016; Carey et al., 2019), anxiety increases with age, and the mechanism linking anxiety and performance might be different at different educational stages. In most of the studies, primary and middle school students were examined (Ma and Xu, 2004; Vukovic et al., 2013; Cargnelutti et al., 2017; Gunderson et al., 2018; Geary et al., 2019), while studies on secondary schools were few (Passolunghi et al., 2016; Wang et al., 2020). This variable could justify some discrepancies found in literature surveys (Pekrun et al., 2017; Szczygiel, 2020). From this point of view, it may be helpful to note that the participants in our study belong to the same age group; thus, also the age-related factor does not seem able to influence the results. Furthermore, it should be emphasized that the age of the students on which our study has been conducted, i.e., adolescence, may represent a critical period, both emotionally and behaviorally. With respect to learning mathematics, the first year of secondary school constitutes a time of transition and a turning point between previous experiences and calls for planning an approach to further learning. In this light, conducting the present investigation at this stage of the school journey could be of particular importance in identifying the risk of math difficulties.

Another aspect to consider in interpreting the results is the practice effect, namely, the improvement in performance that is observed whenever the same test is administered several times to a subject; however, the finding that a high percentage of students showed no improvement in performance in the second assessment suggests that this effect cannot account for the data obtained.

In conclusion, our data, in line with what has been described in the literature (Hembree, 1990; Ma, 1999; Maloney et al., 2011; Wu et al., 2012; Núñez-Peña and Suárez-Pellicioni, 2014), confirms the high frequency in which MA occur among students; furthermore, our findings show that MA negatively correlates with the score on the end-of-year math test and suggest that it unfavorably affects the ability to acquire new calculus skills.

Some limitations of this study must be kept in mind. First, the surveyed population consists of students who all attend a technical vocational school where the teaching of mathematics takes on a different value than it does in other educational institutions. Therefore, the results obtained cannot be generalized to other types of students. Indeed, in Italy, mathematics teaching differs quantitatively and qualitatively according to the type of secondary school; the main division is between schools oriented towards the study of humanities, arts, and social sciences and schools more oriented towards physical sciences and technology; in particular, the technical vocational education is oriented towards practical subjects and enable the students to start searching for a job as soon as they completed their studies. It has been proved that Italian students who had chosen science and technology courses show significantly less math anxiety than students who had chosen humanities and social sciences courses (Primi et al., 2014) but little is known about the possible different effects of MA on the math performance of students following different types of study programs (Schommer-Aikins et al., 2015; Barroso et al., 2021). Future studies would be needed to evaluate the extent to which the data obtained are generalizable to various types of courses (Morsanyi et al., 2017; Paechter et al., 2017).

Second, neither the presence of anxiety in the study of school subjects other than mathematics nor the possibility of MA having a negative effect on other study subjects were investigated. It should also be noted that the score of the test was limited to the number of correct answers, according to the instruction manual; other possible measures, such as the time taken to complete the test, have not been considered; the task was carried out during the time for the mathematics lesson as if it were a usual class assignment. Moreover, with regard to the possible mechanisms by which MA is believed to interfere with performance, the present investigation did not include the administration of cognitive tests, such as those related to executive, verbal, and visuospatial functions; however, given that the test administered for the assessment of attainment in mathematics requires adequate working memory functioning (Lucangeli, 1999), the data obtained seem to be in line with what is already known in the literature about the negative influence MA can exert on executive functions (Ashcraft and Kirk, 2001; Hopko et al., 2002; Ashcraft and Krause, 2007; Caviola et al., 2012; Passolunghi et al., 2016; Mammarella et al., 2018; Skagerlund et al., 2019; Soltanlou et al., 2019; Pelegrina et al., 2020; Van den Bussche et al., 2020). Lastly, based on the experimental design used, no conclusion can be inferred about the role played by the difficulty of solving the math test in the genesis of MA, which is a topic that would have required a different experimental design. Similarly, the causal order remains unresolved, that is, whether it is MA that initially reduces performance or *vice-versa* (Ma and Xu, 2004; Carey et al., 2016). It is likely, as suggested by the so-called "reciprocal theory," that MA and performance continuously exchange roles of cause and effect, thus triggering a vicious cycle and progressively reinforcing each other (Maloney and Beilock, 2012; Carey et al., 2016; Gunderson et al., 2018). In any case, our data suggest that once triggered, regardless of what the "primum movens" was, MA contributes significantly to hindering the acquisition of mathematical skills.

Conclusion

The growing interest in MA appears to be justified by the fact that acquiring competence in this field takes on value in choosing not only the path of continuation of one's schooling but also one's professional future (Dowker et al., 2016; Morsanyi et al., 2017; Carey et al., 2019; Vargas, 2021). A negative attitude to mathematics, in addition to causing a great limitation in life choices (it may lead one to avoid professions that require a commitment to a context related to aspects of mathematics), certainly also entails great difficulties in daily life, especially in social contexts that are highly characterized by technology, such as those of the present day (Suri et al., 2013). Thus, the alarm expressed by international institutions about the decline of math competence in schools seems justified (Invalsiopen, 2022; Save the Children, 2022), especially because the decline is accompanied by an apparent lack of interest in finding solutions (Fondazione Rocca, 2022).

The present investigation can contribute to the understanding that the obstacle to acquiring ever-better math skills may be represented not only by cognitive issues but also, and especially, by emotional issues (Devine et al., 2018; Abín et al., 2020; Passolunghi et al., 2020). Considering anxiety as one of the variables at play in the genesis of learning difficulties may induce teachers to significantly modify their teaching methodology and strategies (Phelps-Gregory et al., 2020), on the one hand abandoning the idea that either you have skills or you do not, and on the other hand gaining awareness that promoting logical reasoning alone is not enough to achieve success (Brown et al., 2008; Devine et al., 2018). There are numerous studies in the literature that aim to propose modes of an intervention designed to take into account the impact of the emotional dimension on learning (Jamieson et al., 2010; Yeager and Dweck, 2012; Brunyé et al., 2013; Park et al., 2014; Supekar et al., 2015; Sokolowski and Necka, 2016; Barroso et al., 2021; Hausman et al., 2021; Samuel and Warner, 2021). From this perspective, an assessment of MA performed during high school entry could become a useful tool to identify which and how many students in a class are at risk of poor acquisition of math skills in the usual course of education. Further studies are needed to confirm this hypothesis, generalize the results, and attempt to answer the many outstanding questions about the relationships between MA and mathematics performance.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by CEAS Umbria—Comitato Etico Aziende Sanitarie Umbria. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

MP contributed to the conception and design of the study and wrote the first draft of the manuscript. GL contributed to the conception and design of the study. CA and SI collected and organized the database. LB, BC, and AL performed the statistical analysis. PD'A and SE contributed to manuscript revision. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abín, A., Núñez, J. C., Rodríguez, C., Cueli, M., García, T., and Rosário, P. (2020). Predicting mathematics achievement in secondary education: the role of cognitive, motivational, and emotional variables. *Front. Psychol.* 11:876. doi: 10.3389/fpsyg.2020.00876

Artemenko, C., Daroczy, G., and Nuerk, H. C. (2015). Neural correlates of math anxiety–an overview and implications. *Front. Psychol.* 6:1333. doi: 10.3389/fpsyg.2015.01333

Ashcraft, M. H., and Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *J. Exp. Psychol. Gen.* 130, 224–237. doi: 10.1037//0096-3445.130.2.224

Ashcraft, M. H., and Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychon. Bull. Rev.* 14, 243–248. doi: 10.3758/bf03194059

Baccaglini-Frank, A., Perona, M., Bettini, P., Caviola, S., and Lucangeli, D. (2013) *Test ABCA 14–16. Prove di abilità di calcolo avanzato per la scuola secondaria di secondo grado*. Italy: Erickson Ed

Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., and Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychol. Bull.* 147, 134–168. doi: 10.1037/bul0000307

Boyer, C. B., and Merzbach, U. C. (2010). *A history of mathematics*. San Francisco, CA: Jossey-Bass Ed.

Brown, M., Brown, P., and Bibby, T. (2008). I would rather die: reasons given by 16-year-olds for not continuing their study of mathematics. *Res. Math. Educ.* 10, 3–18. doi: 10.1080/14794800801915814

Brunyé, T. T., Mahoney, C. R., Giles, G. E., Rapp, D. N., Taylor, H. A., and Kanarek, R. B. (2013). Learning to relax: evaluating four brief interventions for overcoming the negative emotions accompanying math anxiety. *Learn. Individ. Differ.* 27, 1–7. doi: 10.1016/j.lindif.2013.06.008

Buratta, L., Piccirilli, M., Lanfaloni, G. A., Ilicini, S., Bedetti, C., and Elisei, S. (2019). Mathematics anxiety and cognitive performance in adolescent students. *Psychiatr. Danub.* 31, 479–485.

Butterworth, B. (1999). *The mathematical brain*. London: Macmillan Ed.

Carey, E., Devine, A., Hill, F., Dowker, A., McLellan, R., and Szűcs, D. (2019). Understanding mathematics anxiety: investigating the experiences of UK primary and secondary school students. Available at: https://doi.org/10.17863/CAM.37744.

Carey, E., Devine, A., Hill, F., and Szűcs, D. (2017). Differentiating anxiety forms and their role in academic performance from primary to secondary school. *PLoS One* 12:e0174418. doi: 10.1371/journal.pone.0174418

Carey, E., Hill, F., Devine, A., and Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Front. Psychol.* 6:1987. doi: 10.3389/fpsyg.2015.01987

Cargnelutti, E., Tomasetto, C., and Passolunghi, M. C. (2017). How is anxiety related to math performance in young students? A longitudinal study of grade 2 to grade 3 children. *Cogn. Emot.* 31, 755–764. doi: 10.1080/02699931.2016.1147421

Caviola, S., Mammarella, I. C., Lucangeli, D., and Cornoldi, C. (2012). The involvement of working memory in children's exact and approximate mental additions. *J. Exp. Child Psychol.* 112, 141–160. doi: 10.1016/j.jecp.2012.02.005

Choe, K. W., Jenifer, J. B., Rozek, C. S., Berman, M. G., and Beilock, S. L. (2019). Calculated avoidance: math anxiety predicts math avoidance in effort-based decision-making. *Sci. Adv.* 5:eaay1062. doi: 10.1126/sciadv.aay1062

Cipora, K., Szczygieł, M., Willmes, K., and Nuerk, H. C. (2015). Math anxiety assessment with the abbreviated math anxiety scale: applicability and usefulness: insights from the polish adaptation. *Front. Psychol.* 6:1833. doi: 10.3389/fpsyg.2015.01833

Dehaene, S. (1996). *La bosse des maths*. Paris: Odile Jacob Ed.

Dehaene, S. (2021). *How we learn: the new science of education and the brain*. Paris: Odile Jacob Ed.

Devine, A., Fawcett, K., Szűcs, D., and Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behav. Brain Funct.* 8:33. doi: 10.1186/1744-9081-8-33

Devine, A., Hill, F., Carey, E., and Szűcs, D. (2018). Cognitive and emotional math problems largely dissociate: prevalence of developmental dyscalculia and mathematics anxiety. *J. Educ. Psychol.* 110, 431–444. doi: 10.1037/edu0000222

Dowker, A., Sarkar, A., and Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years? *Front. Psychol.* 7:508. doi: 10.3389/fpsyg.2016.00508

Dowker, A., and Sheridan, H. (2022). Relationships between mathematics performance and attitude to mathematics: influences of gender, test anxiety, and working memory. *Front. Psychol.* 13:814992. doi: 10.3389/fpsyg.2022.814992

Dreger, R. M., and Aiken, L. R. (1957). The identification of number anxiety in a college population. *J. Educ. Psychol.* 48, 344–351. doi: 10.1037/h0045894

Fennema, E., and Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales; instruments designed to measure attitudes towards the learning of mathematics by females and males. *J. Res. Math. Educ.* 7:324. doi: 10.2307/748467

Fondazione Rocca. (2022) *Scuola, i numeri da cambiare*. Florence: Giunti Ed.

Furner, J. M., and Duffy, M. L. (2002). Equity for all students in the new millennium: disabling math anxiety. *Interv. Sch. Clin.* 38, 67–74. doi: 10.1177/10534512020380020101

Galdi, S., Cadinu, M., and Tomasetto, C. (2013). The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child Dev.* 85, 250–263. doi: 10.1111/cdev.12128

Geary, D. C., Hoard, M. K., Nugent, L., Chu, F. W., Scofield, J. E., and Hibbard, D. F. (2019). Sex differences in mathematics anxiety and attitudes: concurrent and longitudinal relations to mathematical competence. *J. Educ. Psychol.* 111, 1447–1461. doi: 10.1037/edu0000 355

Geist, E. (2010). The anti-anxiety curriculum: combating math anxiety in the classroom. *J. Instruct. Psychol.* 37, 24–31.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., and Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychol. Sci.* 24, 2079–2087. doi: 10.1177/0956797613486989

Gough, M. F. (1954). Why failures in mathematics? Mathemaphobia: causes and treatments. *Clear. House* 28, 290–294. doi: 10.1080/00098655.1954.11476830

Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., and Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *J. Cogn. Dev.* 19, 21–46. doi: 10.1080/15248372.2017.1421538

Hausman, H., Myers, S. J., and Rhodes, M. G. (2021). Improving metacognition in the classroom. *Z. Psychol.* 229, 89–103. doi: 10.1027/2151-2604/a000440

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* 21, 33–46. doi: 10.2307/749455

Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., and Szűcs, D. (2016). Maths anxiety in primary and secondary school students: gender differences, developmental changes and anxiety specificity. *Learn. Individ. Differ.* 48, 45–53. doi: 10.1016/j.lindif.2016.02.006

Hopko, D. R., Mahadevan, R., Bare, R. L., and Hunt, M. A. (2003). The abbreviated math anxiety scale (AMAS): construction, validity, and reliability. *Assessment* 10, 178–182. doi: 10.1177/1073191103010002008

Hopko, D. R., Mcneil, D. W., Gleason, P. J., and Rabalais, A. E. (2002). The emotional Stroop paradigm: performance as a function of stimulus properties and self-reported mathematics anxiety. *Cogn. Ther. Res.* 26, 157–166. doi: 10.1023/A:1014578218041

Invalsiopen. (2022). Rapporto nazionale 2022. Available at: https://www.invalsiopen.it/risultati/risultati-prove-invalsi-2022/

Jamieson, J. P., Mendes, W. B., Blackstock, E., and Schmader, T. (2010). Turning the knots in your stomach into bows: reappraising arousal improves performance on the GRE. *J. Exp. Soc. Psychol.* 46, 208–212. doi: 10.1016/j.jesp.2009.08.015

JASP Team (2022). JASP (Version 0.16.3). Available at: https://jasp-stats.org/

Kellogg, J. S., Hopko, D. R., and Ashcraft, M. H. (1999). The effects of time pressure on arithmetic performance. *J. Anxiety Disord.* 13, 591–600. doi: 10.1016/s0887-6185(99)00025-0

Lucangeli, D. (1999). *Il farsi ed il disfarsi del numero: conoscenze, modelli e metodi a confronto*. Roma: Borla Ed.

Lyons, I. M., and Beilock, S. L. (2012). When math hurts: math anxiety predicts pain network activation in anticipation of doing math. *PLoS One* 7:e48076. doi: 10.1371/journal.pone.0048076

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *J. Res. Math. Educ.* 30, 520–540. doi: 10.2307/749772

Ma, X., and Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: a longitudinal panel analysis. *J. Adolesc.* 27, 165–179. doi: 10.1016/j.adolescence.2003.11.003

Maloney, E., Ansari, D., and Fugelsang, J. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *Q. J. Exp. Psychol.* 64, 10–16. doi: 10.1080/17470218.2010.533278

Maloney, E. A., and Beilock, S. (2012). Math anxiety: who has it, why it develops, and how to guard against it. *Trends Cogn. Sci.* 16, 404–406. doi: 10.1016/j.tics.2012.06.008

Mammarella, I. C., Caviola, S., Giofrè, D., and Borella, E. (2018). Separating math from anxiety: the role of inhibitory mechanisms. *Appl. Neuropsychol. Child* 7, 342–353. doi: 10.1080/21622965.2017.1341836

McCloskey, M., Caramazza, A., and Basili, A. (1985). Cognitive mechanisms in number processing and calculation: evidence from dyscalculia. *Brain Cogn.* 4, 171–196. doi: 10.1016/0278-2626(85)90069-7

Morsanyi, K., Mammarella, I. C., Szűcs, D., Tomasetto, C., Primi, C., and Maloney, E. A. (2017). Mathematical and statistics anxiety: educational, social, developmental and cognitive perspectives. *Front. Psychol.* 7:1083. doi: 10.3389/fpsyg.2016.01083

Núñez-Peña, M. I., and Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: an ERP study of the size and distance effects. *Biol. Psychol.* 103, 176–183. doi: 10.1016/j.biopsycho.2014.09.004

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., and Papousek, I. (2017). Mathematics anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. *Front. Psychol.* 8:1196. doi: 10.3389/fpsyg.2017.01196

Park, D., Ramirez, G., and Beilock, S. (2014). The role of expressive writing in math anxiety. *J. Exp. Psychol. Appl.* 20, 103–111. doi: 10.1037/xap0000013

Passolunghi, M. C. (2011). Cognitive and emotional factors in children with mathematical learning disabilities. *Int. J. Dis. Dev. Educ.* 58, 61–73. doi: 10.1080/1034912X.2011.547351

Passolunghi, M. C., Caviola, S., De Agostini, R., Perin, C., and Mammarella, I. C. (2016). Mathematics anxiety, working memory, and mathematics performance in secondary-school children. *Front. Psychol.* 7:42. doi: 10.3389/fpsyg.2016.00042

Passolunghi, M. C., De Vita, C., and Pellizzoni, S. (2020). Math anxiety and math achievement: the effects of emotional and math strategy training. *Dev. Sci.* 23:e12964. doi: 10.1111/desc.12964

Pedrabissi, L., and Santinello, M. (1989). Verifica della validità dello STAI forma Y di Spielberger. *Boll. Psicol. Appl.* 191-192, 11–14.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., and Goetz, T. (2017). Achievement emotions and academic performance: longitudinal models of reciprocal effects. *Child Dev.* 88, 1653–1670. doi: 10.1111/cdev.12704

Pelegrina, S., Justicia-Galiano, M. J., Martín-Puga, M. E., and Linares, R. (2020). Math anxiety and working memory updating: difficulties in retrieving numerical information from working memory. *Front. Psychol.* 11:669. doi: 10.3389/fpsyg.2020.00669

Phelps-Gregory, C. M., Frank, M., and Spitzer, S. M. (2020). Prospective elementary teachers' beliefs about mathematical myths: a historical and qualitative examination. *Teach. Educ.* 55, 6–27. doi: 10.1080/08878730.2019.1618423

Pizzie, R. G., and Kraemer, D. J. M. (2019). The academic anxiety inventory: evidence for dissociable patterns of anxiety related to math and other sources of academic stress. *Front. Psychol.* 9:2684. doi: 10.3389/fpsyg.2018.02684

Primi, C., Busdraghi, C., Tommasetto, C., Morsanyi, K., and Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: validity, reliability and gender invariance of the abbreviated math anxiety scale (AMAS). *Learn. Individ. Differ.* 34, 51–56. doi: 10.1016/j.lindif.2014.05.012

Rattan, A., Good, C., and Dweck, C. S. (2012). "It's ok — not everyone can be good at math": instructors with an entity theory comfort (and demotivate) students. *J. Exp. Soc. Psychol.* 48, 731–737. doi: 10.1016/j.jesp.2011.12.012

Richardson, F. C., and Suinn, R. M. (1972). The mathematics anxiety rating scale: psychometric data. *J. Counsel. Psychol.* 19, 551–554. doi: 10.1037/h0033456

Samuel, T. S., and Warner, J. (2021). "I can math!": reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Community Coll. J. Res. Pract.* 45, 205–222. doi: 10.1080/10668926.2019.1666063

Save the Children (2022). Alla ricerca del tempo perduto. Save the children Italia Ed. Available at: https://www.savethechildren.it/cosa-facciamo/pubblicazioni/alla-ricerca-del-tempo-perduto

Schommer-Aikins, M., Unruh, S., and Morphew, J. (2015). Epistemological belief congruency in mathematics between vocational technology students and their instructors. *J. Educ. Train. Stud.* 3, 137–145. doi: 10.11114/jets.v3i4.859

Skagerlund, K., Östergren, R., Västfjäll, D., and Träff, U. (2019). How does mathematics anxiety impair mathematical abilities? Investigating the link between math anxiety, working memory, and number processing. *PLoS One* 14:e0211283. doi: 10.1371/journal.pone.0211283

Sokolowski, H. M., and Necka, E. A. (2016). Remediating math anxiety through cognitive training: potential roles for math ability and social context. *J. Neurosci.* 36, 1439–1441. doi: 10.1523/JNEUROSCI.4039-15.2016

Soltanlou, M., Artemenko, C., Dresler, T., Fallgatter, A. J., Ehlis, A. C., and Nuerk, H. C. (2019). Math anxiety in combination with low visuospatial memory impairs math learning in children. *Front. Psychol.* 10:89. doi: 10.3389/fpsyg.2019.00089

Spielberger, C. D. (1989). *State-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press

Suárez-Pellicioni, M., Núñez-Peña, M. I., and Colomé, À. (2016). Math anxiety: a review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cogn. Affect. Behav. Neurosci.* 16, 3–22. doi: 10.3758/s13415-015-0370-7

Supekar, K., Iuculano, T., Chen, L., and Menon, V. (2015). Remediation of childhood math anxiety and associated neural circuits through cognitive tutoring. *J. Neurosci.* 35, 12574–12583. doi: 10.1523/JNEUROSCI.0786-15.2015

Suri, R., Monroe, K. B., and Koe, U. (2013). Math anxiety and its effects on consumers' preference for price promotion formats. *J. Acad. Mark. Sci.* 41, 271–282. doi: 10.1007/s11747-012-0313-6

Szczygiel, M. (2020). Gender, general anxiety, math anxiety and math achievement in early school-age children. *Issues Educ. Res.* 30, 1126–1142. Available at: http://www.iier.org.au/iier30/szczygiel.pdf

Turner, J. C., Meyer, D. K., Anderman, E. M., Midgley, C., Gheen, M., Kang, Y., et al. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: a multimethod study. *J. Educ. Psychol.* 94, 88–106. doi: 10.1037//0022-0663.94.1.88

Vahedi, S., and Farrokhi, F. (2011). A confirmatory factor analysis of the structure of abbreviated math anxiety scale. *Iran. J. Psychiatry* 6, 47–53.

Van den Bussche, E., Vanmeert, K., Aben, B., and Sasanguie, D. (2020). Too anxious to control: the relation between math anxiety and inhibitory control processes. *Sci. Rep.* 10:19922. doi: 10.1038/s41598-020-76920-7

Vargas, R. A. V. (2021). A literature review on math anxiety and learning mathematics: a general overview. *J. Educ. Res. Rev.* 9, 102–108. doi: 10.33495/jerr_v9i5.21.112

Vukovic, R. K., Roberts, S. O., and Wright, L. G. (2013). From parental involvement to children's mathematical performance: the role of mathematics anxiety. *Early Educ. Dev.* 24, 446–467. doi: 10.1080/10409289.2012.693430

Wang, Z., Rimfeld, K., Shakeshaft, N., Schofield, K., and Malanchini, M. (2020). The longitudinal role of mathematics anxiety in mathematics development: issues of gender differences and domain-specificity. *J. Adolesc.* 80, 220–232. doi: 10.1016/j.adolescence.2020.03.003

Wu, S. S., Barth, M., Amin, H., Malcarne, V., and Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Front. Psychol.* 3:162. doi: 10.3389/fpsyg.2012.00162

Yeager, D. S., and Dweck, C. S. (2012). Mindsets that promote resilience: when students believe that personal characteristics can be developed. *Educ. Psychol.* 47, 302–314. doi: 10.1080/00461520.2012.722805

Young, C. B., Wu, S. S., and Menon, V. (2012). The neurodevelopmental basis of math anxiety. *Psychol. Sci.* 23, 492–501. doi: 10.1177/0956797611429134

frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 01 August 2023
DOI 10.3389/feduc.2023.1096148

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Xiang Hu,
Renmin University of China, China
Katja Upadyaya,
University of Helsinki, Finland

*CORRESPONDENCE
Katharina Reschke
✉ Reschke@ibw.uni-heidelberg.de

RECEIVED 11 November 2022
ACCEPTED 17 July 2023
PUBLISHED 01 August 2023

CITATION
Reschke K, Steinmayr R and Spinath B (2023) Predicting students' math self-concepts to explain gender differences through teachers' judgments and students' perceived teachers' judgments.
Front. Educ. 8:1096148.
doi: 10.3389/feduc.2023.1096148

Predicting students' math self-concepts to explain gender differences through teachers' judgments and students' perceived teachers' judgments

Katharina Reschke[1]*, Ricarda Steinmayr[2] and Birgit Spinath[3]

[1]Institute for Education Studies, Heidelberg University, Heidelberg, Germany, [2]Department of Psychology, Technical University Dortmund, Dortmund, Germany, [3]Department of Psychology, Heidelberg University, Heidelberg, Germany

The present study examined to what extent teachers' judgments of students' aptitude and students' perceived teachers' judgments explain gender differences in the early development of students' math self-concepts. A sample of $N = 519$ elementary school students was investigated at four measurement occasions from the end of third until the end of fourth grade. We assessed students' self-concepts and their perceived teachers' judgments of their aptitude in math. Teachers ($N = 27$) judged students' aptitude in math and provided students' math grades. First, we found significant gender differences in students' math self-concepts, teachers' judgments, and students' perceived teachers' judgments, but not in students' math grades. Second, structural equation models showed that teachers' judgments of students' aptitude as well as students' perceived teachers' judgments of students' aptitude longitudinally predicted students' self-concepts. Mediation analyzes demonstrated that teachers' judgments and students' perceived teachers' judgments contributed to gender differences in students' math self-concepts. Implications for future research and practice are discussed.

KEYWORDS

academic self-concept, gender differences, teachers' judgments, math, elementary school

1. Introduction

The underrepresentation of women in STEM (science, technology, engineering, and math) careers and girls in such majors in high school are persistent challenges (see Lauermann et al., 2015). Women make up only 28% of the workforce in STEM related professions (National Science Board, 2018). This underrepresentation is problematic both because women disproportionately fail to benefit from lucrative, high-status careers, and because this reduces diversity that could increase technological and scientific innovations (Hill et al., 2010). Accordingly, the question arises whether girls and women are underrepresented in the STEM sector? In addition, it would be desirable and necessary to recruit girls and women for this field in view of the shortage of skilled workers and the declining number of students in the STEM sector (Federal Statistical Office, 2023). In the current study, we focus on the math domain, because math is an important element not only for mathematicians, but in all STEM disciplines. As the roots of this underrepresentation start early (Master and Meltzoff, 2020), we focus on elementary school students in order to examine effects on the gender gap in math early during the school trajectory. Empirical evidence shows that in the early school years boys and girls do not significantly differ in their math achievement and if so, the differences were very small (e.g., Hyde et al., 1990; Herbert and Stipek, 2005; Else-Quest et al., 2010; Reilly et al., 2015; Heyder et al., 2019). Therefore, students' math achievement cannot be the initial starting point for the gender gap in STEM. Aside from students' actual achievement, it seems more likely that students' beliefs about their own competencies, namely their self-concepts, might play an important role for choosing STEM majors and careers (see Lauermann et al., 2015, 2017). Indeed, there is ample evidence for gender differences in students' math self-concepts from elementary school on, with boys showing higher self-concepts in math than girls (e.g., Else-Quest et al., 2010; Heyder et al., 2019). But how can these gender differences in students' math self-concepts be explained? There is a strong need to bring forth insights in order to understand when and why girls develop lower math self-concepts.

The development of gender differences in students' math self-concepts is complex with multiple causes. Expectancy-Value Theory (Eccles et al., 1983) assumes that previous achievement, socializers' judgments (e.g., from teachers and parents) as well as students' perceptions of these judgments have an impact among others on students' self-concepts. Various studies have showed that students' achievement cannot explain gender differences in students' self-concepts (e.g., Else-Quest et al., 2010; Heyder et al., 2019). Instead, it was shown that teachers' judgments of students' aptitude might partially explain the relationship between gender and students' self-concept in a cross-sectional study (Heyder et al., 2019). Beyond these prior studies, we not only examined teachers' judgments of students' aptitude but also students' perceived teachers' judgments as possible factors that might explain gender differences in students' math self-concepts. We also extend the literature by taking a longitudinal approach focusing on the early stage of math education. Understanding the onset of gender differences in math can help to prevent them from increasing during the school trajectory. First, we analyzed mean differences in students' math grades, teachers' judgments of students' aptitude in math, students' perceived teachers' judgments, as well as students' math self-concepts. Second, we computed a longitudinal structural equation model with teachers' judgments, as well as students' math self-concepts. Third, we computed a longitudinal structural equation model with teachers' judgments of students' aptitude and students' perceived teachers' judgments as predictors of students' math self-concepts. In this model, it was tested whether teachers' judgments and students' perceived teachers' judgments predicted students' subsequent self-concepts and whether the effects of teachers' judgments on students' self-concepts were mediated by students' perceived teachers' judgments. Finally, we checked whether gender differences in mathematical ability self-concepts were mediated by teachers' judgments and students' perceived teachers' judgments. The findings of this study contribute to a better knowledge about underlying processes of gender differences in students' math self-concepts.

2. Theoretical background

2.1. Academic self-concept: definition, development and gender differences

Academic self-concepts are defined as individuals' evaluation or self-perceptions of their competence in certain domains (Shavelson et al., 1976; Eccles et al., 1983). Academic self-concepts are typically measured via self-report and assessed by asking students' how good they think they are in a specific domain such as Math (Eccles et al., 1983). Therefore, the construct is domain-specifically operationalized and students' self-concepts vary across different subjects. For the development of academic self-concepts, researchers have commonly identified three prominent comparison processes: social, temporal, and dimensional comparisons (Wolff et al., 2018). The social comparison means that students compare their own achievement with that of their classmates, whereas the temporal comparison addresses the comparison of one's actual achievement with previous results. The dimensional comparison process refers to a comparison of one' achievement across domains (e.g., math versus verbal achievement). Due to this dimensional comparison effects, the I/E model (Marsh, 1986) assumes that math achievement negatively relates to students' verbal self-concepts, whereas verbal achievement has a negative effect on students' math self-concept.

At the beginning of primary school, students have quite high, positive self-concepts, which decrease and stabilize over the school years (Jacobs et al., 2002). During this time, students are very sensitive to social comparisons, feedback and evaluations from their teachers (Spinath and Spinath, 2005a; Natale et al., 2009), which is reflected in an adjustment of the self-concept. Because students' self-concepts develop and form in elementary school, it was important for us to examine at such an early stage what factors can predict gender differences in students' math self-concepts. This is why findings are so important and necessary, so that negative developments can be prevented at an early stage.

Empirical evidence highlights gender differences in students' math self-concepts in favor of boys. Studies found boys to have higher self-concepts in math compared to girls already in elementary school with effect sizes ranging from $d=0.36$ to $d=0.52$ (Wigfield et al., 1997; Tiedemann, 2000; Dickhäuser and Stiensmeier-Pelster, 2003; Herbert and Stipek, 2005; Chamorro-Premuzic et al., 2010; Steinmayr et al., 2019). Other studies reported correlations between students' gender and students' math self-concepts with $r=0.20$ for elementary school students (Jacobs et al., 2002). Fredricks and Eccles (2002) as well as Jacobs et al. (2002) found in their longitudinal study gender differences in students' math self-concepts both in elementary school and secondary school. The gender gap was larger in elementary school. Girls had a less steep decline in their self-concepts over time compared to boys which led to a convergence of girls' and boys' math self-concepts in secondary school (Fredricks and Eccles, 2002; Jacobs et al., 2002). Besides these primary studies, two meta-analyzes found significant differences between boys and girls in students' math self-concept with an average effect size of $d=0.25$ (Wilgenbusch and Merrell, 1999) and $d=0.33$ (Else-Quest et al., 2010). Summing up, the presented results show that boys have higher math self-concepts than girls across the school years. Given these gender differences in students' math self-concepts, it is worthwhile to look for factors that can explain why boys have significantly higher self-concepts in math compared to girls about school time. We used the Expectancy-Value-Model by Eccles et al. (1983) as theoretical basis, because this model makes assumptions about specific factors influencing students' self-concepts.

2.2. Expectancy-value theory

Expectancy-Value-Theory (EVT) by Eccles et al. (1983) is a conceptual framework for the development of achievement motivation. The model was originally designed to explain gender differences in achievement-related motivation and behavior. EVT differentiates two motivational constructs that are hypothesized to influence students' achievement and education-related choices (e.g., course plans in secondary school or career choices). These motivational constructs are expectancies and task values. In the present study, we only focus on academic self-concepts as expectancies, because large gender differences were found in this construct. EVT holds that students' self-concepts are influenced by (1) students' stable characteristics such as their aptitude and previous achievement outcomes, (2) socializers' beliefs such as teachers', parents' or peers' beliefs, and (3) students' perception of socializers' beliefs and expectations for them. Furthermore, the model holds that socializers' beliefs influence students' self-concepts indirectly through students' perceptions of socializers' beliefs. In the present study, we focused on teachers as socializers, because they represent a major environmental influence on children's development and might be especially important for the formation of students' self-concepts in elementary school as teachers are the first point of contact for feedback and information on students' own achievement and abilities (Tiedemann, 2000; Gunderson et al., 2012). Moreover, as EVT assumes effects of teachers' judgments and students' perceptions of these judgments on their self-concepts but not on their motivation such as interest or utility value, we focused on students' self-concepts as outcomes in our study.

2.3. Predicting students' math self-concepts to explain gender differences

Based on the theoretical assumptions of EVT and the I/E model, this section will present previous results of empirical studies that examined whether students' achievement, teachers' judgments of students' aptitude as socializers' beliefs as well as students' perceived teachers' judgments can predict students' math self-concepts and whether these factors play a role in explaining gender differences.

2.3.1. Students' achievement

The relationship between students' academic achievement and their self-concepts in a specific domain is well-established. Findings from countless studies showed the association between academic achievement and corresponding self-concepts to vary from moderate to highly positive correlations with $r = 0.30–0.60$ (Helmke and van Aken, 1995; Guay et al., 2003; Valentine et al., 2004; Marsh et al., 2005; Möller et al., 2009; Marsh and Martin, 2011). Furthermore, there is ample evidence that students' previous academic achievement has an impact on students' subsequent self-concepts (e.g., Helmke and van Aken, 1995; Skaalvik and Valås, 1999; Viljaranta et al., 2014). Moreover, the association between academic achievement and students' self-concepts depends on students' age. Students in elementary school often have very positive self-concepts with little variation, which leads to a weak correlation between students' self-concepts and grades. Self-concepts become more realistic in the sense of more in line with external criteria when students grow older, leading to more interindividual variation and a stronger relationship between students' self-concepts and grades (Jacobs et al., 2002; Spinath and Spinath, 2005a,b). Besides the positive correlations between performance and self-concept in the same domain, numerous studies as well as several meta-analyzes found support for the assumption of the I/E model, namely that students' verbal achievement is negatively related to students' math self-concept (Möller et al., 2009, 2020; Marsh et al., 2015). However, this research indicated that social and dimensional comparison effects of academic achievements on self-concepts are much smaller in elementary school children than secondary school students (Möller et al., 2011; Ehm et al., 2014; Lohbeck and Möller, 2017; Möller et al., 2020).

Due to these reported associations between students' achievement and self-concepts, academic achievement can be an explaining factor for gender differences in students' math self-concepts. In order to discuss whether students' achievement might explain gender differences in students' math self-concepts in more detail, the following section will shed light on gender differences in students' math achievement. Results from two meta-analyzes showed very small gender differences in math achievement in favor of boys with effect sizes $d = 0.13$ (Hyde et al., 1990) and $d = 0.07$ (Reilly et al., 2015). The authors also identified significant moderators. In one meta-analysis (Hyde et al., 1990), gender differences in students' math achievement depended on the mathematical content. Boys and girls did not differ in their achievement for algebra, arithmetic, and geometry, but only for calculus ($d = 0.20$). In the other meta-analysis (Reilly et al., 2015), students' age turned out to be a moderator of gender differences in students' math achievement. Whereas boys and girls did not differ in their achievement in elementary school, differences were found at the end of high school. This finding is in line with that of a study by Steinmayr and Spinath (2008) showing boys to outperform girls in math achievement in 11th and 12th grade ($d = 0.19$). In PISA 2018, only very small gender differences in students' math achievement were found in favor of boys with five points across all OECD states (OECD, 2019). In contrast to these findings, other studies found no significant gender differences in students' math achievement for students' grades (Marsh and Yeung, 1998; Tiedemann, 2000; Dickhäuser and Stiensmeier-Pelster, 2003; Chamorro-Premuzic et al., 2010; Wach et al., 2015) and standardized test achievement (Marsh and Yeung, 1998; Herbert and Stipek, 2005; Heyder et al., 2019). A meta-analysis by Else-Quest et al. (2010) and results from the recent Trends in Math and Science Study (TIMSS; Mullis et al., 2016) reported no significant gender differences in students' math achievement. Achievement differences between boys and girls in the verbal domain are small in elementary school, but become even greater in secondary school (Voyer and Voyer, 2014). To sum up, accumulating evidence suggests that gender differences in students' achievement are not existent or if they exist, the differences are only small. Therefore, students' academic achievement cannot serve as an explanation for gender differences in students' math self-concepts. Thus, it is worthwhile to examine other factors that can possibly explain why boys have significantly higher math self-concepts compared to girls.

2.3.2. Teachers' judgments of students' aptitude

Teachers are important socializers for students who have certain assumptions about their students' aptitude and expectations about students' future performance. In the last 50 years, many studies have examined the role of teachers' judgments in determining student outcomes. It is important to note, that teachers' judgments can be operationalized differently, e.g., as teachers' predictions of students' achievement in the near future, as teachers' estimations of students' past or current achievement or as teachers' estimations of students' current aptitude in a specific subject such as mathematics (Heyder et al., 2019). In this study we focused on teachers' judgments of students' aptitude in math. Aptitude is defined as an individual's capacity for learning and proficiency in a specific domain (Snow, 1992; Stemler and Sternberg, 2013). Therefore, students' aptitude is the potential to learn and achieve. At the same time, it is a prerequisite for achievement outcomes such as school grades, but is not always reflected by actual achievement (e.g., underachievers). Teachers can not directly observe students' aptitude in contrast to students' actual achievement. A previous study found that teachers' judgments of students' aptitude predict students' achievement development (Heyder et al., 2019). Anyhow, gender differences were not analyzed. Results of two cross-sectional studies in elementary school contexts showed small associations ($r = 0.12$ / $r = 0.26$) between teachers' judgments of students' aptitude in math and students' math self-concepts (Tiedemann, 2000; Dickhäuser and Stiensmeier-Pelster, 2003). Madon et al. (2001) found in their longitudinal study that teachers' judgments of students' aptitude at the beginning of sixth grade predicted students' self-concepts at the end of sixth grade ($ß = 0.12$). Besides those studies that examined cross-sectional and longitudinal associations between teachers' judgments of students' aptitude and students' self-concepts, one study directly investigated whether teachers' judgments explained gender differences in students' self-concepts (Heyder et al., 2019). The authors showed that teachers' judgments of students' aptitude in math mediated the relationship between gender and students' math self-concepts. In their model, teachers' judgments explained 50% variance of the relationship between gender and students' self-concepts. This was even true after controlling for actual competence, math grades, and parents' estimations of their children's math aptitude (Steinmayr et al., 2019). Nevertheless, both studies were cross-sectional and it is still unclear whether teachers' judgments of students' aptitude have longitudinal effects on students' self-concepts over a longer period of time and can therefore serve as an explaining factor of gender differences in students' self-concepts. In addition to associations between teachers' judgments and students' self-concepts, it is also important to look at gender differences in teachers' judgments of students' aptitude *per se*.

Overall, studies on gender differences in teachers' judgments of students' aptitude in math are rare and their findings are inconsistent. Whereas some study results showed that teachers judged boys and girls as equally talented in math (Dickhäuser and Stiensmeier-Pelster, 2003; Herbert and Stipek, 2005), other findings clearly demonstrated teachers to judge boys' aptitude in math to be higher compared to girls' aptitude with effect sizes of $d = 0.46$ (Heyder et al., 2019; Steinmayr et al., 2019) and $d = 0.26$ (Tiedemann, 2000). Jussim and Eccles (1992) showed an association between students' gender and teachers' judgments of students' talent in math ($ß = 0.06/0.08$).

Taken together, the reported studies showed associations between teachers' judgments of students' aptitude and students' self-concept supporting the assumption that teachers' judgments of students' aptitude have the potential to explain gender differences in students' self-concepts. However, most of the reported studies had a cross-sectional design and could not clarify whether teachers' judgments affect the development of students' self-concepts. Concerning gender differences in teachers' judgments *per se*, findings were inconsistent. Whereas teachers' judged boys to be more talented in math than girls in some studies, other studies did not find significant differences in teachers' judgments for boys and girls. Accordingly, it is worthwhile to examine gender differences in teachers' judgments of students' aptitude in math with another sample and to test whether these judgments can also explain gender differences in students' math self-concepts longitudinally.

2.3.3. Students' perceived teachers' judgments of students' aptitude

Even though EVT holds that socializers' beliefs affect students' self-concepts through students' perception of these socializers' beliefs (Eccles et al., 1983), this part of the model became little attention in educational-psychological research and literature so far. In a study with German elementary school students, students' perceived teachers' judgments significantly mediated the relationship between teachers' judgments of students' math aptitude and students' math self-concepts (Dickhäuser and Stiensmeier-Pelster, 2003). Moreover, students' perceived teachers' judgments predicted students' self-concepts in math more strongly ($ß = 0.47$) compared to students' math grades ($ß = 0.22$). These findings highlight the importance of students' perceived teachers' judgments for their self-concepts. In this study, boys showed significantly higher perceived teachers' judgments of students' math aptitude than girls ($d = 0.30$; Dickhäuser and Stiensmeier-Pelster, 2003). As amply demonstrated, boys show higher perceived teachers' judgments of students' math aptitude in elementary school. These perceived teachers' judgments are strong predictors of students' self-concepts and mediated the relationship between teachers' judgments of students' aptitude and students' self-concepts. Therefore, students' perceived teachers' judgments seem to be an important factor when explaining gender differences in students' math self-concepts. To our knowledge, the reported study is the only one that examined the role of perceived teachers' judgments for students' self-concepts. With a cross-sectional design such as in this reported study, it remains unclear whether boys and girls significantly differ in their perceived teachers' judgments across elementary school years. Furthermore, it is worthwhile to examine whether students perceived teachers' judgments mediate the relationship between teachers' judgments of students' aptitude in math and students' math self-concepts in a longitudinal design and can thus explain gender differences in students' math self-concepts.

3. Research questions and hypotheses

The present study aimed at exploring the reasons underlying the reported gender differences in students' math self-concepts. Therefore, the current study was designed to simultaneously explore whether teachers' judgments of students' aptitude and students' perceived teachers' judgments longitudinally predict students' math self-concepts. Moreover, we tested whether the effects of teachers' judgments on students' self-concepts were mediated by students' perceived teachers' judgments.

Specifically, we addressed the following research questions and derived the following hypotheses:

1. Cross-sectional: Are there gender differences in students' self-concepts and students' perceived teachers' judgments in favor of boys? We expect that boys show higher self-concepts in math and higher perceived teachers' judgments than girls, even

though there are no gender differences in students' math grades. We also expected that teachers judge boys' aptitude in math higher than girls' aptitude in math.

2. Longitudinal: Do teachers' judgments of students' aptitude longitudinally predict students' math self-concepts? We expect prior teachers' judgments of students' aptitude in math to predict students' subsequent math self-concepts.
3. Longitudinal: Do students' perceived teachers' judgments of students' aptitude longitudinally predict students' math self-concepts? We expect prior students' perceived teachers' judgments to predict students' subsequent math self-concepts.
4. Longitudinal: Are the longitudinal effects of teachers' judgments of students' aptitude on students' self-concepts mediated by students' perceived teachers' judgments? We expect that the longitudinal effects of teachers' judgments on students' self-concepts are mediated by students' perceptions of teachers' judgments.
5. Longitudinal: Do teachers' judgments of students' aptitude and students' perceived teachers' judgments mediate the relationship between gender and students' math self-concepts? We expect both teachers' judgments and students' perceived teachers' judgments to explain gender differences in students' self-concepts.

4. Method

4.1. Sample and procedure

The sample consisted of a total of $N = 519$ students (49.9% girls) and $N = 27$ teachers (100% women). Data were collected in 27 classes at 11 elementary schools in the German state of Baden-Württemberg across seven measurement occasions (Lohbeck and Möller, 2017; Weidinger et al., 2018; Heyder et al., 2019). In order to attract schools to participate in the study, the study director contacted the relevant school principals and personally introduced himself and the project. Elementary school in the north of Baden-Württemberg were approached. Comparing the present data with those from the Federal Statistical Office (2023) indicated that the participating children were representative for the population in the federal state where they came from in terms of gender ratio ($\chi^2 = 0.004$, $df = 1$, $p = 0.950$) and immigration background ($\chi^2 = 0.220$, $df = 1$, $p = 0.639$). Because teachers' judgments were assessed at only four measurement occasions, we focused on these four measurement occasions, which were spaced 4 months apart (t1: at the end of third grade; t2: at the beginning of fourth grade; t3: in the middle of fourth grade; t4: at the end of fourth grade). Students mean age was 8.28 years ($SD = 0.54$) at t1 and 9.93 years ($SD = 0.72$) at t4. Participation in the study was voluntary. Parents consented to their children's participation in the study by signing a consent form. All students answered questions about their motivation in mathematics in their classrooms on a regular school day. In order to ensure that all students worked at the same speed, a trained research assistant administered the questionnaire and read all items aloud. Overall, the assessments took about 45 min. During the same time, teachers provided information about students' grades and indicated their judgments of students' aptitude.

4.2. Measures

4.2.1. Students' gender

Students gender was coded as 1 = male and 2 = female.

4.2.2. Students' math self-concept

We assessed students' math academic self-concepts with three items: "How good are you at math?" with a response format ranging from 1 (*very good*) to 5 (*very bad*), "How easy is it for you to learn new things in math?" ranging from 1 (*very easy*) to 5 (*very hard*), and "To which group of students do you belong in your class in math?" ranging from 1 (*the best*) to 5 (*the worst*). These items were based on a questionnaire for assessing students' academic self-concept according to expectancy-value theory (Eccles and Wigfield, 1995) and established in previous studies (e.g., Brunner et al., 2008; Weidinger et al., 2018) that supported the construct validity of academic self-concept. The internal consistency of math self-concepts was good at all four measurement occasions ($\alpha = 0.84$ to 0.85). In order to present the values of this scale more intuitively, the items were recoded so that higher values stood for higher academic self-concept.

4.2.3. Teachers' judgments of students' aptitude in math

We assessed teachers' judgments of students' aptitude in math with one item, namely, "In your opinion, how talented is the following student in math?." Teachers' judgments were given on a 5-point response format ranging from 1 (*not talented*) to 5 (*very talented*). Teachers' judgments of students' aptitude were assessed with one item only, because this construct is a very narrow construct, and additional items would be phrased in the same way. Moreover, other researchers examining teachers' judgments of students' aptitude and achievement have also assessed these constructs with single-item measures (Hoge and Butcher, 1984; Tiedemann, 2000; Kuklinski and Weinstein, 2001; Dickhäuser and Stiensmeier-Pelster, 2003; Fischbach et al., 2013).

4.2.4. Students' perceived teachers' judgments

We assessed students' perceived teachers' judgments with three items, namely, "My teacher believes that I am good at math," "My teacher believes that it is easy for me to learn new things in math," and "My teacher believes that I belong to the best group of students in math in my class" with a response format ranging from 1 (*totally disagree*) to 5 (*totally agree*). The internal consistency of students' perceived teachers' judgments was good at all four measurement occasions ($\alpha = 0.80$ to 0.86).

4.2.5. Math grades

Teachers gave information about students' math grades (from their report cards). In Germany, a one stands for the best grade and six for the worst. We reversed the polarity of math grades such that higher values indicated better math achievement ranging from 1 = *insufficient/fail* (minimum) to 6 = *excellent* (maximum) so that the effects of math grades on other variables in our models can be better interpreted.

4.3. Statistical analyzes

4.3.1. Variance analyzes

In order to examine mean differences between boys and girls on students' math grades, teachers' judgments of students' aptitude, students' perceived teachers' judgments and students' math self-concepts, multivariate analysis of variance were computed for each construct at the four measurement occasions. Post-hoc comparisons were calculated to analyze differences in a construct at one specific measurement occasion. Additionally, we computed effect sizes (Cohen' s *d*) in order to interpret the magnitude of the several mean differences.

4.3.2. Structural equation models

One aim of our study was to examine whether teachers' judgments of students' aptitude and students' perceived teachers' judgments longitudinally predict students' math self-concepts. Therefore, we computed two longitudinal cross-lagged panel models with reciprocal effects between teachers' judgments of students' aptitude and students' self-concepts in the first model. The second model included reciprocal effects between students' perceived teachers' judgments and students' self-concepts. As teachers' judgments of students' aptitude were assessed with one item only, these were modeled as manifest variables. Students' perceived teachers' judgments and students' self-concepts were modeled as latent variables with three indicators each. To correct for the clustering in our data (students nested in different classes and schools) and sampling error, we used the "TYPE = COMPLEX" option in Mplus. We also included the stabilities of the several variables and intercorrelations within one measurement occasion. All structural equation models were computed using Mplus 7.11 (Muthén and Muthén, 1998/2013).

Another aim of our study was to analyze whether teachers' judgments of students' aptitude and students' perceived teachers' aptitude can explain gender differences in students' math self-concepts. Also, we wanted to test whether effects of teachers' judgments of students' aptitude on students' self-concepts are mediated through students' perceived teachers' judgments. For this, we computed a structural equation model with reciprocal effects between teachers' judgments of students' aptitude, students' perceived teachers' judgments and students' math self-concepts (see Figure 1). Also, we included effects from gender on students' math self-concepts and on the independent variables. Stabilities of the variables and intercorrelations within one measurement occasion were included as well. In order to analyze whether the effects from teachers' judgments of students' aptitude on students' self-concepts are mediated through students' perceived teachers' judgments, we included indirect effects of teachers' judgments over students' perceived teachers' judgments on students' self-concepts in our model. Also, we included indirect effects of gender over teachers' judgments and students' perceived teachers' judgments on students' self-concepts. These indirect effects were computed using the bootstrap method. Moreover, we requested confidence intervals.

On a related note, the word "effect" is used in the sense of a statistical prediction in a cross-lagged panel model (one prior variable predicts another subsequent variable) with different measurement occasions and is not meant to be causal.

4.3.3. Evaluation of model fit

For the model fit of the two structural equation models, we took into account the χ^2 value with degrees of freedom, the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA). Because the χ^2 value depends on the sample size and can easily become significant in large samples (Ullman, 2007), this value must be interpreted with caution. For the CFI, values higher than 0.95, and for the RMSEA, values lower than 0.05 are considered excellent. Such an excellent fit occurs when the model fits the data well. If the values for the CFI are lower than 0.90 and the values for RMSEA are higher than 0.10, then the model is not acceptable (Hu and Bentler, 1999; Marsh et al., 2004).

4.3.4. Handling missing data

As in every longitudinal study, we had to deal with missing data in the present study. The main reason for missing data was illness, whereby some students missed one or more measurement occasions. These missing data ranged from 11.4% (t1) to 19.5% (t3) for students' math grades, from 11.4% (t2) to 24.7% (t1) for teachers' judgments of students' aptitude, from 9.8% (t1) to 14.1% (t4) for academic self-concept, and from 9.6% (t1) to 13.7% (t3, t4) for students' perceived teachers' judgments. Students with missing data did not differ from students without missing data. To handle missing data in Mplus, we used the full information maximum likelihood (FIML) method. This is an approach that typically yields less biased estimates under the missing at random assumption than traditional approaches such as listwise or pairwise deletion. Also, this method takes all information into account (i.e., cases with missing values) when estimating the model parameters and maintains statistical power at the same time (Schafer and Graham, 2002; Enders, 2010).

5. Results

5.1. Descriptive statistics and gender differences

Table 1 shows the means (M), standard deviations (SD), and internal consistencies (α) for the examined variables across our four measurement occasions for the whole sample as well as for girls and boys separated. The results from MANOVA showed no significant gender effect for math grades [$F(4,340) = 1.95$, $p = 0.12$, $d = 0.3$]. Post-hoc univariate analyzes showed no gender differences at each measurement occasion. Even there were no significant mean differences in students' math grades, teachers rated students' math aptitude significantly higher for boys [$F(4,306) = 5.36$, $p < 0.001$, $d = 0.54$]. This pattern was found for each measurement occasion. According to this, boys showed significantly higher perceived teachers' judgments of their aptitude in math compared to girls [$F(4,366) = 6.12$, $p < 0.001$, $d = 0.52$]. Post-hoc univariate analyzes confirmed this for each measurement occasion. Also, boys showed significantly higher math self-concepts than girls among all measurement occasions [$F(4,365) = 9.79$, $p < 0.001$, $d = 0.66$]. The differences in teachers' judgments for boy s and girls can be described as small effects (Cohen's d). The effect sizes for the significant differences between boys in girls in their perceived teachers' judgments of their aptitude and in their math self-concept varied from small to medium. Intercorrelations between all examined variables can be seen in Table 2.

5.2. Longitudinal effects from teachers' judgments and students' perceived teachers' judgments on students' self-concepts

In order to analyze longitudinal effects between teachers' judgments of students' aptitude and students' math self-concept

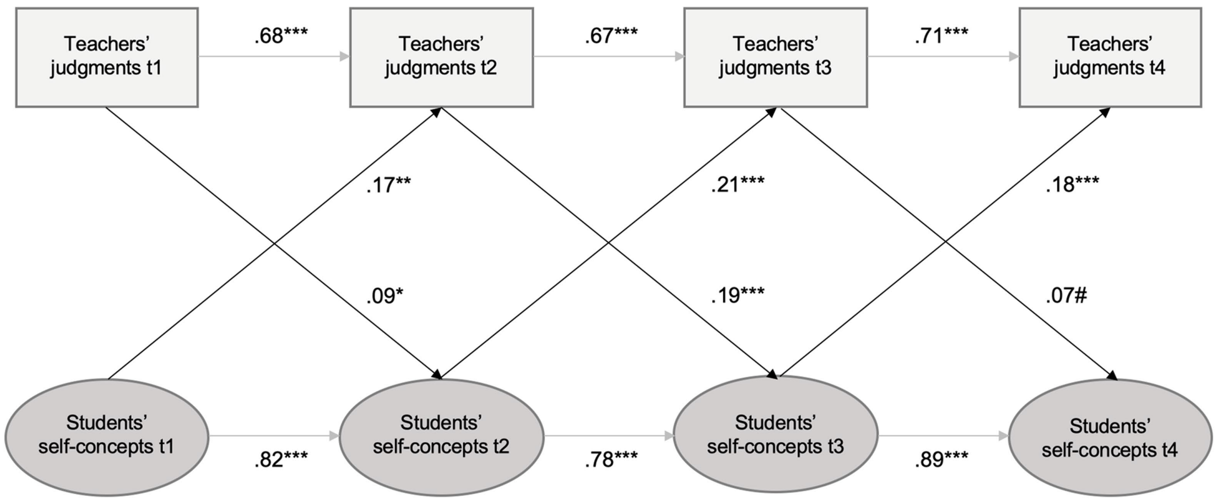


FIGURE 1
Cross-lagged panel model with teachers' judgments of students' aptitude in math and their math-specific self-concepts. For greater clarity, indicators of latent constructs (items), method factors and intercorrelations to each measurement occasion are not depicted in this figure. *** $p < 0.001$; ** $p < 0.01$, * $p < 0.05$, # $p < 0.10$.

as well as between students' perceived teachers' judgments and students' math self-concept, two structural equation models with cross-lagged effects were computed. The first model with teachers' judgments of students' aptitude and students' self-concepts (see Table 3; Figure 1) fitted the data well [$\chi^2(81) = 123.51$, $p < 0.01$; CFI = 0.986; TLI = 0.980; RMSEA = 0.040; SRMR = 0.029]. The high correlations within a construct at different measurement occasions, suggests a relatively high stability of the constructs over the elementary school years. On the one hand, we found significant effects from teachers' prior judgments of students' aptitude on students' subsequent self-concepts (from t1 to t2 $ß = 0.10$, $p < 0.05$; from t2 to t3 $ß = 0.19$, $p < 0.001$). This indicates that teachers' prior judgments of students' aptitude longitudinally predict students' subsequent self-concepts in elementary school. The effect from teachers' judgments to t3 on students' self-concepts to t4 was not significant. On the other hand, we found significant effects from prior self-concepts on subsequent teachers' judgments to all measurement occasions ($ß = 0.17–0.21$, $p < 0.001$). These results provide support that not only teachers' judgments predict students' self-concepts, but also that students' prior self-concepts predict subsequent teachers' judgments of students' aptitude.

The second model with students' perceived teachers' judgments of their aptitude and students' self-concepts (see Table 4; Figure 2) also fitted the data well [$\chi^2(231) = 565.89$, $p < 0.001$; CFI = 0.947; TLI = 0.937; RMSEA = 0.066; SRMR = 0.040]. Again, the constructs were relatively stable over time. In this model, we found one significant path from students' perceived teachers' judgments of their aptitude to t3 on students' self-concepts to t4 ($ß = 0.27$, $p < 0.05$). Also, we found one significant path from students' self-concepts to t2 on students' perceived teachers' judgments of their aptitude to t3 ($ß = 0.34$, $p < 0.05$). We found little support for our hypothesis that students' perceived teachers' judgments of students' aptitude longitudinally predict students' math self-concepts.

5.3. Teachers' judgments and students' perceived teachers' judgments as explaining factors of gender differences in students' academic self-concepts

In order to examine whether teachers' judgments of students' aptitude and students' perceived teachers' judgments of students' aptitude can explain gender differences in students' math self-concepts we computed a longitudinal structural equation model with these variables among all four measurement occasions (see Table 5; Figure 3). The model fit was acceptable with [$\chi^2(327) = 723.70$, $p < 0.001$; CFI = 0.947; TLI = 0.934; RMSEA = 0.059; SRMR = 0.039]. First, we found one significant direct effect from prior teachers' judgments of students' aptitude at t2 on students' subsequent self-concepts at t3 ($ß = 0.17$, $p < 0.01$). Second, there were two significant direct effects from prior teachers' judgments of students' aptitude on students' subsequent perceived teachers' judgments of students' aptitude (from t2 to t3 $ß = 0.19$, $p < 0.001$; from t3 to t4 $ß = 0.14$, $p < 0.05$). Third, we found one significant direct effect from students' perceived teachers' judgments of students' aptitude at t3 on students' subsequent self-concepts at t4 ($ß = 0.26$, $p < 0.05$). In addition to the direct effects of teachers' judgments of students' aptitude and students' perceived teachers' judgments on students' self-concepts, we examined whether these effects are mediated through students' perceived teachers' judgments. We found that teachers' judgments of students' aptitude (t2) had an indirect effect on students' self-concepts (t4) ($ß = 0.11$, $p < 0.01$). But contradictory to our hypothesis, this indirect effect was not mediated through students' perceived teachers' judgments of students' aptitude, but through teachers' judgments of students' aptitude at t3. Finally, we examined the effects of gender on students' self-concepts to all four measurement occasions. We found gender to be a significant predictor at the first measurement occasion ($ß = -0.31$, $p < 0.001$) and at the fourth measurement occasion ($ß = -0.12$, $p < 0.001$). Students' self-concepts at t2, and t3 were not

TABLE 1 Means (M), standard deviations (SD), and internal consistencies (α) for the whole sample as well as for boys and girls separated.

	Overall Sample		Boys		Girls	
	M (SD)	α	*M* (SD)	α	*M* (SD)	α
Math grades t1	4.62 (1.00)	-	4.62 (0.97)	-	4.62 (1.02)	-
Math grades t2	4.66 (0.92)	-	4.72 (0.90)	-	4.59 (0.94)	-
Math grades t3	4.67 (0.90)	-	4.75 (0.90)	-	4.59 (0.89)	-
Math grades t4	4.65 (0.87)	-	4.68 (0.83)	-	4.63 (0.91)	-
Teachers' judgments t1	3.61 (1.01)	-	3.71 (1.00)	-	3.50 (1.01)	-
Teachers' judgments t2	3.68 (0.99)	-	3.80 (1.01)	-	3.56 (0.95)	-
Teachers' judgments t3	3.72 (0.96)	-	3.80 (0.97)	-	3.63 (0.95)	-
Teachers' judgments t4	3.77 (0.97)	-	3.89 (0.96)	-	3.65 (0.96)	-
Perceived teachers' judgments t1	3.52 (0.93)	0.81	3.76 (0.86)	0.74	3.29 (0.94)	0.85
Perceived teachers' judgments t2	3.59 (0.86)	0.80	3.74 (0.83)	0.77	3.44 (0.86)	0.80
Perceived teachers' judgments t3	3.61 (0.87)	0.85	3.74 (0.83)	0.82	3.48 (0.89)	0.86
Perceived teachers' judgments t4	3.56 (0.85)	0.85	3.72 (0.80)	0.83	3.40 (0.87)	0.86
Math self-concept t1	3.80 (0.82)	0.84	3.98 (0.79)	0.83	3.63 (0.82)	0.84
Math self-concept t2	3.92 (0.76)	0.84	4.05 (0.73)	0.81	3.79 (0.77)	0.86
Math self-concept t3	3.84 (0.78)	0.85	3.91 (0.74)	0.81	3.77 (0.81)	0.88
Math self-concept t4	3.78 (0.74)	0.85	3.96 (0.66)	0.81	3.60 (0.78)	0.87

significantly predicted by gender. Moreover, gender predicted teachers' judgments ($ß = -0.11$, $p < 0.05$) as well as students' perceived teachers' judgments at the first measurement occasion ($ß = -0.33$, $p < 0.001$) but was not a significant predictor at the other measurement occasions. The results of the indirect effects showed that the effect of gender on students' self-concepts at t2 was significantly mediated by students' self-concepts at t1. For the effect of gender on students' self-concepts at t3, indirect effects could be found via students' self-concepts at t2 and teachers' judgments at t2. For the effect of gender on students' self-concepts at t4, indirect effects could be found via prior self-concepts, teachers' judgments as well as students' perceived teachers' judgments. These results suggest that gender differences in students' self-concepts can be explained by prior teachers' judgments of students' aptitude, students' perceived teachers' judgments as well as by prior self-concepts.

6. Discussion

The present study was designed to contribute to a better understanding of the reasons behind the underrepresentation of women in STEM domains and girls in STEM-related majors. Therefore, we focused on students' self-concept as one predictor of education-related choices. A primary purpose of this study was to analyze gender differences in students' math self-concepts, their math grades, teachers' judgments of students' aptitude in math and students' perceptions of teachers' judgments. A second and the central purpose of this research was to explain gender differences in students' math self-concepts. Both teachers' judgments and students' perceived teachers' judgments were tested as possible explaining factors. Accordingly, a first point of discussion will refer to gender differences in the examined variables. Afterwards, we will discuss whether teachers' judgments and students' perceptions of these judgments can explain gender differences in students' math self-concepts. Finally, we will discuss limitations of our study, ideas for future research and practical implications.

6.1. Gender differences in the examined variables

In line with our hypothesis, we found significant gender differences in students' math self-concepts in favor of boys at all measurement occasions. These differences can be described as small to medium effect sizes. These empirical findings are consistent with previous reports showing boys to have higher math self-concepts in elementary school than girls (Wigfield et al., 1997; Tiedemann, 2000; Dickhäuser and Stiensmeier-Pelster, 2003; Herbert and Stipek, 2005; Chamorro-Premuzic et al., 2010; Heyder et al., 2019).

Moreover, we found no significant gender differences in students' math grades, which is also in line with previous findings indicating that boys and girls do not differ in their math achievement (e.g., Tiedemann, 2000; Dickhäuser and Stiensmeier-Pelster, 2003; Else-Quest et al., 2010; Heyder et al., 2019). This indicates that gender differences in students' self-concepts cannot be traced back to gender differences in students' achievement. Consequently, there must be other factors influencing students'

TABLE 2 Intercorrelations among all variables.

	MG t2	MG t3	MG t4	TJ t1	TJ t2	TJ t3	TJ t4	PTJ t1	PTJ t2	PTJ t3	PTJ t4	MSC t1	MSC t2	MSC t3	MSC t4
Math Grade (MG) t1	0.76	0.80	0.79	0.76	0.69	0.71	0.71	0.38	0.41	0.48	0.51	0.52	0.42	0.55	0.56
Math Grade (MG) t2		0.87	0.81	0.72	0.82	0.79	0.75	0.39	0.48	0.53	0.59	0.54	0.50	0.58	0.61
Math Grade (MG) t3			0.87	0.73	0.77	0.82	0.80	0.41	0.46	0.56	0.57	0.53	0.47	0.61	0.61
Math Grade (MG) t4				0.68	0.71	0.76	0.79	0.37	0.41	0.53	0.57	0.49	0.43	0.57	0.60
Teachers' judgments (TJ) t1					0.77	0.74	0.71	0.42	0.42	0.50	0.49	0.49	0.43	0.53	0.51
Teachers' judgments (TJ) t2						0.79	0.77	0.42	0.44	0.54	0.59	0.52	0.47	0.56	0.59
Teachers' judgments (TJ) t3							0.82	0.42	0.45	0.57	0.58	0.54	0.50	0.59	0.61
Teachers' judgments (TJ) t4								0.42	0.39	0.54	0.57	0.52	0.44	0.55	0.58
Perceived TJ (PTJ) t1									0.67	0.62	0.62	0.72	0.61	0.59	0.58
Perceived TJ (PTJ) t2										0.72	0.67	0.63	0.77	0.69	0.67
Perceived TJ (PTJ) t3											0.79	0.66	0.72	0.82	0.77
Perceived TJ (PTJ) t4												0.65	0.68	0.75	0.83
Math self-concept (MSC) t1													0.74	0.72	0.71
Math self-concept (MSC) t2														0.75	0.72
Math self-concept (MSC) t3															0.81
Math self-concept (MSC) t4															

All correlations were significant with $p < 0.001$.

self-concepts in a way that boys develop higher self-concepts in math compared to girls.

Although boys and girls in our sample did not differ in their math achievement, teachers judged boys to be significantly more talented in math than girls. At all measurement occasions, teachers' judgments of students' aptitude were significantly higher for boys than for girls. These results are consistent with findings from previous studies that also showed significant differences in teachers' judgments of students' aptitude in math, but no significant gender differences in students' math achievement (Jussim and Eccles, 1992; Tiedemann, 2000; Heyder et al., 2019). The present results, in conjunction with those presented by other researchers, clearly indicate that teachers are subject to a math-male stereotype when they have to judge students' aptitude in math. Based on the fact that teachers judged boys to be more talented in math, it is not surprising that boys showed significantly higher perceived teachers' judgments of students' aptitude in math than girls. These differences were significant at all four measurement occasions. To our knowledge, only one previous study examined gender differences in students' perceived teachers' judgments. This study was cross-sectional and also showed boys to have higher perceived teachers' judgments of students' aptitude in math than girls (Dickhäuser and Stiensmeier-Pelster, 2003). Therefore, students' perception of teachers' judgments might also be an important variable for the explanation of gender differences in students' math self-concepts.

In sum, our findings show consistent gender differences in students' math self-concepts in elementary school already. Moreover, our study provided empirical support for gender differences in teachers' judgments of students' aptitude and students' perceptions of teachers' judgments without having differences in students' achievement. These differences might explain why boys develop higher self-concepts in math during elementary school than girls. The next section will shed light on, whether teachers' judgments and students' perceptions of these judgments predict students' math self-concepts to explain gender differences.

6.2. Teachers' judgments and students' perceived teachers' judgments as explaining factors of gender differences in students' academic self-concepts

In our first longitudinal model, we found significant effects of prior teachers' judgments on students' subsequent self-concepts in math. These findings are in line with the assumption of EVT that socializers' beliefs predict students' self-concepts (Eccles et al., 1983). These effects are also consistent with findings from cross-sectional studies demonstrating significant small associations between teachers' judgments of students' aptitude and students' self-concepts

TABLE 3 Standardized stability and cross-lagged effects with standard errors and confidence intervals for teachers' judgments of students' aptitude in math (TJ), and students' math self-concepts (MSC).

	Stability effects		Cross-lagged effects	
Waves	TJ	MSC	TJ → MSC	MSC → TJ
t 1–2	0.68*** (0.04) [0.60; 0.77]	0.82*** (0.05) [0.72; 0.91]	0.09* (0.05) [0.01; 0.19]	0.17** (0.05) [0.06; 0.27]
t 2–3	0.67*** (0.04) [0.61; 0.74]	0.78*** (0.04) [0.70; 0.86]	0.19*** (0.04) [0.11; 0.28]	0.21*** (0.04) [0.13; 0.29]
t 3–4	0.71*** (0.05) [0.60; 0.81]	0.89*** (0.04) [0.81; 0.96]	0.07# (0.05) [−0.02; 0.16]	0.18*** (0.05) [0.07; 0.28]

TJ, Teachers' judgments of students' aptitude in math; MSC, Math self-concept; $***p<0.001$; $**p<0.01$, $\#p<0.10$.

TABLE 4 Standardized stability and cross-lagged effects with standard errors and confidence intervals for students' perceived teachers' judgments of students' aptitude in math (PTJ), and students' math self-concepts (MSC).

	Stability effects		Cross-lagged effects	
Waves	PTJ	MSC	PTJ → MSC	MSC → PTJ
t 1–2	0.69*** (0.13) [0.44; 0.90]	0.88*** (0.11) [0.65; 0.99]	0.01 (0.12). [−0.24; 0.24]	0.19 (0.13) [−0.07; 0.45]
t 2–3	0.57*** (0.15) [0.27; 0.81]	0.82*** (0.14) [0.54; 0.99]	0.08 (0.15) [−0.21; 0.37]	0.34* (0.15) [0.04; 0.64]
t 3–4	0.88*** (0.19) [0.51; 0.99]	0.68*** (0.15) [0.39; 0.93]	0.27* (0.15) [0.01; 0.52]	0.05 (0.19) [−0.33; 0.42]

PTJ, Students' perceived teachers' judgments of students' aptitude in math; MSC, Math self-concept; $***p<0.001$; $*p<0.05$.

(Tiedemann, 2000; Madon et al., 2001; Dickhäuser and Stiensmeier-Pelster, 2003; Heyder et al., 2019).

Moreover, the results of the second longitudinal model showed one significant effect of students' prior perceived teachers' judgments of students' aptitude on students' subsequent self-concepts (t3 to t4). This finding provides little support for the assumption of EVT, namely, that perceived teachers' judgments as perceptions of socializers' beliefs predict students' self-concepts (Eccles et al., 1983). Whereas only one study reported a cross-sectional association between students' perceived teachers' judgments and students' self-concepts (Dickhäuser and Stiensmeier-Pelster, 2003), this study showed one longitudinal effect. Nevertheless, it is unclear, whether only one of three possible effects of students' perceived teachers' judgments on students' self-concepts became significant. One explanation might be that students' self-concepts were highly stable over time in elementary school and it might be hard for other variables to explain further variance. At this point, our results indicate that both teachers' judgments of students' aptitude in math and students' perceptions of these judgments have direct effects on students' math self-concepts over time.

In our overall model, we again found evidence for direct effects of teachers' prior judgments of students' aptitude and students' perceived teachers' judgments on students' subsequent self-concepts. Based on these findings, it seems highly likely that students develop their self-concepts based on teachers' judgments. Furthermore, we found significant direct effects of teachers' prior judgments on students' perceptions of teachers' judgments. Teachers seem to communicate their judgments to their students in different ways such as through feedback and attributional processes as well as through emotional responses and classroom practices (Stipek et al., 2001; Georgiou et al., 2002; Upadyaya and Eccles, 2015). For example, when a teacher has high expectations for a boy in math, the teacher will challenge and encourage him more in class, for example, giving him more challenging tasks and praising him for the corresponding performance. Subsequently, the boy will transfer this information into his perceived teachers' judgments and assumes that the teacher thinks he is gifted. In contrast, if a teacher assumes a girl to have a lower math aptitude, he or she will challenge and encourage her less in class by providing less challenging tasks and attributing good performance to effort, for example. As a result, the girl will perceive a lower judgment of her math aptitude.

Another interesting finding concerns the mediation effect of teachers' judgments on students' self-concepts through students' perceived teachers' judgments. So, teachers' judgments had an indirect effect on students' self-concepts, but this effect was not mediated through students' perceived teachers' judgments but through teachers' judgments one measurement occasion later. This suggests that the effects of teachers' judgments on students' self-concepts seem to be more directly than mediated through students' perceptions of their teachers' judgments. This finding highlights the importance of teachers' beliefs about students' aptitude for students' self-concepts. In other words, what the teacher thinks about a student's aptitude is an important factor when students develop their self-concepts.

One final result we want to put attention to is that effects of students' gender on students' self-concepts in math completely disappeared at t2 and t3 and partially disappeared at t4 after including teachers' judgments and students' perceived teachers' judgments as predictors of students' self-concepts. Furthermore, we found indirect effects of gender on students' self-concepts via teachers' judgments and students' perceived teachers' judgments. These findings emphasize that teachers' judgments of students' aptitude and the perception of these by the students can explain gender differences in students' self-concepts in math. Our study goes beyond the finding that teachers' judgments explained half of the variance of the relationship between gender and self-concepts (Heyder et al., 2019), because of our longitudinal design with four measurement occasions. Students seem to develop their self-concept based on messages they receive from important socializers such as teachers. Our study provided empirical support for the idea that students in elementary school seem to have internalized their teachers' judgments with a math-male stereotype resulting in lower math self-concepts for girls even though girls and boys do not differ in their math achievement. One can assume that the effects of teachers' judgments on students' perceived teachers'

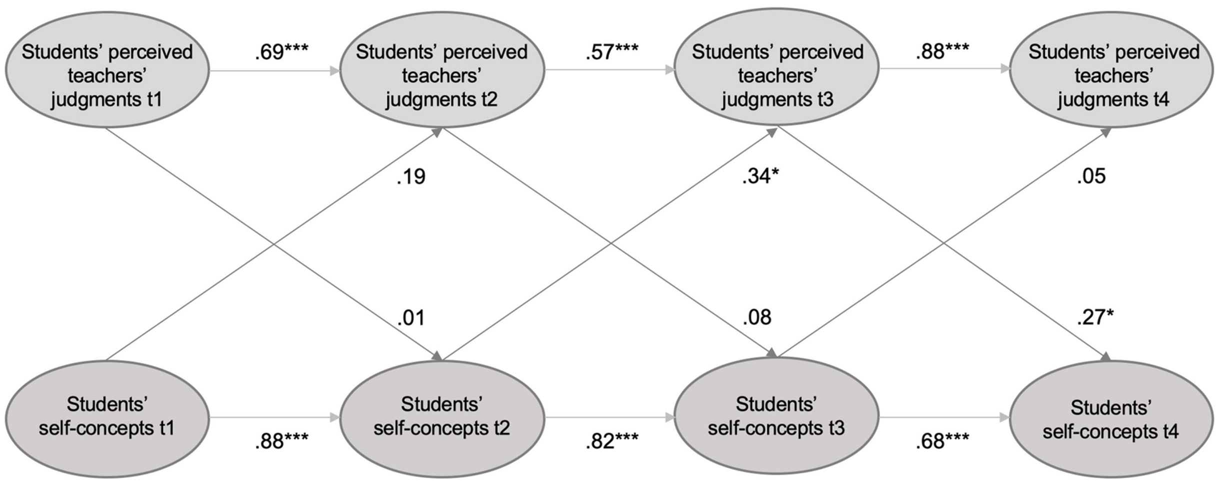


FIGURE 2
Cross-lagged panel model with students' perceived teachers' judgments and their math-specific self-concepts. For greater clarity, indicators of latent constructs (items), method factors and intercorrelations to each measurement occasion are not depicted in this figure. *** $p < 0.001$; * $p < 0.05$.

TABLE 5 Standardized stability and cross-lagged effects with standard errors and confidence intervals for teachers' judgments of students' aptitude in math (TJ), students' perceived teachers' judgments (PTJ), and students' math self-concepts (MSC).

	Stability effects			Cross-lagged effects					
Waves	TJ	PTJ	MSC	TJ → PTJ	PTJ → TJ	TJ → MSC	MSC → TJ	PTJ → MSC	MSC → PTJ
t 1–2	0.68*** (0.04) [0.59; 0.76]	0.66*** (0.20) [0.32; 0.83]	0.79*** (0.19) [0.56; 0.98]	0.06 (0.06) [−0.06; 0.18]	0.03 (0.12) [−0.28; 0.19]	0.10# (0.06) [−0.01; 0.22]	0.18 (0.12) [−0.05; 0.42]	0.03 (0.19) [−0.21; 0.27]	0.19 (0.21) [−0.24; 0.52]
t 2–3	0.67*** (0.08) [0.55; 0.78]	0.52*** (0.37) [0.16; 0.72]	0.77*** (0.49) [0.38; 0.99]	0.19*** (0.06) [0.09; 0.32]	0.32# (0.19) [−0.01; 0.59]	0.17** (0.06) [0.05; 0.27]	0.52*** (0.19) [0.19; 0.86]	0.05 (0.50) [−0.14; 0.37]	0.27* (0.38) [0.05; 0.56]
t 3–4	0.72*** (0.14) [0.51; 0.87]	0.92*** (0.65) [0.32; 0.98]	0.61*** (0.45) [0.35; 0.95]	0.14* (0.08) [0.01; 0.29]	0.28 (0.21) [−0.14; 0.67]	0.08# (0.05) [−0.01; 0.18]	0.10 (0.21) [−0.58; 0.24]	0.26* (0.44) [0.01; 0.52]	0.09 (0.66) [−0.24; 0.45]
Effects of gender on students' math self-concepts				Effects of gender on teachers' judgments			Effects of gender on teachers' judgments		
gender → MSC t1	−0.31*** (0.05) [−0.41; −0.20]			gender → TJ t1	−0.11* (0.06) [−0.20; −0.01]		gender → PTJ t1	−0.33*** (0.05) [−0.42; −0.25]	
gender → MSC t2	0.03 (0.04) [−0.06; 0.11]			gender → TJ t2	−0.05 (0.04) [−0.12; 0.02]		gender → PTJ t2	0.01 (0.05) [−0.06; 0.09]	
gender → MSC t3	0.03 (0.04) [−0.05; 0.11]			gender → TJ t3	−0.04 (0.05) [−0.11; 0.05]		gender → PTJ t3	0.01 (0.05) [−0.07; 0.07]	
gender → MSC t4	−0.12*** (0.04) [−0.18; −0.06]			gender → TJ t4	−0.03 (0.09) [−0.10; 0.08]		gender → PTJ t4	−0.01 (0.06) [−0.09; 0.07]	

TJ, Teachers' judgments of students' aptitude in math; PTJ, Students' perceived teachers' judgments; MSC, Math self-concept; *** $p < 0.001$; ** $p < 0.01$, * $p < 0.05$, # $p < 0.10$.

judgments and students' self-concepts are in line with self-fulfilling prophecy effects (Brophy and Good, 1970). It can be argued that based on their judgment or belief about students' aptitude in math, teachers treat boys and girls differently in class such as providing more support, more challenging tasks and better feedback to students with higher judged aptitudes. As argued above, students perceive their teachers' judgments based on various teacher expressions and behaviors. These student perceptions, in turn, have effects on self-concepts. This is a worrying finding because students' self-concepts themselves have effects on students' achievement as well as their education-related choices such as majors and careers in turn. Consequently, it is not surprising that women or girls are underrepresented in STEM domains.

Gender differences in students' math self-concepts were not fully explained to our last measurement time point at the end of grade 4. This suggests that there must be other variables besides teachers' judgments and students' perceived teachers' judgments that can explain gender differences in students' math self-concepts. At this time, students are quite before their transfer to secondary school. Maybe beliefs from other important socializers such as their parents

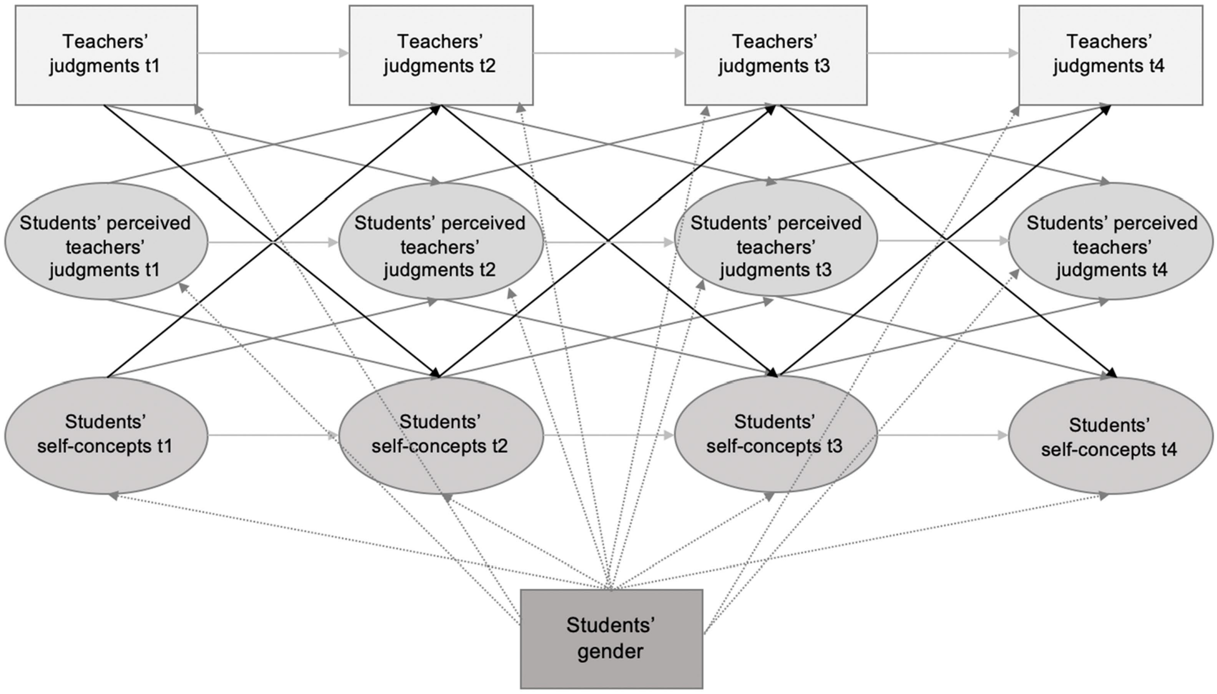


FIGURE 3
Cross-lagged panel model with teachers' judgments of students' aptitude in math, students' perceived teachers' judgments and their math self-concepts. For greater clarity, indicators of latent constructs (items), method factors and intercorrelations to each measurement occasion are not depicted in this figure.

or peers become more important and have effects on students' self-concepts as well. Previous research has shown that, in addition to teachers, parents also have gender-stereotypic perceptions of their children's aptitude. In a study from Herbert and Stipek (2005), parents rated boys' math competencies higher than girls in both third and fifth grades. Moreover, parent ratings of children's competence were a particularly strong predictor of children's judgments of their math skills. One difference between the study mentioned above and our study, however, is that the parents were asked to assess the competencies that the children already have. In our study, we focused on the assessment of students' aptitude in the sense of an underlying potential that does not necessarily translate into achievement. In addition to teachers' and parents' judgments, the judgments of peers also seem to play a role for the development of students' self-concepts. For example, Lorenz et al. (2020) showed with social network analyzes that ninth-grade students adapt their expectations toward the average expectations of their friends. The fact that we did not include parents' and peers' judgments in our study, will also be addressed in the limitation section.

Our results point to the importance of teachers' judgments of students' aptitude and students' perceived teachers' judgments for the development and explanation of gender differences in students' math self-concepts during their time in elementary school.

6.3. Limitations and future research

The current study has a longitudinal design with four measurement occasions in order to examine whether teachers' judgments of students' aptitude and students' perceived teachers' judgments predict students' math self-concepts over a longer period of time in elementary school. Even though we found effects from prior teachers' judgments and students' perceived teachers' judgments on students' self-concepts, these effects cannot be interpreted as causal effects. It is possible that a third variable has an influence on our examined variables. To clarify whether teachers' judgments and students' perceptions of these judgments have a causal influence on students' self-concepts, future studies should compute an experiment, in which students get simulated feedback from teachers.

As reported above, we focused on the math domain as one of the STEM domains, because girls are underrepresented in these majors and careers. Indeed, we found that girls showed significantly lower math self-concepts compared to boys. Moreover, teachers' rated girls' aptitude in math as significantly lower as boys' aptitude in math, although girls and boys did not significantly differ in their math achievement. This indicates that teachers might be biased by math-male stereotypes. Accordingly, it would also be interesting to examine gender differences in teachers' judgments and students' self-concepts in a domain, in which girls might be better such as languages.

It is unknown why teachers judged boys' aptitude in math higher than girls' aptitude in math even though they did not differ in their math achievement. It may help to explore and understand the underlying mechanisms. One possible explanation might be that teachers attribute boys' achievement to their high aptitude and girls' achievement to their high effort. In a study by Fennema et al. (1990) teachers more often named boys to be the best students in class and attributed boys' success to aptitude and girls' success to effort. Furthermore, Tiedemann (2000) showed that teachers attributed boys' failures in math to a lack of effort, but girls' failures to a lack of aptitude. Future studies should systematically examine teachers' judgments of students' aptitude in math, teachers' attributions as well as students' achievement in order to analyze gender differences in the math domain in more detail. In addition, future research is needed to explore possible opportunities to reduce this math-male stereotype, which will be discussed in the section to practical implications in detail.

In our study we focused on elementary school, because during this time students use feedback from different sources such as achievement in exams and direct feedback from their teachers in order to develop a more realistic self-concept. Teachers' judgments can have far-reaching consequences not only for the development of students' self-concepts but also for recommendation for secondary school. Nevertheless, future research could shed light on whether teachers' judgments have significant effects on students' subsequent self-concepts in secondary school and to test whether teachers' judgments become less important over time.

One shortcoming of the present study is that all teachers in our sample were female. It is interesting that the gender differences in teachers' judgments were highly significant even though boys and girls did not differ in their achievement. Maybe female teachers have internalized the math-male stereotype to which they were exposed in former times. Therefore, the question arises whether the gender gap in teachers' judgments might depend on teachers' gender. Robinson-Cimpian et al. (2014) found the gender difference in teachers' judgments of students' ability in math to be larger for female than for male teachers. From this perspective, future research should systematically examine whether gender differences in teachers' judgments depend on teachers' gender and why female teachers might be more math-male stereotyped. The mechanisms behind that phenomenon should be explored in more detail.

Moreover, we focused on teachers' judgments as socializers' beliefs, because teachers play an important role in students' development. The results of our study suggested that gender differences in students' math self-concepts at t4 could not be fully explained by teachers' judgments and students' perceptions of teachers' judgments. Consequently, there must be further variables that might have an influence on students' math self-concepts at the end of elementary school. Prior research showed that parents' math-gender stereotypes can affect their perceptions of their own children's math ability even in early elementary school (Gunderson et al., 2012). Furthermore, Wolff (2021) found in his recent study that gender stereotypes shared by students' classmates can have a substantial impact on students' math self-concepts. Therefore, it would be worthwhile to examine different socializers' beliefs such as from teachers, parents and peers in one study in order to examine their relative importance for students' self-concepts in order to explain these gender differences.

6.4. Practical implications

Beyond our described implications for further research, there are some important practical implications of these findings that should be mentioned in the following. In our study, teachers judged girls' aptitude in math significantly lower than boys' aptitude although girls and boys did not differ in their actual achievement. The underestimation of girls' aptitude in math might be biased by a math-male stereotype and not based on objective criteria. This prevalence of math-male stereotypes can have far-reaching negative consequences for girls' motivation (development), their achievement (in form of self-enhancement effects) and their educational choices such as majors in high school, study programs and careers. From this perspective, interventions are needed to reduce these biased judgments or these lower judgments for girls, respectively. One possibility for an intervention might be to make teachers aware of this math-male stereotype and biased judgments. Another possibility might be to provide teachers with empirical evidence showing that girls and boys do not differ in their math achievement. Teachers should also consciously make sure that they support boys and girls with the same performance in math equally. Heyder et al. (2019) proposed that these interventions can be integrated in the teacher education at university or as additional training for in-service teachers. In addition to prevention and intervention measures for teachers, offers for students should also be established. For example, girls could benefit from a workshop in which they learn to perceive their aptitude for mathematics independently of the assessments of different peers. This workshop was also intended to present female role models in the STEM sector and to make it clear that girls and women can be just as successful as boys and men. Reducing this gender bias might lead to a higher decision of girls to select a STEM career.

7. Conclusion

In conclusion, the current study provides new insights into how gender differences in students' math self-concepts can be explained. Both teachers' judgments of students' aptitude in math and students' perceived teachers' judgments had longitudinal effects on students' math self-concepts and can therefore explain why boys show higher math self-concepts than girls during the time of elementary school.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

BS and RS have provided the data set and supervised the project and supplemented individual points in the manuscript and reviewed the manuscript. KR analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Brophy, J. E., and Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: some behavioral data. *J. Educ. Psychol.* 61, 365–374. doi: 10.1037/h0029908

Brunner, M., Krauss, S., and Kunter, M. (2008). Sex differences in mathematics: does the story need to be rewritten? *Intelligence* 36, 403–421. doi: 10.1016/j.intell.2007.11.002

Chamorro-Premuzic, T., Harlaar, N., Greven, C. U., and Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence* 38, 385–392. doi: 10.1016/j.intell.2010.05.002

Dickhäuser, O., and Stiensmeier-Pelster, J. (2003). Wahrgenommene lehrereinschätzungen und das fähigkeitsselbstkonzept von jungen und mädchen in der grundschule. [Perceptions of teacher evaluations and the academic self-concept of boys and girls in elementary school]. *Psychol. Erzieh. Unterr.* 50, 182–190.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). "Expectancies, values, and academic behaviors" in *Achievement and achievement motivation*. ed. J. T. Spence (San Francisco, CA: W. H. Freeman), 75–146.

Eccles, J. S., and Wigfield, A. (1995). In the mind of the actor: the structure of adolescents' task values and expectancy-related beliefs. *Personal. Soc. Psychol. Bull.* 21, 215–225. doi: 10.1177/0146167295213003

Ehm, J. -H., Lindberg, S., and Hasselhorn, M. (2014). Reading, writing, and math self-concept in elementary school children: influence of dimensional comparison processes. *Eur. J. Psychol. Educ.* 29, 277–294. doi: 10.1007/s10212-013-0198-x

Else-Quest, N. M., Hyde, J. S., and Linn, M. C. (2010). Cross-national patterns of sex differences in mathematics: A meta-analysis. *Psychol. Bull.* 136, 103–127. doi: 10.1037/a0018053

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Federal Statistical Office. (2023). *Statistisches Bundesamt. Pressemitteilung N004 vom 23*. Available at: https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/01/PD23_N004_213.html. (Accessed January 2023).

Fennema, E., Peterson, P. L., Carpenter, T. P., and Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. *Educ. Stud. Math.* 21, 55–69. doi: 10.1007/BF00311015

Fischbach, A., Baudson, T. G., Preckel, F., Martin, R., and Brunner, M. (2013). Do teacher judgments of student intelligence predict life outcomes? *Learn. Individ. Differ.* 27, 109–119. doi: 10.1016/j.lindif.2013.07.004

Fredricks, J. A., and Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Dev. Psychol.* 38, 519–533. doi: 10.1037/0012-1649.38.4.519

Georgiou, S., Christou, C., Stavrinides, P., and Panaoura, G. (2002). Teacher attributions of student failure and teacher behavior toward the failing student. *Psychol. Sch.* 39, 583–595. doi: 10.1002/pits.10049

Guay, F., Marsh, H. W., and Boivin, M. (2003). Academic self-concept and academic achievement: developmental perspectives on their causal ordering. *J. Educ. Psychol.* 95, 124–136. doi: 10.1037/0022-0663.95.1.124

Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2

Helmke, A., and van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *J. Educ. Psychol.* 87, 624–637. doi: 10.1037/0022-0663.87.4.624

Herbert, J., and Stipek, D. (2005). The emergence of sex differences in children's perceptions of their academic competence. *J. Appl. Dev. Psychol.* 26, 276–295. doi: 10.1016/j.appdev.2005.02.007

Heyder, A., Steinmayr, R., and Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain sex differences in children's math ability self-concept? *Front. Educ.* 4, 1–11. doi: 10.3389/feduc.2019.00034

Hill, C., Corbett, C., and St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: AAUW.

Hoge, R. D., and Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *J. Educ. Psychol.* 76, 777–781. doi: 10.1037/ 0022-0663.76.5.777

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., and Hopp, C. (1990). Sex comparisons of mathematics attitudes and affect: A meta-analysis. *Psychol. Women Q.* 14, 299–324. doi: 10.1111/j.1471-6402.1990.tb00022.x

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., and Wigfield, A. (2002). Changes in children's self-competence and values: sex and domain differences across grades one though twelve. *Child Dev.* 73, 509–527. doi: 10.1111/1467-8624.00421

Jussim, L., and Eccles, J. S. (1992). Teacher expectations II: construction and reflection of student achievement. *J. Pers. Soc. Psychol.* 63, 947–961. doi: 10.1037/0022-3514.63.6.947

Kuklinski, M. R., and Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Dev.* 72, 1554–1578. doi: 10.1111/1467-8624.00365

Lauermann, F., Chow, A., and Eccles, J. S. (2015). Differential effects of adolescents' expectancy and value beliefs about math and english on math/science-related and human-services-related career plans. *Int. J. Sci. Technol. Res.* 7, 205–228.

Lauermann, F., Tsai, Y.-M., and Eccles, J. S. (2017). Math-related career aspirations and choices within eccles et al.'s expectancy–value theory of achievement-related behaviors. *Dev. Psychol.* 53, 1540–1559. doi: 10.1037/dev0000367

Lohbeck, A., and Möller, J. (2017). Social and dimensional comparison effects on math and reading self-concepts of elementary school children. *Learn. Individ. Differ.* 54, 73–81. doi: 10.1016/j.lindif.2017.01.013

Lorenz, G., Boda, Z., Salikutluk, Z., and Jansen, M. (2020). Social influence or social selection? Peer effects on the development of educational expectations among adolescents in Germany. *Br. J. Sociol. Educ.* 41, 643–669. doi: 10.1080/01425692.2020.1763163

Madon, S., Smith, A., Jussim, L., Russell, D. W., Eccles, J., Palumbo, P., et al. (2001). Am I as you see me or do you see me as I am? Self-fulfilling prophecies and self-verification. *Personal. Soc. Psychol. Bull.* 27, 1214–1224. doi: 10.1177/0146167201279013

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *Am. J. Educ. Res.* 23, 129–149. doi:10.2307/1163048

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., et al. (2015). The internal/external frame of reference model of self-concept and achievement relations: age-cohort and cross-cultural differences. *Am. Educ. Res. J.* 52, 168–202. doi: 10.3102/ 0002831214549453

Marsh, H. W., Dowson, M., Pietsch, J., and Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *J. Educ. Psychol.* 96, 518–522. doi: 10.1037/0022-0663.96.3.518

Marsh, H. W., and Martin, A. J. (2011). Academic self-concept and academic achievement: relations and causal ordering. *Br. J. Educ. Psychol.* 81, 59–77. doi: 10.1348/000709910X503501

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Dev.* 76, 397–416. doi: 10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., and Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and English constructs. *Am. Educ. Res. J.* 35, 705–738. doi: 10.2307/1163464

Master, A., and Meltzoff, A. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. International journal of gender. *Sci. Technol.* 12, 152–198.

Master, A., and Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in stem. *Int. J. Sci. Technol. Res.* 12, 152–198. Retrieved from: https://genderandset.open.ac.uk/index.php/genderandset/article/view/674

Möller, J., Kuska, S. K., and Zaunbauer, A. (2011). "Einflussfaktoren auf die Entwicklung des Selbstkonzepts im Grundschulalter [Influencing factors of the development of self- concept during elementary school years]" in *Selbstkonzepte im Grundschulalter*. eds. F. Günther and F. Hellmich (Stuttgart: Kohlhammer), 159–172.

Möller, J., Pohlmann, B., Köller, O., and Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self- concept. *Rev. Educ. Res.* 79, 1129–1167. doi: 10.3102/0034654309337522

Möller, J., Zitzmann, S., Helm, F., Machts, N., and Wolff, F. (2020). A Meta-analysis of relations between achievement and self-concept. *Rev. Educ. Res.* 90, 376–419. doi: 10.3102/0034654320919354

Mullis, I. V. S., Martin, M. O., Foy, P., and Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Boston, MA: TIMSS and PIRLS International Study Center.

Muthén, L. K., and Muthén, B. O. (1998/2013) *Mplus version 7.11 [computer software]*. Los Angeles, CA: Muthén & Muthén.

Natale, K., Viljaranta, J., Lerkkanen, M.-K., Poikkeus, A.-M., and Nurmi, J.-E. (2009). Cross! Lagged associations between kindergarten teachers' causal attributions and children's task motivation and performance in reading. *Educ. Psychol.* 29, 603–619. doi: 10.1080/01443410903165912

National Science Board. (2018). Science and Engineering indicators. Available at: https://nsf.gov/statistics/2018/nsb20181/assets/nsb20181.pdf

OECD. (2019). *PISA 2018 results. Where all students can succeed. 2.* Paris: OECD Publications.

Reilly, D., Neumann, D. L., and Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of national assessment of educational progress assessments. *J. Educ. Psychol.* 107, 645–662. doi: 10.1037/ edu0000012

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., and Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early sex gaps in achievement. *Dev. Psychol.* 50, 1262–1281. doi: 10.1037/a0035073

Schafer, J. L., and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychol. Methods* 7, 147–177. doi: 10.1037//1082-989X.7.2.147

Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Self-concept: validation of construct interpretations. *Rev. Educ. Res.* 46, 407–441. doi: 10.3102/ 00346543046003407

Skaalvik, E. M., and Valås, H. (1999). Relations among achievement, self-concept and motivation in mathematics and language arts: A longitudinal study. *J. Exp. Educ.* 67, 135–149. doi: 10.1080/00220979909598349

Snow, R. E. (1992). Aptitude theory: yesterday, today, and tomorrow. *Educ. Psychol.* 27, 5–32. doi: 10.1207/s15326985ep2701_3

Spinath, B., and Spinath, F. M. (2005a). Development of self-perceived ability in elementary school: the role of parents' perceptions, teacher evaluations, and intelligence. *Cogn. Dev.* 20, 190–204. doi: 10.1016/j.cogdev.2005.01.001

Spinath, B., and Spinath, F. M. (2005b). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learn. Instr.* 15, 87–102. doi: 10.1016/j.learninstruc.2005.04.008

Steinmayr, R., and Spinath, B. (2008). Sex differences in school achievement: what are the roles of personality and achievement motivation? *Eur. J. Personal.* 22, 185–209. doi: 10.1002/per.676

Steinmayr, R., Weidinger, A. F., Heyder, A., and Bergold, S. (2019). Warum schätzen Mädchen ihre mathematischen Kompetenzen geringer ein als Jungen? *Z. Entwicklungspsychol. Pädagog. Psychol.* 51, 71–83. doi: 10.1026/0049-8637/a000213

Stemler, S. E., and Sternberg, R. J. (2013). "The assessment of aptitude" in *APA handbook of testing and assessment in psychology. Testing and assessment in school psychology and education: Vol. 3*. eds. K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel and S. P. Reise (Washington, DC: American Psychological Association), 281–296.

Stipek, D. J., Givvin, K. B., Salmon, J. M., and MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teach. Teach. Educ.* 17, 213–226. doi: 10.1016/S0742-051X(00)00052-4

Tiedemann, J. (2000). Sex-related beliefs of teachers in elementary school mathematics. *Educ. Stud. Math.* 41, 191–207. doi: 10.1023/A:1003953801526

Ullman, J. B. (2007). "Structural equation modeling" in *Using multivariate statistics*. eds. B. G. Tabachnick and L. S. Fidell, vol. *5* (Boston, MA: Allyn & Bacon/Pearson Education), 676–780.

Upadyaya, K., and Eccles, J. (2015). Do teachers' perceptions of children's math and reading related ability and effort predict children's self-concept of ability in math and reading? *Educ. Psychol.* 35, 110–127. doi: 10.1080/01443410.2014.915927

Valentine, J. C., DuBois, D. L., and Cooper, H. (2004). The relation between self-beliefs and academic achievement: a meta-analytic review. *Educ. Psychol.* 39, 111–133. doi: 10.1207/s15326985ep3902_3

Viljaranta, J., Tolvanen, A., Aunola, K., and Nurmi, J.-E. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scand. J. Educ. Res.* 58, 734–756. doi: 10.1080/00313831.2014.904419

Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol. Bull.* 140, 1174–1204. doi: 10.1037/a0036620

Wach, F. S., Spengler, M., Gottschling, J., and Spinath, F. M. (2015). Sex differences in secondary school achievement - the contribution of self-perceived abilities and fear of failure. *Learn. Instr.* 36, 104–112. doi: 10.1016/j.learninstruc.2015.01.005

Weidinger, A. F., Steinmayr, R., and Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Dev.* 89, e138–e156. doi: 10.1111/cdev.12806

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., et al. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *J. Educ. Psychol.* 89, 451–469. doi: 10.1037/0022-0663.89.3.451

Wilgenbusch, T., and Merrell, K. W. (1999). Sex differences in self-concept among children and adolescents: a meta-analysis of multidimensional studies. *Sch. Psychol. Q.* 14, 101–120. doi: 10.1037/h0089000

Wolff, F. (2021). How classmates' gender stereotypes affect students' math self-concepts: A multilevel analysis. *Front. Psychol.* 12:599199. doi: 10.3389/ fpsyg.2021.599199

Wolff, F., Helm, F., and Möller, J. (2018). Testing the dimensional comparison theory: when do students prefer dimensional comparisons to social and temporal comparisons? *Soc. Psychol. Educ.* 21, 875–895. doi: 10.1007/s11218-018-9441-2

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 26 October 2023
DOI 10.3389/fpsyg.2023.1239933

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
John Mark R. Asio,
Gordon College, Philippines
Grace Vesga,
Universidad Antonio Nariño, Colombia

*CORRESPONDENCE
Mohd Effendi Ewan Mohd Matore
✉ effendi@ukm.edu.my

RECEIVED 14 June 2023
ACCEPTED 10 October 2023
PUBLISHED 26 October 2023

CITATION
Wan Hussin WAS and Mohd Matore MEE (2023) The influence of learning styles on academic procrastination among students in mathematics.
Front. Psychol. 14:1239933.
doi: 10.3389/fpsyg.2023.1239933

The influence of learning styles on academic procrastination among students in mathematics

Wan Anis Syamimi Wan Hussin[1] and Mohd Effendi Ewan Mohd Matore[2,3]*

[1]Faculty of Education, The National University of Malaysia (UKM), Bangi, Selangor, Malaysia, [2]Research Centre of Education, Leadership and Policy, Faculty of Education, The National University of Malaysia (UKM), Bangi, Selangor, Malaysia, [3]University Research Groups (KPU), Educational Evaluation, The National University of Malaysia (UKM), Bangi, Selangor, Malaysia

Introduction: Procrastination is a complex psychological and behavioral construct that is strongly influenced by certain personality traits. In mathematics learning, students find it difficult to master the concepts because of less exposure to learning styles. Poor knowledge of mathematical concepts leads to academic procrastination in the subject of Mathematics among students. Therefore, this study aims to identify students' learning styles in Mathematics, identify the stages of students' academic procrastination in Mathematics, and determine whether there is a significant influence of learning styles (visual, auditory, and kinesthetic) on academic procrastination among secondary school students in Mathematics.

Methods: A quantitative approach with a survey was applied. A total of 500 Form Two and Form Four students in five national secondary schools in the Kota Bharu district, Kelantan, were selected using simple random sampling. The duration of data gathering started from 4 October 2022 until 31 January 2023. The Learning Styles Questionnaire and the Academic Procrastination Questionnaire were adapted and verified by eight experts in psychology and counseling. Descriptive and multiple regression tests were carried out using IBM SPSS version 26.0.

Results: The results revealed that the visual learning style was the most dominant learning style among students in the subject of Mathematics, followed by auditory and kinesthetic. The level of students' academic procrastination in Mathematics was low. Besides, multiple regression showed that visual and kinesthetic learning styles were significant contributors or predictors, which amounted to 14.1% of the variation in students' academic procrastination in Mathematics.

Discussion: The implications of this study highlight the possibility to improve programs in schools by exposing students to suitable learning styles so that they can practice effective learning styles in Mathematics and consequently overcome academic procrastination. Further research can be carried out by identifying other factors that encourage academic procrastination in the subject of Mathematics in order to increase students' motivation and self-efficacy.

KEYWORDS

learning styles, academic procrastination, VAK modality, mathematical concepts, secondary school students

1. Introduction

Good cognitive skills are essential in mathematics learning so that students can understand and master mathematical concepts. Different learning styles are identified in mathematics learning. There are three types of learning styles that students often practice in the subject of Mathematics, namely visual, auditory, and kinesthetic (VAK) learning styles (Kurniawan and Hartono, 2020). This is because VAK learning styles are more famous and easy to use in identifying one's learning style (Febriani, 2018; Zulkipli et al., 2019). Thus, appropriate learning styles in the subject of Mathematics help students to understand mathematical concepts more effectively (Cimermanová, 2018), in addition to creating a more fun mathematics learning atmosphere since it is more interactive (Osman et al., 2019). Study by Asio (2020) found that respondents tend to procrastinate with their academics. There were significant differences reported when the respondents' academic procrastination response is grouped according to the program, scholarship status, and religion. This means that the difference in academic procrastination requires further study to identify the cause.

Characteristics of students' learning styles can be identified during mathematics learning. Students with a visual learning style learn mathematical concepts through images such as by depicting formulas (Machromah et al., 2021). For example, in the topic of geometry, students with a visual learning style analyze geometric shapes through pictures. Meanwhile, students with an auditory learning style listen to the teacher's explanation of the mathematical topic being studied (Rahman and Ahmar, 2017), and they always repeat mathematical concepts such as saying mathematical formulas regularly so that they can remember the formulas better. Finally, students who practice a kinesthetic learning style often apply mathematical topics to everyday life (Irvine, 2019). For example, in algebra topics, students are given real situations to form algebraic expressions (Indraswari et al., 2018). Therefore, with an effective learning style, students can master mathematical concepts more quickly and easily.

When students can understand the concepts and content of mathematics effectively, they can complete the tasks given by the teacher excellently without delaying them (Fulano et al., 2021). However, due to the lack of understanding of difficult mathematical concepts, students are not interested in mathematical tasks (Gonda et al., 2021) because they can neither re-explain the concepts nor apply those concepts to solve mathematical problems. Students also find it difficult to relate newly learned mathematical concepts to other mathematical concepts (Ulum and Pujiastuti, 2020).

Most students only memorize formulas and steps to solve mathematical problems without understanding them. This causes students to experience difficulties in solving questions or other mathematical problem-solving exercises even if they involve the same question form (Yuanita et al., 2018). Students also think that mathematics assignments are too complicated and difficult because mathematics learning is more abstract (Hui and Rosli, 2021). Therefore, practicing the right learning style can create a positive atmosphere when learning mathematics (Mundia and Metussin, 2019) and subsequently reduce the tendency of students to practice academic procrastination in the subject of Mathematics. Moreover, academic procrastination in the subject of Mathematics also occurs due to a lack of self-confidence and possible fear of the subject (Agustin and Winarso, 2021).

Based on a study of academic procrastination involving students in Indonesia, the late submission of Mathematics assignments was the most prominent (44%), followed by Physics (31%), and other subjects (28%; Setiyowati et al., 2020). Mathematics education in Malaysia involves many exercises to strengthen students' understanding of mathematical concepts (Sarudin et al., 2022). Teachers usually give assignments or mathematical exercises in textbooks or reference books as homework so that problem-solving skills and mathematical theories and concepts can be honed and improved (Azhar and Rosli, 2021). These mathematics assignments or exercises will then be discussed in the next class, and students may also need to complete them before class. Hence, the suitability of students' learning styles plays a role in encouraging students to complete mathematics assignments.

One of the reasons for less effective learning styles is that students are not given enough exposure to learning styles at an early stage (Cholifah et al., 2018; Ismajli and Imami-Morina, 2018). This is because students are not given specific guidance on the way or style of learning. A less effective learning style causes students to be less interested in the subject of Mathematics and this will lead to academic procrastination because students do not have sufficient basic knowledge of Mathematics (Fulano et al., 2021). Differences in the learning style of each student require teachers to identify and adapt mathematics teaching methods to students. One of the factors influencing students' inability to master the learning content well is the teacher's teaching style, which is not a match to the variety of learning styles that students have (Mustaffa et al., 2021). Thus, teachers need to identify students' learning style patterns in order to better plan teaching and learning strategies based on the characteristics of each learning style (Zulkipli et al., 2019; Abd Shukor and Masroom, 2020).

Studies have found that many teachers still prioritize the conventional method of teaching Mathematics by giving explanations in class (Voon and Amran, 2021). This teaching method is very suitable for students who practice visual and auditory learning styles. This is because students with an auditory modality receive information through hearing, such as discussions and explanations of mathematical concepts. Meanwhile, students with a visual learning style are more likely to learn through sight by looking at the symbols, graphs, or mathematical solutions shown by the teacher on the blackboard (Zulkipli et al., 2019). However, this is quite difficult for kinesthetic students who need hands-on or objects to manipulate in order to understand abstract mathematical concepts that are difficult to explain (Nurrahmah et al., 2021). As such, visual and auditory students may be able to understand mathematical concepts better than kinesthetic students (Umam and Azhar, 2019).

In addition, past studies have also focused more on the relationship between learning styles and other factors such as motivation (Abd Shukor and Masroom, 2020; Noervadila, 2020; Mohamed and Hassan, 2021) and mathematics achievement (Schulze and Bosman, 2018; Altun and Serin, 2019), while studies on the relationship between learning styles and academic procrastination are still limited. Studies on learning styles against academic procrastination should be given attention because students need to identify the learning style patterns to be applied when learning Mathematics. There are previous studies examining the relationship between academic procrastination and student self-regulation in Malaysia (Jerry et al., 2021) as well as the relationship between students' learning styles and academic procrastination in Istanbul (Gunduz, 2020). However, these two studies do not focus on the context of Mathematics education. Studies related to learning styles and academic procrastination are also reportedly limited and have not been widely studied (Çakır et al., 2014; Gunduz, 2020).

There are several important areas where this study makes an original contribution to identify the aspect of learning styles that are able to influence students' academic procrastination in Mathematics. Secondly, this contributes to the body of knowledge by expanding the academic procrastination replications onto Mathematics. Besides, these results can benefit schools by developing the training module for students to manage time properly to avoid procrastination. For teachers, this can help them to modify the pedagogy and assessment in class to match with procrastination problems. Besides, parents also can learn how to train their children's by getting bigger understanding on procrastination.

Therefore, in order to improve the understanding of mathematical concepts among students, this study was conducted to address the following objectives:

1. To identify students' learning styles in the subject of Mathematics in secondary schools.
2. To identify the level of students' academic procrastination in the subject of Mathematics.
3. To determine whether there is a significant influence of learning styles (visual, auditory, kinesthetic) on the academic procrastination of secondary school students in the subject of Mathematics.

2. Literature review

Learning styles refer to how students acquire knowledge and skills as well as respond to the environment in order to process and interpret information (El-Bishouty et al., 2019; Ahmad and Ambotang, 2020; Shamsuddin and Kaur, 2020). Learning styles also refer to how students focus, store, and process new knowledge (Mohd Zamri et al., 2022). Knowledge can be processed when students plan their learning system such as how to revise lessons and how to complete assignments (Mohamed and Hassan, 2021). Students' learning styles differ according to their respective inclinations such as visual, auditory, and kinesthetic. Therefore, in this study, learning styles refer to the inclinations of students in processing mathematical knowledge through visual, auditory, or kinesthetic learning styles.

Apipah (2018) stated that visual, auditory, and kinesthetic student characteristics can be identified in mathematics learning. Students with a visual learning style record the steps to solve mathematical problems in a systematic and clear manner. They also think a lot by describing solutions before solving mathematical problems (Machromah et al., 2021). Meanwhile, students with an auditory learning style are also able to write steps to solve mathematical problems. However, they find it difficult to write complete or detailed solution steps. Finally, students with a kinesthetic learning style are more inclined to movements or touch such as holding objects or making body movements when learning mathematics (Irvine, 2019), causing them to be less careful in solving mathematical problems.

Several studies have examined the most dominant student learning style in the subject of Mathematics. For instance, a study by Virgana (2019) found that most students adopted an auditory learning style in mathematics learning, followed by kinesthetic and visual learning styles. However, a study by Nithya Dewi et al. (2019) showed that students with a visual learning style were more dominant, while the kinesthetic learning style recorded the lowest percentage in the subject of Mathematics. Additionally, based on research by Jurenka et al. (2018), the auditory learning style showed the highest percentage compared to kinesthetic and visual learning styles. Evidently, these three different research findings have shown that students tend to adopt a learning style according to their suitability and ability in mathematics learning (Ahmad Damanhuri et al., 2020).

One of the main mathematical learning skills includes making mathematical connections. This skill helps students to connect mathematical ideas in order to formulate knowledge between topics so that mathematical problem-solving can be performed (Savitri and Rochmad., 2022). Indeed, difficulty in understanding the relevance of Mathematics is influenced by learning styles (Apipah, 2018). This is because learning styles typically determine how students develop mathematical knowledge.

Through learning styles, students learn to absorb, organize, and process mathematical knowledge. A study by Agnes and Yunis (2021), which compared visual, auditory, and kinesthetic students in making connections between social arithmetic and addition operations, showed that kinesthetic and visual students were able to connect mathematical concepts well. This study is different from those conducted by Apipah (2018) and Savitri and Rochmad. (2022) in which kinesthetic students were found to lack the skills to connect mathematical concepts. Nonetheless, it should be noted that all three studies found that visual students were able to make mathematical connections excellently.

In addition, Jasmi and Sulaiman (2018) found that students lacked knowledge about their learning styles, making it difficult for them to determine the learning style that suits them and apply an effective learning style in accordance with the mathematics learning process. As a result, students become lazy and less interested in learning Mathematics in class. Problems also occur when students practice a learning style that is not compatible with the teacher's teaching method (Ismajli and Imami-Morina, 2018). Teachers who tend to explain mathematical concepts without holding activities or discussions in class cause students with different learning styles to pay less attention. When students find it difficult to concentrate during mathematics learning, they will be easily compelled to not complete mathematics assignments as instructed by the teacher and this consequently leads to academic procrastination.

Academic procrastination in Mathematics refers to the act of postponing or delaying mathematics tasks such as delaying mathematical exercises and doing other activities that are more enjoyable (Zacks and Hen, 2018). At school, teachers are not only focused on explaining the concepts and content of mathematics lessons but also on giving assignments in the form of exercises or homework to students to complete at home to hone their mathematical skills. Usually, the assignments or exercises given are taken from mathematics textbooks or exercise books and examples of past questions (Lim and Rosli, 2021). The exercises given are also sometimes considered a form of revision for the topics taught in class to improve the students' memory of a mathematical topic and familiarize students with mathematical problem-solving. These mathematics exercises are then discussed the next day at school.

Some students were also found to intentionally or unintentionally fail to complete the assignments given by their teachers. Surprisingly, academic procrastination in the subject of Mathematics is considered normal among school students. Previous studies have identified two types of academic procrastination, i.e., sending assignments beyond the cut-off time and delaying the completion of assignments (Rabin et al., 2021). Procrastination in academic assignments is caused by a lack of motivation and the existence of various negative emotions (Rahimi and Vallerand, 2021).

Students who experience anxiety in Mathematics are the cause of increased academic procrastination in Mathematics among students (Paechter et al., 2017). In addition to psychological factors, factors such as lack of understanding and basic knowledge of mathematical concepts also cause students to delay their mathematics assignments. This is because learning involves students' existing knowledge; therefore, students need to connect it with new information. In this regard, students who have complete basic knowledge understand mathematics lessons better, and this is evident when students complete the mathematics tasks given by their teachers excellently (Fulano et al., 2021).

Studies have also shown that students tend to procrastinate on mathematics assignments because they do not deem the assignments important. For instance, a study by Asri et al. (2017) found that students were less aware of the importance of mathematics assignments given by their teachers; they would only finish the assignments a few minutes before the class starts, and some students would even copy their friends' steps and answers. Based on data from interviews with several Mathematics teachers, it was found that students were not serious about completing mathematics assignments and would only carry out the assignments for the purpose of fulfilling the teachers' instructions.

In addition, teachers who are less strict with the submission of mathematics assignments may also cause students to not take the assignments seriously (Nurhadi, 2018). This is because students are not subject to any fines or penalties for submitting their assignments late. In fact, some teachers are only aware that students have submitted their assignments, but they have no idea whether students perform their assignments as soon as given or at the final time of the assignment submission (Asri et al., 2017).

Academic procrastination can affect students' mental health, as evidenced in a study by Maria-Ioanna and Patra (2022). Therefore, intervention in helping students deal with academic procrastination in learning is very necessary to help students have good mental and physical health so that their learning process is not interrupted. Previous studies by Dewi et al. (2015) and Toker and Avcı (2015) also emphasized the importance of developing cognitive, emotional, and behavioral skills in dealing with academic procrastination. Academic procrastination affects the emotional, behavioral, and cognitive aspects of students, and its impact on students' emotions can be seen when students find it difficult to initiate, maintain, or complete assignments on time. Procrastination from a behavioral aspect can be translated as a tendency to delay the completion of tasks, while procrastination from a cognitive aspect is seen as the highest level of procrastination. Repetitive procrastination in the emotional, behavioral, and cognitive dimensions is also seen to occur frequently due to the way individuals think.

Cognitive behavioral theory (CBT) states that irrational thinking causes procrastination. Therefore, in the context of dealing with academic procrastination, CBT emphasizes efforts to change beliefs and irrational thoughts to rational thoughts. Zhang et al. (2022) found that negative and irrational thinking leads to academic procrastination. One of the factors influencing negative thinking includes an intervention program that is less effective and does not represent the entire domain, which is cognitive and behavioral (Toker and Avcı, 2015). The approach to overcoming academic procrastination is also deemed obsolete. Particularly, in Malaysia, the approach is only focused on counseling and motivational workshops for students to create positive thinking. However, such an approach is deemed less effective because it only focuses on the emotional aspect. The cognitive behavioral theory is also less commonly used in the context of education in Malaysia, unlike Turkiye, which has created many intervention programs such as psycho-education programs that use a cognitive-behavioral approach to overcome academic procrastination (Toker and Avcı, 2015). Essentially, the CBT program is believed to improve students' basic skills to control cognitive and behavioral academic procrastination.

Therefore, a study related to the influence of learning styles on academic procrastination in the subject of Mathematics must be carried out to examine this issue among secondary school students in detail. Accordingly, the following hypotheses were proposed in the current study:

> $H_0$1: There is no significant influence of visual learning style on the academic procrastination of secondary school students in the subject of Mathematics.
>
> $H_0$2: There is no significant influence of auditory learning style on the academic procrastination of secondary school students in the subject of Mathematics.
>
> $H_0$3: There is no significant influence of kinesthetic learning style on the academic procrastination of secondary school students in the subject of Mathematics.

3. Methodology

3.1. Research design and sample

A quantitative approach with a survey research design was used in this study. Survey methods are appropriate based on the objectives of the study and coincide with the goals of quantitative research to make predictions between variables (Chua, 2021). The quantitative approach was used to determine the influence of learning styles on students' academic procrastination in Mathematics. The survey research design is also capable of obtaining data from a large population (Azizi et al., 2021). The research population includes Form Two and Form Four secondary school students at National Secondary Schools in Kota Bharu district, Kelantan. The total student population in this study was 10,071, while the sample involves 500 respondents. Therefore, the survey method is most appropriate, especially since the study involves a large population (Chua, 2021). In addition, the survey method was chosen because information can be gathered in a short time (Darusalam and Hussin, 2021). This will make the data and information easier to analyze, in addition to saving time.

The respondents of this study were selected through simple random sampling so that each respondent in the population has an equal chance of being selected as a respondent (Chua, 2021). This study involved Form Two and Form Four secondary school students at National Secondary Schools in Kota Bharu district, Kelantan. The rationale for selecting Form Two and Form Four students is that these students are considered mature (Ahmad and Ambotang, 2020) enough to provide rational feedback or answers and can better understand the instructions given in the questionnaire. In addition, Form Two students were selected because they would sit for special assessments such as entrance tests to specialized schools, while Form Four students would sit for the Malaysian Certificate of Education (SPM) examination next year (Afzan and Abd Rahman, 2021). In terms of the selection of the research location, Kelantan was chosen because the state showed an increase in the achievement of the Malaysian Certificate of Education (SPM) 2021. Furthermore, Harian Metro also reported 100% achievement of passing SPM 2021 in 31 schools in Kelantan (Idris, 2022). This shows that students in the state of Kelantan have high enthusiasm and motivation in their learning. The duration of data gathering started from 4 October 2022 until 31 January 2023.

3.2. Questionnaire instrument

The questionnaire instrument in this study was adapted from Mamickam (2010) and Ghazal (2012). The three main sections of the questionnaire include demographics (3 items), students' learning styles (31 items), and academic procrastination (21 items). The instrument was adapted and translated into Malay to meet the context of mathematics learning. Specifically, the questionnaire was translated to meet the values and norms of the local community as well as to determine the validity and reliability of the items in the questionnaire (Yusoff et al., 2018). The demographic section includes the respondents' backgrounds such as age, gender, and socioeconomic status. Sections related to students' learning styles were categorized into three modalities, namely visual, auditory, and kinesthetic, where visual and auditory modalities comprise 11 items, respectively, followed by kinesthetic with 9 items.

A three-point Likert scale was used (1 = rarely practiced, 2 = sometimes practiced, 3 = often practiced). The use of the three-point Likert scale is in line with the aim to identify students' learning styles in Mathematics. Meanwhile, the items for academic procrastination are based on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The five-point Likert scale was used because the reliability coefficient increases with the increasing number of response options (Taherdoost, 2019). The mean score interpretation for the five-point Likert followed the Zaidatol (2009) that more than 3.80 (as high), range 3.40 to 3.79 (as moderate), and less than 3.39 (as low). The 5-point scale is also easy to understand and the respondents can express their views freely owing to the neutral option provided (Park et al., 2021). This will increase the response rate from respondents.

3.3. Validity and reliability

Eight experts in the field of psychology and counseling were appointed to verify the content of the instrument, which coincides with Lynn (1986) who advocated a minimum of three experts. Feedback from experts was used as a reference to improve the research instrument in terms of language and appropriateness of sentences to suit the context of the study and students in Malaysia. Content validity analysis was carried out using content validity ratio (CVR) analysis. Specifically, CVR was chosen to measure the content validity of items because it is more user-friendly and transparent and has been widely used by researchers. CVR is also easier to use because it involves simple calculations and the determination of the critical cut-off value is also available (Lindell and Brandt, 1999).

Before the content validity process was carried out, initially, the total number of items used in the instrument was 51, i.e., 30 items for learning styles and 21 items for academic procrastination. In general, the expert panel certified that the items could measure the content of each construct; however, some issues were presented by some experts to improve the content of the items in the instrument. Based on the calculation of CVR values per the evaluation of the expert panel, 15 items recorded a CVR value less than the critical level of 0.75. Table 1 shows the distribution of CVR values for all items according to each variable in detail. Eight items in the learning style variable did not reach the required cut-off level, while seven items in the academic procrastination variable must be re-evaluated.

For the 15 items that did not meet the CVR critical value, experts raised several issues such as item clarity (8 items), repetitive items (4 items), and layered items (3 items). Based on expert comments and suggestions, some items with issues from the aspect of item clarity, repeated items, and layered items were improved and refined to further enhance the quality of the instrument. One of the refined items was also divided into two separate items per the experts' recommendations. Therefore, following a total of 51 initial items that were distributed to experts for content validity, there were 52 final items consisting of 31 items learning style variable including (11 items for visual and auditory each, with 9 items for kinesthetic) and 21 items for academic procrastination.

A pilot study was conducted in a secondary school in Bachok district, Kelantan, involving 60 students consisting of Form Two and Form Four students. Table 2 shows the reliability test results based on constructs, where Cronbach's alpha coefficient value for the entire learning style constructs was 0.737. Cronbach's alpha coefficient value for the visual modality was 0.677, followed by auditory with 0.573 and kinesthetic with 0.583. Academic procrastination overall showed Cronbach's alpha coefficient of 0.905. This value is more than 0.7 and consider as a valid, good and acceptable value (Sekaran and Bougie, 2016).

A total of 600 copies of the questionnaire were distributed to the five schools involved; however, only 500 sets of questionnaires had complete data for analysis. The collected data were then analyzed using Statistical Package for Social Sciences (SPSS) software version 26.0. Descriptive analysis was used to identify the learning styles of students in the subject of Mathematics and their level of academic procrastination in the subject, while the multiple regression analysis determined the influence of learning styles on the academic procrastination of Form Two and Form Four students in Mathematics.

TABLE 1 CVR value distribution of items.

CVR	Number of items for learning styles	Number of items for academic procrastination	Action
1.00	9	2	Accepted
0.75	13	12	Accepted
<0.75	8	7	Re-evaluated

TABLE 2 Cronbach's alpha coefficient value analysis for each construct.

Construct	Cronbach's alpha	Number of items
Learning style	0.795	31
■ Learning style - Visual	0.677	11
■ Learning style - Auditory	0.573	11
■ Learning style - Kinesthetic	0.583	9
Academic procrastination	0.907	21

4. Research findings

The objectives of this study are (a) to identify the learning styles of students in the subject of Mathematics in secondary schools, (b) to identify the level of students' academic procrastination in the subject of Mathematics, and (c) to determine whether there is a significant influence of learning styles (visual, auditory, and kinesthetic) on the academic procrastination of secondary school students in Mathematics.

The discussion of the descriptive analysis results was based on the first and second objectives. The results for the first objective were obtained by analyzing the overall scale score based on scales of 1 = rarely practiced, 2 = sometimes practiced, and 3 = often practiced. Data for each item was analyzed based on frequency (*f*), percentage (%), and median. Next, the overall scale score for each construct, i.e., visual, auditory, and kinesthetic, was analyzed using mean values and standard deviation. The analysis of each learning style construct in the subject of Mathematics is shown in Table 3.

Overall, based on the mean scores, the visual learning style was the most dominant among Form Two and Form Four students in Mathematics, followed by auditory and kinesthetic learning styles.

The next analysis results were based on the second objective, which is to identify the level of students' academic procrastination in the subject of Mathematics. Items in the academic procrastination construct were measured using a 5-point Likert scale. The overall scale score for academic procrastination was also analyzed based on the mean value and standard deviation. Table 4 shows the analysis results for the construct of academic procrastination. Evidently, the findings showed that the level of students' academic procrastination in the subject of Mathematics was low.

Finally, the inferential analysis results were based on the third research objective. A multiple regression analysis was carried out to determine the influence of learning styles on students' academic procrastination in the subject of Mathematics. Assumptions for the use of multiple regression analysis have been met, including data measured on an interval scale, normal data distribution based on the normality test through skewness and kurtosis values, homoscedasticity, and the existence of a linear relationship between variables. Tables 5–7 show the multiple regression analysis results.

Overall, based on the analysis, the regression model was significant [$F(3,496) = 27.17$, $p < 0.001$, and $R^2 = 0.141$]. Visual ($b = 0.062$, $t = 8.621$, $p < 0.001$) and kinesthetic ($b = -0.021$, $t = -2.576$, $p < 0.001$) learning styles were significant predictors of academic procrastination. The analysis also showed that learning styles could explain 14.1% of the variation in academic procrastination ($R^2 = 0.141$), while the remaining 85.9% were explained by other factors that were not examined in this study.

TABLE 3 Analysis of learning style constructs.

Learning style construct	Mean	Standard deviation
Visual	23.89	3.83
Auditory	22.87	3.44
Kinesthetic	16.66	3.15

TABLE 4 Analysis of academic procrastination construct.

Academic procrastination construct	Mean	Standard deviation
Academic procrastination	3.04	0.55

TABLE 5 Regression model summary.

Model	*R*	R^2	Adjusted R^2
1	0.376	0.141	0.136

TABLE 6 Variance analysis of regression model.

	Sum of squares	*df*	Mean squared	*F*	*p*-value
Regression	21.141	3	7.047	27.167	0.000
Error	128.663	496	0.259		
Total	149.803	499			

Analysis of the Beta standardized coefficient value in Table 7 showed that the standardized coefficient value for the visual learning style (0.433) was greater than that of the kinesthetic learning style (−0.123). The null hypothesis was successfully rejected. There was a significant influence of visual and kinesthetic learning styles on students' academic procrastination in the subject of Mathematics. The findings also showed that the visual learning style had a greater effect on students' academic procrastination in the subject of Mathematics compared to the kinesthetic learning style.

Overall, the contribution of learning styles to the variance in students' academic procrastination in the subject of Mathematics can be formed through a regression equation model, as follows:

$$\textit{Academic procrastination} = 2.164 + (0.062)\,\textit{Visual learning style} + (-0.021)\,\textit{Kinesthetic learning style}$$

Based on the above equation, for every increase of one unit in visual learning style, students' academic procrastination in the subject of Mathematics would increase by 0.062 units. Meanwhile, for every increase of one unit in kinesthetic learning style, students' academic procrastination in the subject of Mathematics would decrease by 0.021 units.

TABLE 7 Multiple regression coefficient analysis.

Variable	Unstandardized coefficient *B*	Unstandardized coefficient *SE*	Standardized coefficient beta	*t*	*p*-value
Constant	2.164	0.175		12.335	0.000
Visual	0.062	0.007	0.433	8.621	0.000
Auditory	−0.011	0.008	−0.068	−1.314	0.189
Kinesthetic	−0.021	0.008	−0.123	−2.576	0.010

5. Discussions

5.1. Objective 1: to identify students' learning styles in the subject of mathematics in secondary schools

The first research objective is about the learning styles of students in the subject of Mathematics in secondary schools. Overall, the findings showed that the majority of Form Two and Form Four students practiced the visual learning style, followed by auditory and kinesthetic learning styles. This agrees with Nithya Dewi et al. (2019) who examined student learning styles in Mathematics in which different learning styles were identified according to the students' inclinations. Although there were differences in student learning styles in the subject of Mathematics, Nithya Dewi et al. (2019) in their study showed that the majority of the students adopted a visual learning style when learning mathematics. This study was supported by Apipah (2018) as well as Savitri and Rochmad. (2022) who found that students typically begin to build an understanding of mathematical concepts through visuals. Moreover, the same findings were also reported by Ahmad et al. (2021) who conducted a study on Form Five students in a cluster secondary school in Pahang. In essence, all the above past findings have strengthened the findings of the current study on the learning styles of students in the subject of Mathematics.

To master mathematical concepts well, critical thinking in mathematics is essential. Ahmad et al. (2021) stated that the right learning style helps sharpen students' critical thinking in the subject of Mathematics. Critical thinking in mathematics learning requires a high level of cognitive skills, and solving mathematical problems involves critical thinking skills. Therefore, students need to know how to solve mathematical problems by applying the correct formulas and concepts as well as explaining concepts and finding solutions to mathematical problems. Accordingly, when students adopt an appropriate learning style, they can improve their critical thinking skills in the subject of Mathematics.

Better mastery of mathematical concepts was demonstrated by students with a visual learning style because Mathematics is closely related to formulas, graphs, and diagrams. Students with a visual learning style visualize these images to ensure that the mathematical knowledge acquired can be stored in long-term memory through careful observation (Apipah, 2018). As such, mathematics learning requires strong visual skills in interpreting information so that students can master mathematical concepts better. Apipah (2018) also explained that the visual learning style helps students to make good mathematical connections—a skill that helps students to connect each mathematical idea and makes it easier for them to summarize each mathematical concept. As a result, students can improve their understanding of the mathematical concepts learned in class. Choosing the wrong learning style makes it difficult for students to understand the relevance of mathematics. This is because, through learning styles, the process of learning mathematics is absorbed, processed, and developed more deeply.

Based on the findings reported by Apipah (2018), students who practice the visual learning style record the steps to solve mathematical problems in a systematic, organized, and clear manner. This is because visual students are more likely to visualize or imagine the steps in a mathematical solution before attempting to solve a mathematical problem. Thus, the findings of the study coincide with the characteristics of the visual learning style where visual students learn better by looking at diagrams or imagining concepts or formulas before translating them into real mathematical problem-solving.

In addition, the findings of the study showed that the auditory learning style was the second-highest learning style in the subject of Mathematics. This is different from the study conducted by Virgana (2019) where the auditory learning style was the dominant learning style in this subject. However, in this study, the levels of visual and auditory learning styles did not show significant differences. This is closely related to the teacher's teaching method in the classroom, which is still contingent on the traditional teaching style where the teacher merely explains topics and mathematical concepts in the classroom. Furthermore, the teaching method of using explanations as the main source of mathematics learning also causes students to practice auditory learning style in the subject of Mathematics.

Meanwhile, the kinesthetic learning style recorded the lowest level in Mathematics, and this finding is supported by Savitri and Rochmad. (2022) who stated that kinesthetic students find it difficult to make mathematical connections well because they learn by manipulating objects or applying them to everyday life. Briefly, since mathematics learning is more abstract (Hui and Rosli, 2021), it is difficult for these students to relate the mathematical concepts learned to their real life. Furthermore, the teaching method of Mathematics teachers, which is entirely contingent on explanations in class without diversifying the teaching methods, limits kinesthetic students' understanding of abstract mathematical concepts. Students with a kinesthetic learning style also learn to understand mathematical concepts with body movements such as moving fingers or hands. This makes it difficult for them to maintain focus or concentration in order to understand mathematical concepts well (Sulisawati et al., 2019).

Mathematics learning, which is abstract, makes it difficult for students to process mathematical content without guidance and explanation from the teacher. According to Pramudya et al. (2021), due to mathematical concepts that are too abstract, most teachers tend to teach students by explaining mathematical concepts in class and then reinforcing students' understanding with mathematical exercises. However, students tend to experience difficulties in understanding the terms or terminology and instructions for solving mathematical problems (Sarudin et al., 2019). Consequently, the problems encountered by the students make it difficult for them to master mathematical concepts well.

This situation further causes mathematics learning in the classroom to become passive. Thus, to create an active mathematics learning process, teachers need to act as facilitators in guiding the students' mathematics learning process. However, in Malaysia, teachers serve as the main source or driver in the delivery of mathematics content. Students' self-learning, which has yet to reach a satisfactory level, also makes it difficult for teachers to change their teaching methods to one that is more student-centered (Pramudya et al., 2021). Thus, the findings of past studies are in line with the current findings, which highlighted the learning styles of students in the subject of Mathematics. Evidently, students preferred to practice visual and auditory learning styles when learning mathematics.

From a theoretical point of view, some students have good cognitive skills owing to effective learning style factors. By integrating cognitive behavioral theory (CBT) into mathematics learning, students are able to think positively by adopting an appropriate learning style based on their inclinations. The right learning style will reduce negative emotions in mathematics learning such as anxiety as well as low levels of motivation and self-efficacy in Mathematics. Students with high mathematics anxiety have difficulty controlling their negative emotions. Thus, several studies have used cognitive behavioral interventions to reduce students' mathematics anxiety, such as by changing the way students think (Asanjarani and Zarebahramabadi, 2021). Mathematics anxiety typically occurs due to the way mathematical content is introduced to students and it does not necessarily stem from students' difficulties in mastering the mathematical content itself (Samuel and Warner, 2021).

In this regard, it is very important for teachers to change students' thinking and steer it toward a more positive direction by helping them plan effective strategies when learning mathematics. One of the effective mathematics learning strategies involves practicing an effective learning style that is compatible with students' inclinations so that they can receive and process mathematical knowledge; students will also feel more confident when they can master mathematical concepts and skills well. As a result, a positive and active mathematics learning atmosphere among students will be realized.

5.2. Objective 2: to identify the level of students' academic procrastination in the subject of mathematics

The second research objective highlights the level of students' academic procrastination in the subject of Mathematics. The findings showed that students' academic procrastination in Mathematics was at a low level, indicating that students had good study practices in the subject of Mathematics.

In general, the low level of students' academic procrastination in the subject of Mathematics reflects the students' positive spirit in learning this subject. Students also had a good attitude toward their learning time to ensure that every mathematics assignment can be completed on time. Good time management helps students plan their learning more efficiently so that every task given by the teacher can be completed perfectly (Ramadhan et al., 2021). Besides, the low level of academic procrastination in the subject of Mathematics suggests that students were aware of every instruction and mathematics assignment given to them, and they also had a high awareness of the importance of a mathematical task or exercise in order to improve their academic performance in the subject of Mathematics.

The positive behavior of students, such as not delaying mathematics assignments, also reflects their high motivation to deepen and learn mathematical topics seriously. Smart students take the opportunity to improve their understanding of mathematical concepts through mathematical exercises given by teachers. As such, when learning mathematics, students do not consider difficult mathematical tasks an obstacle for them to complete the tasks. In fact, they see the difficulty as a challenge and motivation to better strengthen their understanding of mathematical concepts (Hasbullah, 2021; Jeremy et al., 2021). For example, when students are unable to solve certain mathematical problems on some practice questions, they can effectively identify their weaknesses in the mathematical topic.

Mathematical exercises with various difficulty levels also help students to identify their level of understanding of a mathematical topic, making it easier for them to focus on the topic. When students are able to complete mathematics assignments, they will feel more confident in the mathematical concepts learned. Consequently, negative emotions in mathematics learning can be overcome.

Moreover, the low level of academic procrastination in the subject of Mathematics among students helps teachers plan their learning strategies. Teachers are more enthusiastic to teach mathematical topics in class because students take every task entrusted to them seriously. Teachers can also reflect on teaching through the mathematics assignments given to students (Ariffin, 2017), besides identifying students' weaknesses from the way they solve mathematical problems in a given task. Next, teachers can help students to focus more on questions that are considered difficult (Listiwikono, 2022). For instance, teachers can thoroughly and frequently explain the mistakes that students always make when solving mathematical problems based on their review of the students' mathematics assignments.

In addition, teachers can also emphasize difficult mathematical topics by increasing the mathematical exercises to be done so that students' understanding of the topics can be enhanced. This can be ensured when students are able to complete their mathematics assignments well because mathematical questions or exercises can be used as benchmarks for teachers to assess students' understanding of the subject of Mathematics. Besides, due to the nature of mathematics learning, which mainly depends on the ability and skills of students to solve mathematical problems, students can be given many mathematical exercises to help build their understanding of the subject and strengthen mathematical concepts by applying the correct mathematical formulas to solve mathematical problems.

Since this study is only focused on secondary school students, the findings may differ from the previous studies focusing on students' academic procrastination in the subject of Mathematics at the university level. Research on academic procrastination in Mathematics in Malaysia is also very limited compared to Indonesia. Based on the findings of previous studies, the level of students' academic procrastination in the subject of Mathematics was high (Setiyowati et al., 2020; Agustin and Winarso, 2021). The findings of the current study are different, as a decrease in the level of students' academic procrastination in the subject of Mathematics was observed, i.e., from a high level to a low level.

The significant difference in the level of students' academic procrastination in the subject of Mathematics in this study may be due to the respondents and the research design used. Previous studies employed university students as respondents compared to the current study, which only involved secondary school students. In addition, this study only used a survey research design and involved a large number of respondents (i.e., 500) compared to the previous studies using interviews and observations as well as involving a small number of respondents. Although these factors may explain the difference in results, further research is still needed to strengthen and confirm the difference.

Furthermore, the difference in the level of students' academic procrastination in the subject of Mathematics may be due to the positive changes taking place in the Malaysian education system. Various efforts and initiatives have been taken by the Ministry of Education Malaysia (MoE) to empower the country's education, such as introducing a special Educational TV channel through KPM's DidikTV to help students study at home and provide mathematics teaching and learning (PdP) support materials for the use of students and parents, which can be accessed online. The initiative taken by the MoE encourages continuous learning of mathematics. Students do not have to wait to acquire the full source of information through teachers at school; instead, students can practice self-learning when completing mathematics assignments to improve their understanding of mathematical concepts.

From a theoretical perspective, students' academic procrastination is closely related to negative emotions and irrational thoughts that students face when learning mathematics (Hima and Sari, 2021), which affects their cognitive ability. Students who procrastinate on mathematics assignments are motivated by fear of failure or high mathematics anxiety. This happens because students are not confident in their abilities and mathematical skills; hence, they will try to avoid thinking about and doing mathematical tasks. In this vein, through the application of the cognitive behavioral theory, students are made aware of the importance of overcoming academic procrastination in the subject of Mathematics. In addition, effective cognitive behavior can also foster students' positive thinking toward mathematics learning (Asanjarani and Zarebahramabadi, 2021). Essentially, positive thinking among students can make them more efficient and confident to learn mathematics and this further increases their efforts to improve their weaknesses in difficult mathematical concepts.

Overall, although the level of students' academic procrastination in the subject of Mathematics was low, focus still needs to be given to ensure that mathematics achievement can be improved. The factors affecting students' academic procrastination in the subject of Mathematics must also be examined in detail to determine the long-term effect of these factors on student learning, particularly in the subject of Mathematics so that measures can be taken to curb academic procrastination in Malaysia.

5.3. Objective 3: to determine whether there is a significant influence of learning styles (visual, auditory, and kinesthetic) on the academic procrastination of secondary school students in the subject of mathematics

The multiple regression analysis results showed that visual and kinesthetic learning styles had a significant influence on the academic procrastination of students in the subject of Mathematics. The positive beta standardized coefficient value indicates that students who adopted a visual learning style contributed to academic procrastination in the subject of Mathematics. Generally, visual students are more likely to learn with the help of texts or pictures and by looking at diagrams or real objects; therefore, when there are pictures or illustrations that are deemed more interesting than their reading materials at that time, visual students are more easily distracted by the attraction of colorful illustrations or diagrams found in other reading materials.

As a result, visual students tend to lose focus of their initial learning goals easily. This is supported by Mahadi et al. (2022) who stated students with a visual learning style are more easily distracted than those with other learning styles. When visual students find it difficult to fully focus on mathematics learning in class, they cannot understand the mathematical concepts being presented and this will lead to academic procrastination because they cannot apply mathematical concepts in mathematical tasks.

Since kinesthetic students also had a significant influence on academic procrastination in the subject of Mathematics, the negative beta standardized coefficient value suggests that students' academic procrastination in Mathematics decreased with kinesthetic learning style. This is probably because kinesthetic students use many hands-on or applications with real situations, which increases their interest in learning mathematics. Students with a kinesthetic learning style are also more confident in their learning when touching objects or engaging in hands-on learning activities (Mahadi et al., 2022). For example, in the topic of polygons, kinesthetic students can better understand the concept of polygons when they relate the combinations of polygon shapes around them. A good student's mastery of mathematical concepts is demonstrated through mathematical exercises; when students can complete mathematics assignments perfectly, academic procrastination can be overcome.

The third objective finding supports the flow of theoretical supports of ideas from VAK learning styles model and cognitive behavioral theory. Although the influence of learning styles on academic procrastination among students in mathematics is only 14 percent, this finding shows the existence of the interaction of emotional, cognitive and behavioral components that result in the practice of student academic procrastination in the subject of Mathematics. In this research context, cognitive behavioral theory provides the students with a way of understanding to the world experience or mathematics learning, and enabling students to make changes if they need to. It does this by dividing the students experience into central components: namely thoughts (cognitions), feelings (emotions), and behaviors except for physiology (biology) that has not been discussed in this paper. All these elements show the interaction of students' irrational thinking, unstable emotions and less efficient learning styles causing academic procrastination in Mathematics.

Through cognitive behavioral theory, students' knowledge in Mathematics can be built through positive thinking and behavior. Cognitive behavioral theory states that irrational thinking causes procrastination. In the context of dealing with academic procrastination, CBT emphasizes efforts to change beliefs and irrational thoughts to rational thoughts. Zhang et al. (2022) found that negative and irrational thinking leads to academic procrastination. One of the factors influencing negative thinking includes an intervention program that is less effective and does not represent the entire domain, which is cognitive and behavioral (Toker and Avcı, 2015).

Cognitive behavioral specifically is a form of psychological part that has been demonstrated to be effective for a range of problems including depression, anxiety disorders, alcohol and drug use problems, marital problems, eating disorders, and severe mental illness. The students involved in this study are Form Two and Form Four students in secondary schools among 14 to 16 years face many challenges including in learning Mathematics. In fact, students are likely to have some characteristics of depression, anxiety disorders, eating disorders, and severe mental illness as a result of stress in learning mathematics which can cause academic procrastination.

This finding also supported by VAK learning styles model by proving that visual and kinesthetic learning styles have a significant influence on students' academic procrastination in Mathematics. Visual learning style is the most dominant learning style. Visual in this research highlighted on learners respond to images and graphics in learning Mathematics. Students begin to build an understanding of Mathematical concepts through visuals. Mastery of Mathematics concepts is better demonstrated by students with a visual style because Mathematics is closely related to formulas, formulas, graphs and diagrams. The visual learning style shows the ability to visualize images to ensure that the Mathematical knowledge gained can be stored for long-term memory through careful observation (Apipah, 2018). This shows that mathematics learning uses a lot of strong visual skills in interpreting information so that students can master Math concepts better. To master the concept of Mathematics well, critical thinking in Mathematics is really necessary.

In this context, respondents with a visual learning style indirectly show the need for cognitive behavioral theory. Apart of visual style requires long-term memory and strong visual skills, students have the potential to feel a little pressure and depression in learning mathematics which can cause academic procrastination. Students tend to not perform very well in their Mathematics assignments. Due to this, they may face psychological distress and thinks that they would fail the assignment. This is distorted thinking because there is also a chance that they would perform better in future. In this regard, the immediate action that need to be taken is improving the visual learning styles techniques for overcoming students' academic procrastination in Mathematics.

6. Implication

As for the implications, this study suggests improvements in excellence programs in schools by giving exposure to students on how to identify the appropriate learning styles so that they can practice effective learning styles in the subject of Mathematics and overcome academic procrastination.

7. Limitation and future research

This study has certain limitations as it only focuses on Form Two and Form Four students in learning Mathematics; therefore, the findings of the study cannot be generalized to other subjects at school. This study is also limited to one independent variable, which is the learning styles of students toward their academic procrastination in Mathematics. The findings have shown that the learning style factor could only explain 14.1% of the total variance in students' academic procrastination in the subject of Mathematics. This shows that there are other factors that can still influence students' academic procrastination in the subject of Mathematics, which have yet to be examined. Future research can be carried out by identifying other factors that lead to academic procrastination in the subject of Mathematics in order to increase the motivation and self-efficacy of students in this subject.

8. Conclusion

Different learning styles of students have been identified based on their inclinations in learning Mathematics. There are three main types of learning styles practiced by students in the subject of Mathematics, namely visual, auditory, and kinesthetic learning styles. Visual learning style was the dominant learning style among students in the subject of Mathematics, followed by auditory and kinesthetic learning styles. The level of students' academic procrastination in Mathematics was also low. In addition, the study found that visual and kinesthetic learning styles had a significant influence on students' academic procrastination in the subject of Mathematics. Therefore, exposure to the different learning styles must be improved through excellence programs in schools to provide knowledge to students in order to allow them to identify their learning styles and practice an effective learning style so that the quality of student learning in the subject of Mathematics can be improved.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ministry of Education (MoE), Malaysia. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

WW was involved in the collection of data, data analysis, and producing the first draft of this study. MM was responsible for reviewing and editing, as well as supervising. All authors are responsible for the concept and design of the study. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

My deepest appreciation and gratitude go to my supervisor, MM, and UKM for giving me the opportunity to carry out this meaningful study and for guiding me from the beginning to the end of the study. Special thanks also to Kumpulan Penyelidikan Universiti (KPU), Penilaian Pendidikan (Educational Evaluation) UKM.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abd Shukor, N. S., and Masroom, M. N. (2020). Gaya Pembelajaran dan Motivasi Dalam Kalangan Calon SPM Di Sebuah Sekolah Menengah Agama di Negeri Johor. *J. Kemanusiaan* 18, 77–90.

Afzan, R. A., and Abd Rahman, I. (2021). Isu dan Cabaran dalam Pelaksanaan Pendidikan Peringkat Rendah dan Menengah: Pendekatan Malaysia Semasa Pandemik Covid-19. *Malays. J. Soc. Sci. Hum.* 6, 1–13. doi: 10.47405/mjssh.v6i9.1043

Agnes, T. H., and Yunis, S. (2021). Analisis Kemampuan Koneksi Matematis Ditinjau dari Gaya Belajar Siswa. *Laplace: J. Pendidikan Matematika* 4, 158–168. doi: 10.31537/laplace.v4i2.550

Agustin, A., and Winarso, W. (2021). Profile of student academic procrastination behavior in problem solving and mathematical digital literacy. *J. Didaktik Matematika* 8, 190–207. doi: 10.24815/jdm.v8i2.21367

Ahmad, A. A., and Ambotang, A. S. (2020). Gaya Pembelajaran dan Persekitaran Sekolah serta Perkaitannya Dengan Kecerdasan Emosi Pelajar Sekolah Menengah Luar Bandar. *Malays. J. Soc. Sci. Hum.* 5, 141–148. doi: 10.47405/mjssh.v5i5.406

Ahmad Damanhuri, N. H., Jamlus, S., and Ahmad, S. S. (2020). Faktor-faktor yang Mempengaruhi Pencapaian Subjek Matematik Pengurusan bagi Pelajar Sijil Pengoperasian Perniagaan di Kolej Komuniti Hulu Langat. *Int. J. Educ. Pedagogy* 2, 275–287.

Ahmad, S., Nassir, A. A., Tarmuji, N., Yusob, K., and Yacob, N. (2021). "Assessing the relationship between students' learning styles and mathematics critical thinking ability in a "cluster school."" in *E-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (ICMS 2021)*.

Altun, H., and Serin, O. (2019). Determination of learning styles and achievements of talented students in the fields of science and mathematics. *Cypriot J. Educ. Sci.* 14, 80–89. doi: 10.18844/cjes.v14i1.3441

Apipah, S. (2018). An analysis of mathematical connection ability based on student learning style on visualization auditory kinesthetic (VAK) learning model with self-assessment. *J. Phys. Conf. Ser.* 983:012138. doi: 10.1088/1742-6596/983/1/012138

Ariffin, N. E. (2017). Persepsi murid tahun lima terhadap penggunaan kaedah model bar dalam penyelesaian masalah matematik berayat tajauk pecahan. *Proc. ICECRS* 1, 287–304. doi: 10.21070/picecrs.v1i1.613

Asanjarani, F., and Zarebahramabadi, M. (2021). Evaluating the effectiveness of cognitive-behavioral therapy on math self-concept and math anxiety of elementary school students. *Preven. Sch. Failure* 65, 223–229. doi: 10.1080/1045988X.2021.1888685

Asio, R. J. M. (2020). The relationship between academic procrastination and academic performance of freshmen students from a teacher education institution. *J. Hum. Soc. Sci.* 2, 105–115. doi: 10.36079/lamintang.jhass-0203.156

Asri, D., Setyosari, P., Hitipeuw, I., and Tutut, C. (2017). The academic procrastination in junior high school students' mathematics learning: a qualitative study. *Int. Educ. Stud.* 10, 70–77. doi: 10.5539/ies.v10n9p70

Azhar, N. A., and Rosli, R. (2021). Analisis kandungan topik tambah dan tolak dalam buku teks matematik tahap 1 sekolah kebangsaan. *J. Dunia Pendidikan* 3, 394–405.

Azizi, A., Peter, V., Ismail, M., and Mohd Dahlan, A. M. (2021). *Kaedah Penyelidikan dalam Pendidikan. 5th Edn*. Tanjong Malim: Penerbit Universiti Pendidikan Sultan Idris.

Çakır, S., Akça, F., Kodaz, A., and Tulgarer, S. (2014). The survey of academic procrastination on high school students in terms of school burn-out and learning styles. *Procedia Soc. Behav. Sci.* 114, 654–662. doi: 10.1016/j.sbspro.2013.12.763

Cholifah, T., Degeng, I., and Utaya, S. (2018). Analisis gaya belajar siswa untuk peningkatan kualitas pembelajaran. *J. Nat. Sci. Educ.* 1, 65–74. doi: 10.31002/nse.v1i2.273

Chua, Y. (2021). *Kaedah dan Statistik Penyelidikan: Kaedah Penyelidikan Buku 1. 4th Edn*. Kuala Lumpur: McGraw-Hill Education.

Cimermanová, I. (2018). The effect of learning styles on academic achievement in different forms of teaching. *Int. J. Instr.* 11, 219–232. doi: 10.12973/iji.2018.11316a

Darusalam, G., and Hussin, S. (2021). *Metodologi Penyelidikan dalam Pendidikan: Amalan dan Analisis Kajian. 3rd Edn*. Kuala Lumpur: Penerbit Universiti Malaya.

Dewi, N., Sedanayasa, G., and Dharmayanti, P. (2015). Teori Konseling Kognitif-Behavioral dengan Teknik Self-Management untuk Meminimalisasi Prokrastinasi Akademik. *J. Ilmiah Bimbingan Konseling Undiksha.* 3, 1–11. doi: 10.23887/jibk.v3i1.5471

El-Bishouty, M. M., Aldraiweesh, A., Alturki, U., Tortorella, R., Yang, J., Chang, T. W., et al. (2019). Use of Felder and Silverman learning style model for online course design. *Educ. Technol. Res. Dev.* 67, 161–177. doi: 10.1007/S11423-018-9634-6

Febriani, S. (2018). Pengaruh model guided discovery learning dengan metode eksperimen terhadap hasil belajar fisika peserta didik MAN 1 Mataram ditinjau dari gaya belajar VAK. *J. Pendidikan Fisika Dan Teknologi* 5, 82–90. doi: 10.29303/jpft.v5i1.872

Fulano, C., Magalhães, P., Núñez, J., Marcuzzo, S., and Rosario, P. (2021). As the twig is bent, so is the tree inclined: lack of prior knowledge as a driver of academic procrastination. *Int. J. Sch. Educ. Psychol.* 9, S21–S33. doi: 10.1080/21683603.2020.1719945

Ghazal, M. (2012). Academic procrastination: prevalence and causes from the point of view of undergraduate students. *Jordanian J. Educ. Sci.* 8, 131–149.

Gonda, D., Pavlovičová, G., Tirpáková, A., and Ďuriš, V. (2021). Setting up a flipped classroom design to reduce student academic procrastination. *Sustainability* 13:8668. doi: 10.3390/su13158668

Gunduz, G. (2020). The relationship between academic procrastination behaviors of secondary school students, learning styles and parenting behaviors. *Int. J. Contemp. Educ. Res.* 7, 253–266. doi: 10.33200/ijcer.731976

Hasbullah, H. (2021). Pengaruh Prokrastinasi Akademik dan Motivasi Belajar terhadap Penguasaan Konsep Matematika (Survei Pada Siswa SMP Negeri di Kota Tangerang). *Alfarisi* 2, 189–197.

Hima, L. R., and Sari, L. N. I. (2021). Procrastination of students in working mathematics. *IJO-Int. J. Math.* 4, 9–17.

Hui, E., and Rosli, R. (2021). Kebimbangan dan Efikasi Kendiri Terhadap Pembelajaran Matematik dalam kalangan Pelajar Tingkatan Empat. *Malays. J. Soc. Sci. Hum.* 6, 41–53. doi: 10.47405/mjssh.v6i3.690

Idris, S. R. (2022). SPM: Kelulusan 100% di 31 Sekolah Kelantan. Harian Metro. Available at: https://www.hmetro.com.my/mutakhir/2022/06/852369/spm-kelulusan-100-di-31-sekolah-kelantan [Accessed June 16, 2022].

Indraswari, N. F., Budayasa, I. K., and Ekawati, R. (2018). Algebraic reasoning in solving mathematical problem based on learning style. *J. Phys. Conf. Ser.* 947:012061. doi: 10.1088/1742-6596/947/1/012061

Irvine, J. (2019). Strategies for enhancing mathematics learning for students who are kinesthetic learners. *Gazette-Ontario Assoc. Math.* 57, 41–44.

Ismajli, H., and Imami-Morina, I. (2018). Differentiated instruction: understanding and applying interactive strategies to meet the needs of all the students. *Int. J. Instr.* 11, 207–218. doi: 10.12973/iji.2018.11315a

Jasmi, K., and Sulaiman, Z. (2018). Kajian Kes Penggunaan Gaya Pembelajaran Pelajar Pusat Pendidikan Islam Andalus, Singapura. *Attarbawiy: Malays. Online J. Educ.* 2, 36–46.

Jeremy, D., Natalia, S., and Lumbantoruan, J. (2021). Faktor Procrastination Penyebab Mahasiswa Terlambat Menyelesaikan Tugas Akhir di Program Studi Pendidikan Matematika. *Brillo J.* 1, 1–10. doi: 10.56773/bj.v1i1.5

Jerry, A., Rahman, S., and Surat, S. (2021). Prokrastinasi Akademik dan Pembelajaran Regulasi Kendiri dalam Kalangan Murid Sekolah Rendah. *J. Dunia Pendidikan* 3, 592–601.

Jurenka, R., Stareček, A., Vraňaková, N., and Caganova, D. (2018). "The learning styles of the generation group Z and their influence on learning results in the learning process." in International Conference on Emerging eLearning Technologies and Applications (ICETA). pp. 251–260.

Kurniawan, A., and Hartono, S. (2020). The effect of learning style on academic achievement of prospective teachers in mathematics education. *J. Math. Pedagogy* 2, 26–31. doi: 10.26740/jomp.v2n1.p%25p

Lim, C., and Rosli, R. (2021). Analisis Contoh dan Latihan bagi Topik Penambahan dan Penolakan dalam Buku Teks Matematik SJKC. *Malays. J. Soc. Sci. Hum.* 6, 254–270. doi: 10.47405/mjssh.v6i9.1008

Lindell, M., and Brandt, C. (1999). Assessing interrater agreement on the job relevance of a test: a comparison of CVI, T, rWG(J), and r*WG(J) indexes. *J. Appl. Psychol.* 84, 640–647. doi: 10.1037/0021-9010.84.4.640

Listiwikono, E. (2022). Meningkatkan Hasil Belajar Matematika Siswa Kelas Ix A Smp Negeri 2 Negara dengan Pendekatan Dril. *J. Bina Ilmu Cendekia* 1, 35–43. doi: 10.46838/jbic.v1i1.4

Lynn, M. (1986). Determination and quantification of content validity. *Nurs. Res.* 35, 382–385. doi: 10.1097/00006199-198611000-00017

Machromah, I., Ishartono, N., Mirandhani, A., Samsudin, M., and Basry, W. (2021). PISA problems solving of students with visual learning styles. *J. Phys. Conf. Series 5th PROFUNEDU*:012010. doi: 10.1088/1742-6596/1720/1/012010

Mahadi, F., Husin, M., and Hassan, N. M. (2022). Gaya Pembelajaran: Visual, Auditori atau Kinestetik. *J. Hum. Soc. Sci.* 4, 29–36. doi: 10.36079/lamintang.jhass-0401.340

Mamickam, K. (2010). The learning styles and achievement of students in science using visual, auditory and kinesthetic (VAK) modality. [Universiti Teknologi Malaysia/master's thesis]. Iskandar Puteri: University of Technology Malaysia.

Maria-Ioanna, A., and Patra, V. (2022). The role of psychological distress as a potential route through which procrastination may confer risk for reduced life satisfaction. *Curr. Psychol.* 41, 2860–2867. doi: 10.1007/S12144-020-00739-8

Mohamed, J. S. F. W., and Hassan, M. M. (2021). Konsep Kendiri dan Gaya Pembelajaran terhadap Motivasi Akademik dalam kalangan Mahasiswa. *Malays. J. Soc. Sci. Hum.* 6, 75–105. doi: 10.47405/mjssh.v6i2.683

Mohd Zamri, N. S., Yusof, A., Sarudin, A., and Ismail, M. S. (2022). Gaya Pembelajaran Pelajar Bagi Subjek Bahasa Melayu Semasa Pandemik. *E-J. Bahasa Dan Linguistik* 4, 58–79. doi: 10.53840/ejbl.v4i2.118

Mundia, L., and Metussin, H. (2019). Exploring factors that improve mathematics achievement in Brunei. *Stud. Educ. Eval.* 60, 214–222. doi: 10.1016/j.stueduc.2018.10.003

Mustaffa, Z., Hussin, Z., and Sulaiman, A. M. (2021). Pedagogi Terbeza Untuk Pengajaran Guru Terhadap Kepelbagaian Murid. *Malays. J. Soc. Sci. Hum.* 6, 202–214. doi: 10.47405/mjssh.v6i9.997

Nithya Dewi, S., Lina, H., Noor Azida, S., Zuraina, A., Norhasyimah, H., Nur Shamsiah, A. R., et al. (2019). Learning styles and teaching styles determine students' academic performances. *Int. J. Eval. Res. Educ.* 8, 610–615. doi: 10.11591/ijere.v8i3.20345

Noervadila, I. (2020). Pengaruh Gaya Belajar dan Motivasi Belajar terhadap Hasil Belajar Siswa Pada Mata Pelajaran Matematika Kelas X IPS Semester Genap di MA Fathus Salafi Tahun. *J. IKA PGSD (Ikatan Alumni PGSD) UNARS* 8, 48–56. doi: 10.36841/pgsdunars.v8i1.582

Nurhadi, M. (2018). Studi Deskriptif Prokrastinasi Akademik pada Siswa Kelas XI IPS 4 SMA Negeri 11 Banjarmasin. *J. Pelayanan Bimbingan Dan Konseling* 1, 87–93. doi: 10.20527/jpbk.2018.1.2.1495

Nurrahmah, A., Rochmad, R., and Isnarto, I. (2021). Kemampuan Berpikir Abstraksi Matematis Mahasiswa pada Mata Kuliah Statistika Matematika ditinjau dari Gaya Belajar. PRISMA, Prosiding Seminar Nasional Matematika, 67–64.

Osman, Z., Sarudin, A., Janan, D., and Omar, A. (2019). The teaching of Malay essay writing based on an authentic approach. *Int. J. Innov. Tech. Explor. Eng.* 9, 175–181. doi: 10.6007/IJARBSS/v9-i1/5894

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., and Papousek, I. (2017). Mathematics anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. *Front. Psychol.* 8:1196. doi: 10.3389/fpsyg.2017.01196

Park, S., Park, J., Kang, J., and Cha, M. (2021). The presence of unexpected biases in online fact-checking. *Harvard Kennedy School Misinformation Rev.* 2, 1–11. doi: 10.37016/mr-2020-53

Pramudya, I., Mardiyana, M., Sutrima, S., Sujatmiko, P., Aryuna, D. R., Umbarno, U., et al. (2021). Pelatihan Praktek Pembelajaran Matematika dengan Pendekatan "Mikir" untuk Meningkatkan Kemampuan Berpikir Logis Matematis. *J. Pendidikan Matematika Dan Matematika SOLUSI* 5, 281–287.

Rabin, L. A., Krishnan, A., Bergdoll, R., and Fogel, J. (2021). Correlates of exam performance in introductory statistics courses: basic math skills along with self-reported psychological/ behavioral and demographic variables. *Stat. Educ. Res. J.* 20, 1–21. doi: 10.52041/serj.v20i1.97

Rahimi, S., and Vallerand, R. (2021). The role of passion and emotions in academic procrastination during a pandemic (COVID-19). *Personal. Individ. Differ.* 179:110852. doi: 10.1016/j.paid.2021.110852

Rahman, A., and Ahmar, S. A. (2017). Relationship between learning styles and learning achievement in mathematics based on genders. *World Trans. Eng. Tech. Educ.* 15, 74–77. doi: 10.26858/wtetev15i1y2017p7477

Ramadhan, K., Alipatan, M., and Khotimah, H. (2021). Penyebab prokrastinasi siswa pada pembelajaran matematika. *Riemann* 3, 64–72. doi: 10.38114/riemann.v3i2.151

Samuel, T. S., and Warner, J. (2021). "I can math!": reducing math anxiety and increasing math self-efficacy using a mindfulness and growth mindset-based intervention in first-year students. *Commun. College J. Res. Prac.* 45, 205–222. doi: 10.1080/10668926.2019.1666063

Sarudin, A., Redzwan, H., Osman, Z., Maa, R. N. F. A. R., and Albakri, I. S. M. A. (2019). Menangani kekaburan kemahiran prosedur dan terminologi awal Matematik: Pendekatan leksis berdasarkan Teori Prosodi Semantik. *Malays. J. Learn. Instr.* 16, 255–294. doi: 10.32890/mjli2019.16.2.10

Sarudin, A., Sufian, I., Abdullah, Y., Husna Faredza, M. R., Zulkifli, O., and Hermandra, H. (2022). Peranan Leksis dalam Mengukuhkan Pemahaman Konsep Awal Matematik. *Malays. J. Soc. Sci. Hum.* 7:e001770. doi: 10.47405/mjssh.v7i9.1770

Savitri, Y. A., and Rochmad, . (2022). Kemampuan Koneksi Matematis Berdasarkan Gaya Belajar Siswa pada Model Pembelajaran Connecting, Organizing, Reflecting, and Extending. In Prosiding Seminar Nasional Pendidikan Matematika, Universitas Mulawarman, 29–38.

Schulze, S., and Bosman, A. (2018). Learning style preferences and mathematics achievement of secondary school learners. *S. Afr. J. Educ.* 38, 1–8. doi: 10.15700/saje.v38n1a1440

Sekaran, U., and Bougie, R. (2016). *Research methods for business: A skill building approach. 7th Edn*. West Sussex: John Wiley & Sons.

Setiyowati, A., Triyono, T., and Rachmawati, I. (2020). Academic procrastination of high school students in East Java. *PSIKOPEDAGOGIA: J. Bimbingan Dan Konseling* 9, 53–52. doi: 10.12928/psikopedagogia.v9i1.17907

Shamsuddin, N., and Kaur, J. (2020). Students' learning style and its effect on blended learning, does it matter? *Int. J. Eval. Res. Educ.* 9, 195–202. doi: 10.11591/ijere.v9i1.20422

Sulisawati, D. N., Lutfiyah, L., Murtinasari, F., and Sukma, L. (2019). Differences of visual, Auditorial, kinesthetic students in understanding mathematics problems. *Malikussaleh J. Math. Learn.* 2, 45–51. doi: 10.29103/mjml.v2i2.1385

Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. *Int. J. Acad. Res. Manag.* 8, 1–10.

Toker, B., and Avcı, R. (2015). Effect of cognitive-behavioral-theory-based skill training on academic procrastination behaviors of university students. *Educ. Sci. Theory Prac.* 15, 1157–1168. doi: 10.12738/estp.2015.5.0077

Ulum, M., and Pujiastuti, H. (2020). Learning styles against students understanding mathematical concepts. *Edumatica: Jurnal Pendidikan Matematika* 10, 38–44. doi: 10.22437/edumatica.v10i2.9185

Umam, K., and Azhar, E. (2019). Peningkatan pemahaman konsep matematis siswa melalui pendekatan SAVI (somatic, auditory, visual and intellectual). *J. Pendidikan Matematika Indonesia* 4, 53–57. doi: 10.26737/jpmi.v4i2.1038

Virgana, V. (2019). Understanding of mathematical concepts through cooperative learning, and learning styles. *J. Educ. Learn.* 13, 212–218. doi: 10.11591/edulearn.v13i2.9917

Voon, S., and Amran, M. (2021). Pengaplikasian Teori Pembelajaran Konstruktivisme Dalam Pembelajaran Matematik: application of constructivism learning theory in mathematical learning. *Sains Insani* 6, 73–82. doi: 10.33102/sainsinsani.vol6no2.285

Yuanita, P., Zulnaidi, H., and Effandi, Z. (2018). The effectiveness of realistic mathematics education approach: the role of mathematical representation as mediator between mathematical belief and problem solving. *PLoS One* 13:e0204847. doi: 10.1371/journal.pone.0204847

Yusoff, F., Shahimin, M., Saat, N., and Salehuddin, K. (2018). Kesahan dan Kebolehpercayaan Soal Selidik "Pengalaman Pembelajaran Berbantukan Rakan" Versi Bahasa Melayu Untuk Pembelajaran Teknik Lampu Celah Biomikroskopi. *J. Pers. Pelajar* 21, 13–22.

Zacks, S., and Hen, M. (2018). Academic interventions for academic procrastination: a review of the literature. *J. Prev. Interven. Commun.* 46, 117–130. doi: 10.1080/10852352.2016.1198154

Zaidatol, A. L. (2009). Entrepreneurship as a centre choice: an analysis of entrepreneurial self efficiency and intention of university student. *Eur. J. Soc. Sci.* 9, 338–346.

Zhang, R., Chen, Z., Hu, B., Zhou, F., and Feng, T. (2022). The anxiety-specific hippocampus–prefrontal cortex pathways links to procrastination through self-control. *Hum. Brain Mapp.* 43, 1738–1748. doi: 10.1002/HBM.25754

Zulkipli, N. Z. M., Mohamad, W. M. R. W., and Surat, S. (2019). Pengaplikasian Gaya Pembelajaran VAK Bagi Pembelajaran Budaya Melayu Dalam Kalangan Pelajar Di Tufs, Jepun. *J. Melayu Sedunia* 2, 112–142.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 May 2024
DOI 10.3389/fpsyg.2024.1276914

Check for updates

OPEN ACCESS

EDITED BY
Yiming Cao,
Beijing Normal University, China

REVIEWED BY
Xiaobo Xu,
Shanghai Normal University, China
Michael B. Steinborn,
Julius Maximilian University of Würzburg,
Germany

*CORRESPONDENCE
Dawei Liu
✉ dawei.liu@whu.edu.cn

RECEIVED 13 August 2023
ACCEPTED 19 January 2024
PUBLISHED 20 May 2024

CITATION
Chen S, Liu D, Yan H and Ma Y (2024) The effect of a Schema-based method on correcting persistent errors in mental arithmetic: an experimental study.
Front. Psychol. 15:1276914.
doi: 10.3389/fpsyg.2024.1276914

The effect of a Schema-based method on correcting persistent errors in mental arithmetic: an experimental study

Shufang Chen[1], Dawei Liu[2]*, Huifen Yan[3] and Yong Ma[2]

[1]College of Mathematics and Statistics, Hubei University of Education, Wuhan, China, [2]College of Teacher Education, Hubei University of Education, Wuhan, China, [3]Wuhan Luoyi Education Research Institute, Wuhan, China

Introduction: Arithmetic calculation is a fundamental skill for mathematical learning and daily life. However, elementary school students often make errors in practice.

Methods: Grounded in the schema theory and the memory retrieval theory of mental arithmetic, this study employs a controlled experiment to investigate the effect of a schema-based method in correcting persistent errors in mental arithmetic, specifically in the context of simple addition operations. The experimental group utilizes a schema-based method to help participants rectify incorrect answers in memory retrieval, while the control group did not receive this treatment.

Results: The results showed that significant differences emerged between the experimental and control groups in both the post-test performance and the reduction of persistent error count, indicating that the experimental group had rectified incorrect answers in memory; and persistent errors in simple addition were indeed caused by interference from incorrect answers during memory retrieval; and the schema-based method proves to be effective.

Discussion: The findings of this study contribute to enhancing practical mental arithmetic instruction, assisting students in correcting relevant errors, and improving their mental arithmetic abilities. Not only does it offer directive guidance for teaching practices, but it also provides an enlightening reference for promoting innovative teaching methods.

KEYWORDS

elementary school students, mental arithmetic, memory retrieval, schema theory, simple addition

1 Introduction

The arithmetic calculation, as a fundamental part of mathematics learning, serves as its foundation and helps students develop logical reasoning abilities. Meanwhile, it is also an essential life skill. Overly slow calculating speeds or frequently occurring arithmetic errors will inevitably lead to difficulties in their mathematics learning and present challenges for teachers in their instruction. National Council of Teachers of Mathematics, Commission on Standards for School Mathematics (2006) explicitly highlights the importance of early-acquired arithmetic ability in the elementary school curriculum and teaching. Therefore, improving the students' speed and accuracy in arithmetic calculations is of paramount importance for both mathematics education and the individuals' development.

Unlike general arithmetic calculations, mental arithmetic is a crucial cognitive activity in daily life, emphasizing the process of performing arithmetic operations without the use of external tools such as pens, paper, or calculators (Maclellan, 2001; Erdem, 2017). Grounded in cognitive structures of numerical understanding, mental arithmetic enhances comprehension of numerical systems and operations (Lemonidis, 2015; Jurić and Pjanić, 2023). Prior to the 1970s, cognitive psychology did not consider mental arithmetic as a research topic. Notwithstanding, some earlier works had explored the relationship between mathematics (arithmetic) and psychology. For example, the German philosopher and mathematician Husserl (1891) investigated the psychological foundation of mathematical thinking and how these thoughts were reflected in the logical structure of mathematics. Thorndike (1922) examined the process of human learning and understanding of arithmetic, as well as the application of psychology in education. However, it was not until the publication of the paper, *A chronometric analysis of simple addition* by Groen and Parkman (1972) that cognitive psychology officially recognized the beginning of its research on mental arithmetic, leading to gradually increasing interest in pertinent studies thereafter. Nowadays, researchers have recognized the importance of mental arithmetic and have successively made attempts to explore the impact of mental arithmetic abilities on elementary school students (Varol and Farran, 2007).

At the same time, considerable research findings about arithmetic errors have been revealed, such as the phenomenon of Post-Error Slowing, where individuals exhibit a slowing down after making errors in mental arithmetic. The slowing down reflects both internal reactions to errors and correction processes, with higher-accuracy individuals who focus more on internal reactions showing a more pronounced slowing effect (Steinborn et al., 2012). Recently, Sokolov (2023) has explored the reasons for arithmetic errors occurring during children's developing stage of mathematical concepts. Lestiana et al. (2017) delved into students' errors in fraction calculations, and classified the errors into three types based on Brown and Skow's (2016) framework: factual errors, procedural errors, and computational errors. In this context, the present study aims to investigate the effect of a schema-based method in rectifying persistent errors in mental arithmetic through a controlled experiment. The method is designed to address the issue of persistent errors, and will be illuminated herein, to provide a more comprehensive understanding of errors to research on mental arithmetic.

2 Literature review

2.1 Mathematical arithmetic errors

For the study on arithmetic errors, it is a crucial step to identify the types of errors (Reisman, 1972; Ashlock, 1986; Morales, 2014; Pankow et al., 2018). The systematic handling of errors made by students is a vital stage in teacher's work, involving the spotting and explanation of errors (Rozak and Nurwiani, 2020; Ngcobo, 2021; Sanina and Sokolov, 2021). In further research on arithmetic errors, scholars have summarized the error types, including Errors of perseveration, Errors of association, Errors of interference, knowledge errors, and comprehension errors, etc. (Radatz, 1979; Sarwadi and Shahrill, 2014). The classic classification of error mentioned in *Human Error* includes mistakes, false-rule applications, and slips of action (Reason, 1990).

Moreover, in mental arithmetic, simple addition serves as the foundation of arithmetic proficiency and the cornerstone for developing mathematical skills (Ashcraft, 1992; National Council of Teachers of Mathematics, Commission on Standards for School Mathematics, 2000; Peng et al., 2017). Obtaining the actual mastery of basic addition facts is one of the children's learning objectives (Office for Economic Co-operation and Development, 2014). The development of complex addition and multiplication skills relies on proficiency in simple addition (Geary and Brown, 1991; Geary et al., 2004). Thus, investigating errors in simple addition is essential for developing students' mental arithmetic abilities.

This paper aims to explore effective methods for addressing the problem of persistent errors in simple addition. As for persistent errors, Myers (1924) first introduced the concept. He made regular tests of basic arithmetic for students and found that they made the same errors repeatedly. For example, students might often write the result of $7+4$ as 10 instead of any other digit. This paper posits that this phenomenon is related to interference from incorrect answers during memory retrieval. Memory retrieval involves extracting information from memory (Jensen and Whang, 1994). When subjects solve arithmetic problems involving single-digit numbers, they retrieve arithmetic answers from long-term memory (Geary and Brown, 1991; FÜrst and Hitch, 2000; Noël et al., 2001; Geary et al., 2004; Swanson et al., 2008; Paul and Reeve, 2016).

2.2 The relationship between memory retrieval and persistent errors

The efficiency of mental calculation is associated with two critical factors: the organization of simple arithmetic facts in long-term memory, and the processing of information in working memory (De Rammelaere et al., 1999; Ding et al., 2019). The hypothetical model proposed by Jonides (1995) regarding the mental calculation process indicates that working memory and long-term memory are concurrently involved. Long-term memory provides various mental calculation knowledge and solution strategies, while working memory processes and stores information during mental calculation (Jonides, 1995; DeStefano and LeFevre, 2004). Research suggests that students are apt to develop their own mental arithmetic strategies rather than solely rely on the methods imparted at school. Given the fact that some students may adopt inefficient mental calculation strategies, learning systematic mental arithmetic strategies is supposed to be emphasized (Joung, 2018; Baranyai et al., 2019; Månsson, 2023).

For mental calculation, two classic strategies are commonly employed: counting strategy and memory retrieval strategy. The transition from the counting strategy to the memory retrieval strategy typically occurs in the third grade of primary school when half of the students use counting and the other half use memory retrieval, approximately. After the third grade, the role of memory retrieval will turn to be increasingly pronounced (Ashcraft and Fierman, 1982; Geary and Brown, 1991; Lucangeli et al., 2003). Existing memory retrieval models can be categorized into two types: table-search models and associate models. Table-search models propose that arithmetic knowledge in long-term memory is organized in the form of tables, where entries represent the sums of numbers (Ashcraft and Battaglia, 1978; Butterworth et al., 2001; Verguts and Fias, 2005). Therefore, the total steps depend on the sum of the numbers, which often leads to a slow and cumbersome process.

Associate models include network retrieval models, distribution of associations models, and network interference models. The network retrieval model posits that answers to simple addition problems are retrieved from a memory network. The addition network is represented as a matrix stored at the intersection points of entry nodes corresponding to the two addends (Ashcraft and Battaglia, 1978; Ashcraft, 1992). In this model, the difficulty in retrieving arithmetic answers depends on the strength of the association between numbers and answers which increases with the frequency of retrieval. The distribution of associations model suggests that the difficulty depends on the interference from other answers, originating from the stage of acquiring arithmetic skills and being related to incorrect answers (Siegler, 1988). The network interference model attempts to combine associative strength and interference to explain adult problem size effects while simulating the developmental changes in mental calculation skills, to elucidate the transition from counting to memory extraction (Zbrodoff, 1995).

Memory retrieval is considered as a strategy that can easily achieve high accuracy and high speed of calculation (Steel and Funnell, 2001; Ophuis-Cox et al., 2023). As evident from the previous discussion, the difficulty in memory retrieval is related to associative strength and interference. Therefore, it can be inferred that the persistent errors made by students are attributed to weak associations between equations and correct answers, coupled with interference from incorrect answers. To facilitate a more accurate and dexterous application of memory retrieval strategies, effective learning strategies or methods are required to guide children to encode arithmetic facts in long-term memory (Adesope et al., 2017; Agarwal et al., 2021; Ophuis-Cox et al., 2023). This study exactly employs a schema-based method to assist them in achieving this and investigates its specific effects on rectifying persistent errors. The following sections will further elaborate on the theories related to a schema-based method.

2.3 Mental arithmetic and schema theory

In his research, Bartlett found that participants tend to use their familiar schemas to understand new information, so he defined schema as "an active organization of past reactions or experiences of the individuals" (Bartlett and Bartlett, 1995). Piaget defined schema as "a structure or organization of actions, which, through repetition in similar or analogous situations, leads to the transfer or generalization of these actions" (Piaget, 2002), and demonstrated the critical role of schema in thinking and creative processes. Brown and Yule (1983) argued that schemata are well-organized background knowledge that guides the prediction of discourse content. In the 1970s, psychologist Rumelhart developed the concept of schemata in his schema theory which categorized schemata into three types: linguistic schemata, rhetorical schemata, and content schemata (Carrell and Eisterhold, 1983). Nowadays, the schema has become a core concept in cognitive psychology, widely applied in memory research to help understand advanced cognitive processes such as thinking.

The automation of schema and rule automation do not require additional resources and can compensate for the limited capacity of an individual's working memory, reducing cognitive load and improving transferability and transfer stability. In cognitive development theory, mental schema play an important role in cognitive processes, allowing for the construction of past experiences in a simplified manner, and allowing for spontaneous filtering, sorting, and organizing of external stimuli in new contexts to form organized modular knowledge structures. Through the recall of prior knowledge and experiences, individuals can gain symbolic meanings and emotional experiences of objective objects (Simonton, 2000). Thus, strengthening the construction of schema can help learners establish a sound cognitive structure. As can be extrapolated from the above demonstration, schema-based memory is highly efficient. Therefore, this study adopts the schema-based mental arithmetic method to correct persistent arithmetic errors in the participants' memories, and subsequently examines whether the phenomenon of persistent errors is improved.

According to the schema theory, schema is a high-level knowledge structure formed based on long-term memory, which represents an abstraction of things in life that a learner has experienced. Schema can bind multiple elements simultaneously, with overlapping and associative relationships being built between the elements (Gilboa and Marlatte, 2017). Therefore, based on the characteristics of numbers, students' cognitive features, and memory rules, the numbers in the arithmetic equation are encoded as real-life elements, and the links among the elements are also established to simulate the arithmetic relationships among the numbers. As shown in Figure 1, the red flag represents the number "4", the pencil represents the number "1", and the weighing hook represents the number "5". The equation represented by this image is $4+1=5$.

The schema-based mental arithmetic method not only possesses the superiority of schema but also visualizes arithmetic facts in an image-oriented manner, enabling students to connect mathematical expressions with real-life objects, thereby enhancing their memory retention. In the context of visualization research, in the late 1960s to early 1970s, Canadian psychologist Paivio (2013) proposed the Dual Coding Theory. He posited that the human cognitive system typically comprises two coding systems—verbal coding and non-verbal coding.

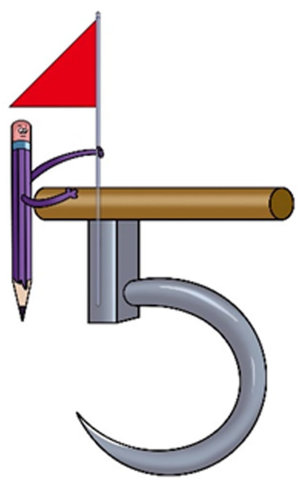


FIGURE 1
The image for equation $4+1=5$.

The two systems possess different forms and representational units for information perception. Experimental evidence indicates that a non-verbal coding system exhibits outstanding performance in the extraction and free recall of information. For instance, reading combining text and images show markedly higher efficiency compared to reading pure text. Different encoding of explicit knowledge plays distinct roles in the dissemination of information among individuals.

The essence of visual representation of knowledge lies in the explicit visual coding of knowledge products. The visual and the auditory coding of explicit knowledge products provide complements for each other, and simultaneously fundamental impetus together for visualized thinking. Eppler and Burkhard (2008) defined "knowledge visualization" as: "In general, the field of knowledge visualization research focuses on the application of visual representation in improving knowledge creation and transmission between two or among more individuals." Therefore, knowledge visualization can be considered as the application of visual representation to present knowledge in the form of charts, assisting in instructional design, and promoting the reflections of both teachers and students in teaching activities. Whether in traditional teaching with language as the primary medium or in visualization teaching aided by multimedia, the ultimate goal is to facilitate learners' knowledge acquisition and skill enhancement.

Thus, based on the schema theory, this paper just adopts the visualized strategy that uses images to present arithmetic facts as shown in Figure 1. Meanwhile, the images can stimulate neural activity through association, resulting in a significant enhancement in memory. This is because images enable learners to establish meaningful connections between visual and verbal information (Ahmadi and Zarei, 2021). Numerous studies indicate that transforming materials required to be memorized into visually intuitive images often results in better memory and recognition of images compared to mere text, highlighting the advantage of image memory (Ritchey, 1980). Image memory enables learners to engage in more cognitive activities, and deepens the information processing, rendering memories more stable and robust (Craik and Lockhart, 1972; Kobayashi, 1986).

3 Methods

3.1 Participants

The present study focused on elementary school students in grades 4, 5, and 6. The reason for the selection of these participants is to ensure that all the participants will use more memory retrieval strategies when performing calculations. Specifically, we selected 120 students from primary schools in rural areas of Tianmen City (in Hubei Province, China). These students were randomly assigned to either the experimental group or the control group, with each group comprising 60 students. The experimental group consisted of 25 boys and 35 girls, and the control group had 26 boys and 34 girls. The average ages of the control group and the experimental group are 10.06 and 10.25, with standard deviations of 0.97 and 1.34, respectively, indicating similarity in age between the two groups. The gender ratios for the two groups are 0.71:1 and 0.76:1, respectively, demonstrating similarity in gender distribution, as well. Additionally, all the participants have a similar level of educational attainment. In summary, the participants in the control group and experimental group are comparable in terms of demographic characteristics.

3.2 Instruments

3.2.1 Design of a test paper for single-digit addition

This experiment used an addition scale (refer to Appendix 1). The scale consisted of six sets of single-digit addition test questions, with the same questions in each set but in a different sequence of augend and addend, which were used to count the number of errors made by students on the same question and to detect the occurrence of persistent arithmetic errors. The test questions on this scale were used for both pretest and post-test in the experiment. But to avoid bias, the order of the six sets of test questions for the pretest and post-test were randomly assigned.

3.2.2 Creation of the schema-based teaching aids

Based on the cognitive load theory, knowledge exists in the form of cognitive schema in human's long-term memory. The purpose of learning is to intensify the training, automate cognitive activities, and reduce the cognitive load of an individual. The development of mental arithmetic schema teaching aids is to visualize the most basic combinations/blocks of arithmetic in mental arithmetic. The first step is to encode the numbers into common physical images through images or homophonic ways. For instance, the number "7" is shaped like a "sickle," the number "9" is shaped like a "balloon," and the pinyin "SHÍ LIÙ" for the number "16" can be encoded as a homophone of "SHÍ LIU" (pomegranate). The second step is to integrate the coded images of the three numbers in the mental arithmetic formula into a single picture through a brief story. It should be noted that the images of attend and augend should be significantly smaller than the image of the "sum." For example, the formula $7+9=16$ has three image elements (i.e., "sickle," "balloon," and "pomegranate," respectively), and the image of the pomegranate should be obviously larger than the other two. Next, combined with a story that "a sickle from the giant pomegranate dug out a balloon ", the final graphic aid has been developed with these images.

We have developed a total of 45 schemata (see Appendix 1) ranging from $1+1=2$ to $9+9=18$, arithmetic pictorials that contain all single-digit addition. In addition, because each graphic aid distinguishes between "addition" and "sum" by the size of the image, each schema actually represents four equations, e.g., Figure 2 represents $7+9=16$, $9+7=16$, $16–7=9$, and $16–9=7$. Compared to traditional mental arithmetic, the advantage of schema is manifested not only in reducing cognitive load and rendering learning process more interesting, but also in that it can integrate elements such as kinesthetic sense, five senses, images, interest, and uniqueness, etc., so as to deepen the connection to the working memory and transform the working memory to the long-term memory, thus enhancing the retention of the memory. What's more, such schema-based memorization is error-free and helpful in correcting persistent errors.

The employment of these schema-based aids for mental arithmetic teaching is relatively simple. First of all, let the students be familiar

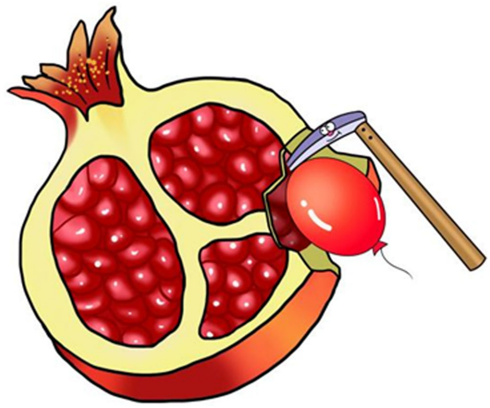


FIGURE 2
The image for equation 7 + 9 = 16.

with and memorize the number codes 1–18. Secondly, present the mental arithmetic schemata one by one to the students and encourage the students to create their own stories or scenarios according to the schemata (9 for each group) to mobilize the initiative of the students to deepen their working memory connection. Thirdly, ask the students to revert the schema to arithmetic equations according to the size and category of the elements; Finally, through calculation training, improve the students' proficiency in the inter-converting between schemata and arithmetic equations.

3.3 Experimental procedures

Before the experiment, a pretest was conducted on the experimental and control groups to identify the students' arithmetic errors. Research has shown that under the instruction to emphasize speed, students are more likely to use memory retrieval strategies (Ashcraft, 1982). Therefore, during the test, the instruction "at your fastest speed" was given, ensuring that most students chose memory retrieval strategies while doing calculations. Then, preliminary statistics were conducted on the results of the pretest, and students who exhibited persistent errors were selected for further experiments. In the experiment, a controlled trial was conducted. Graphic teaching aids were adopted in the experimental group used for simple addition teaching, where the teacher described and explained each graphic teaching aid and asked the students to repeat them for 20 min every day for a total training period of 15 days. The control group did not receive any training. Tests on the two groups were conducted in different classrooms, and teachers also took supervisory roles to prevent communication between the two groups of students. This effectively prevented the two groups from discovering their differences and achieved the placebo effect. After the experiment, a post-test was conducted on both groups, during which the instruction "at your fastest speed" was given again. The experimental group used graphic teaching aids. They used the schema-based method to rectify persistent errors in memory. By comparing the experimental group with the control group, if the former performs better, it proves that students' persistent errors in mental arithmetic are related to the interference of incorrect answers in memory retrieval.

3.4 Data collection and analysis

In this study, persistent errors were counted separately in the pretest and post-test for both the experimental and control groups. A fixed incorrect answer that emerged two or more times for the same question was considered as a persistent error. For example, if a student answered $3+5=7$ three times, $8+1=10$ two times, and $2+5=8$ two times, he would be recorded as having made three persistent errors. Statistical analysis was conducted with the help of SPSS software on four variables:

1 group: experimental and control groups;
2 the number of persistent errors made in the pretest by each student;
3 the number of persistent errors made in the post-test by each student;
4 the reduction in the number of persistent errors from pretest to post-test.

First of all, a significance test was conducted on the pretest results of both groups to ensure that the initial arithmetic levels of students were comparable. Likewise, a significance test was also needed in terms of the post-test results of both groups. Finally, one more significance test was made on the reduction in the number of persistent errors between the experimental and control groups to further examine the experimental effect.

4 Experimental results

Both the experimental group and the control group consist of 60 individuals. A preliminary analysis was conducted on the number of students whose pretest results showed persistent errors. It turned out to be that 16 participants in the experimental group (26.67%) and 17 participants in the control group (28.33%) exhibited persistent errors during the pretest. The similar percentage of participants facing persistent errors in both groups in the pretest indicated that there was indeed a phenomenon of persistent arithmetic errors among students and that the initial arithmetic levels of the two groups were comparable. Next, the students with persistent errors were picked out to form the experimental and control groups which were comprised of 16 and 17 students, respectively. The assignment would remain unchanged for subsequent experiments.

4.1 Significance test

4.1.1 Pretest evaluation

To begin with, whether the data for the number of errors in the pretest was normally distributed was checked. Since the sample size was $N=33<5{,}000$, we used the Shapiro–Wilk test at first and yielded the W-value (0.60), $p<0.001$, which indicated that the data did not follow a normal distribution. Therefore, we turned to using the Mann–Whitney U-test for independent samples.

As shown in Table 1, both the median values for the pretest frequencies in the control group and experimental group are 1.0. The Median Absolute Deviation (MAD) of 0 indicates no difference between the two groups at the median level. Their respective mean

TABLE 1 Analysis results table for pretest and post-test assessments.

Variable name	Variable value	Sample size	Mean	Standard deviation	*u*	*p*	MAD	Cohen's *d*
The number of errors in the pretest	Control group	17	1	0.61	141.5	0.80	0	0.07
	Experimental group	16	1.31	0.60				
The number of errors in the post-test	Control group	17	1	0.66	231.5	<0.001	1	1.59
	Experimental group	16	0.31	0.48				

Significance level is set at 0.05, and a *p*-value less than 0.05 is considered statistically significant.

values are 1.0 and 1.31, indicating a slight difference between the two groups at the mean level. The standard deviations are 0.61 and 0.60 for the control group and experimental group, suggesting a minimal disparity in the dispersion of data. The results of the U-test showed a *p*-value of 0.80 and a u-value of 141.5, indicating non-significance. The Cohen's *d* value for the difference is 0.07. When Cohen's *d* value is less than 0.2, it denotes a small effect; between 0.2 and 0.5, it signifies a medium effect; while around 0.8 or larger, a large effect is considered (Cohen, 1988). Therefore, Cohen's *d* of 0.07 suggested a small effect size, indicating a very small difference. Consequently, there was no significant difference between the control and experimental groups in pretest frequencies, which showed that there was no substantial disparity in the initial arithmetic proficiency between the students in the two groups, ensuring that the experimental results were not unduly influenced by significant differences in students' arithmetic proficiency.

4.1.2 Post-test evaluation

In the first place, the normality of the data was tested to determine whether it was normally distributed. Since the sample size for the number of errors was $N = 33 < 5{,}000$, the Shapiro–Wilk test was first adopted. The significance level was found to be $p < 0.001$, and the *W*-value was 0.79, indicating a significant departure from normality. Therefore, the Mann–Whitney *U*-test should be employed.

As shown in Table 1, the median values for post-test frequencies in the control group and experimental group are 1.0 and 0.0, respectively. The Median Absolute Deviation of 1 suggests a certain difference between the two groups at the median level. Their respective mean values are 1.0 and 0.31, indicating a significant difference between the two groups at the mean level. Meanwhile, the results from the experimental group exhibited mean values of 1.31 and 0.31 in the pretest and post-test, respectively, suggesting a reduction in error frequencies, and thus a favorable outcome. The standard deviations are 0.66 and 0.48 for the control group and experimental group, signifying a certain level of disparity in the dispersion of data between the two groups. The data distribution in the experimental group is relatively concentrated. The *U*-test results yielded a *p*-value of less than 0.001, and the u-value was 231.5, demonstrating statistical significance. The effect size, measured by Cohen's *d*, was 1.59, denoting a very large magnitude of difference. Therefore, there was a significant difference between the control and experimental groups in post-test frequencies. It can be easily observed that the graphic aids have effectively corrected the persistent errors in the memory of experimental group students. Furthermore, as most students rely on memory retrieval when doing mental arithmetic, it provided proof from the reverse side for the former hypothesis that the occurrence of persistent errors in the mental calculating process is associated with interference from incorrect answers in memory retrieval.

4.1.3 Test on reduction of error count

Primarily, the data was checked whether it followed a normal distribution with the Shapiro–Wilk test since the sample size of the error reduction count was $N = 33 < 5{,}000$. The *W*-value was 0.82, $p < 0.001$, indicating a significant departure from normality and negating the null hypothesis. Likewise, we conducted a Mann–Whitney *U*-test then.

As shown in Table 2, the median reduction in the number of errors for the control group and the experimental group are 0.0 and 1.0, The Median Absolute Deviation of 1 suggests a marked difference between the two groups at the median level. Their means are 0 and 1, with the experimental group exhibiting more reduction in the number of errors compared to the control group. This indicated there was a significant decrease in error rates for the experimental group, suggesting a positive experimental effect. Their standard deviations were 0.49 and 0.63, indicating a certain level of disparity in the variability between the two groups, with the data distribution in the control group being relatively concentrated. The test result was $p < 0.001$, and the *u*-value is 43.5, indicating statistical significance. The effect size, Cohen's *d* value, was 1.572, which suggests a very large difference between the control and experimental groups in terms of decreased errors. These results demonstrate jointly that the experimental group's persistent errors were significantly reduced with the help of the schema-based teaching aids whereas the control group's persistent errors remained largely unchanged. This further confirms the initial hypothesis.

5 Discussion and conclusion

5.1 Discussion

This study aimed to investigate the effect of the schema-based method in correcting students' persistent errors in mental arithmetic practice. Through results analysis of a controlled experiment, the results revealed that correcting students' persistent errors in mental arithmetic through the schema-based method is effective. Despite the initial similarity between the experimental and control groups in pretests, significant disparity was found in post-tests. And the reduction degree in the number of errors also showed significant differences. This illustrates a significant decrease in persistent error rates in the experimental group during the post-test, unraveling a positive experimental effect compared to the control group. It points

TABLE 2 Mann–Whitney *U*-test for reduction in the number of persistent errors for the control and experimental groups.

Variable name	Variable value	Sample size	Mean	Standard deviation	*u*	*p*	MAD	Cohen's *d*
The reduction in the number of errors	Control group	17	0	0.49	43.5	<0.001	1	1.57
	Experimental group	16	1	0.63				

Significance level is set at 0.05, and a *p*-value less than 0.05 is considered statistically significant.

to the fact that the experimental group effectively corrected persistent arithmetic errors in memory through the schema-based method.

Although Myers (1924) had ever studied similar error phenomena, the cause of the phenomenon was ignored. This study, through a controlled experiment, testified the hypothesis in a reverse way and investigated the persistent errors which were not ever mentioned in previous studies on arithmetic error types (Engelhardt, 1977; Radatz, 1979; Herholdt and Sapire, 2014; Sarwadi and Shahrill, 2014; Muthukrishnan et al., 2019). When a student makes an arithmetic error only once, it may be a random one. However, when he gives the same incorrect answer multiple times for the same arithmetic problem, it is unlikely to be random. Therefore, the author hypothesized that such persistent errors in mental calculation are caused by interference from incorrect answers in memory retrieval. This hypothesis is based on the fact that students are apt to retrieve answers from memory (memory retrieval) while doing mental calculations. There have been many studies on memory retrieval models for mental calculation and arithmetic error types, but none has linked memory retrieval models to persistent errors in mental calculation before this study. In this research, by rectifying the incorrect answers in memory retrieval, the negative interference can be removed and the persistent arithmetic errors can be rectified. Meanwhile, the accuracy and efficiency of mental arithmetic can be improved, which would provide significant help for teaching and learning activities.

The schema-based mental arithmetic method is implemented through non-verbal encoding by visualizing arithmetic facts and connecting mathematical concepts with real-life objects for students. The use of schema can be viewed as a form of non-verbal encoding, aligning with the relevant theory proposed by Paivio (2013). In the learning process, through visual representation, students can gain a deeper understanding and memorization of arithmetic facts, as the images can provide meaningful associations and contexts. The concept of knowledge visualization further supports the schema-based mental arithmetic method. In this method, presenting arithmetic facts with images is exactly an application of knowledge visualization. It can not only aid teachers in the designing of more effective instruction but also encourage students to actively make reflections during educational activities. By combining graphic design and cognitive science, the schema-based mental arithmetic method endows the students with a more engaging and efficient learning way. Moreover, the research results mentioned in this paper support the advantages of visual mnemonics. Transforming material that needs to be memorized into visually intuitive images helps deepen information processing and enhance memory stability. The emphasis on visual memory in cognitive activities diverges from traditional text-centered teaching. To be specific, the positive role of visual memory in learners' cognitive activities is highlighted, aligning with the schema theory that focuses on perception and cognition in learning.

As for why visual representation of numbers through images can be feasible and effective, the underlying cognitive mechanisms warrant more consideration. This study hypothesizes a potential association with the brain's associative functions. O'Malley and Chamot (1990) proposed associative memory strategies, referring to linking new information with acquired knowledge or establishing connections between individuals and information to enhance understanding and memorization of learning materials, similar to the schema-based method employed in this study, in which numbers are linked to daily experiences and objects by resorting to images that closely resemble real-life objects to represent every single number and combining these images to form mental representations of arithmetic facts. Consequently, students can automatically convert these images into arithmetic facts through association and store them in long-term memory. When solving similar arithmetic problems next time, students can swiftly associate these mental representations of arithmetic facts to retrieve answers.

5.2 Unique contribution and future outlook

Many scholars have engaged in in-depth discussions on the best teaching methods for mathematics, giving rise to a variety of instructional strategies (Gulson, 2021). The unique contribution of this study is evident in its profound impact on future educational strategies. Teachers can actively apply schema-based teaching tools, and integrate them into mental arithmetic instruction, to help students correct persistent errors and enhance their mental arithmetic abilities. This study provides valuable guidance, practical and theoretical, to arithmetic education. By further exploring these findings, a more comprehensive understanding of the learning process can be gained, which helps lay a solid foundation for the advent of more innovative teaching methods and strategies, thus driving continuous progress in the field of arithmetic education. In the future, emphasis should be placed on teacher training to enhance their proficiency in schema-based teaching. As teachers play a guiding role in the implementation of schema-based arithmetic teaching, their expertise and teaching skills are essential for the success of schema-based instruction. Additionally, further research should delve into the impact of schema-based teaching on students' learning interests, motivation, and attitudes, as well as how to stimulate students' initiative and self-directed learning through schema-based instruction. This will contribute to optimizing relevant pedagogy, and improving student's learning experiences and performance.

This study simultaneously fills the gaps in previous research. Scholars have explored the impact of schema-based instruction on mathematical problem-solving abilities (Fuchs et al., 2004). Xin's (2008) research aimed to explore the effects of schema-based teaching strategies on the arithmetic applications of algebraic concepts among primary school students who encounter learning obstacles or issues (LP). Mudrikah and Hakim (2016) implemented Problem-Based Learning (PBL) combined with Action, Process, Object, Schema (APOS) theory for 26 prospective mathematics teachers. Gembong and Suwarsono (2018) study aimed to investigate the schemata established by elementary school students while solving fraction addition problems. In Gulson's study Gulson (2021), teachers can embed abstract concepts like equation solving within students' existing knowledge structures using schema theory. In summary, schema theory has garnered extensive attention in the field of mathematics education. However, there has been relatively less specific application of schema theory in correcting continuous errors in mental arithmetic within these studies. Our research fills this knowledge gap, underscoring the potential value of schema theory in rectifying mental arithmetic. Furthermore, our study's outcomes not only confirm the effectiveness of schema-based methods in mental arithmetic but also advance the development of schema theory. Through experimental results, we may offer new insights into how schema influences the cognitive processes and learning outcomes in mental arithmetic. This paves an interesting path for future research, delving deeper into the specific mechanisms of schema-based methods in mental arithmetic and potentially triggering more profound theoretical discussions on mental arithmetic instructional approaches.

Despite valuable findings yielded in this study, there are still various shortcomings in its research design and results analysis. Firstly, in the experiment, the relatively small number of participants who are primarily concentrated in rural schools in fixed regions limits the external validity of the results. Future research can enhance external validity by increasing the number of participants and diversifying the distribution of source regions. Additionally, extending the testing period and conducting long-term observations can more comprehensively validate the long-standing reliability of experimental results. Moreover, the experiment still embraces uncontrolled factors, such as the intentional behaviors of students, which may have influenced the results more or less. To better understand the results, future research can employ more strict and detailed control measures to eliminate potential confounding factors. While this study confirms the effect of the schema theory in addressing persistent errors in mental arithmetic, the applicability to other error types remains to be validated.

This study intends to further develop schema-based teaching tools in different subjects under existing conditions and explore the application effects and implementation strategies of schema-based teaching in various subjects and grade levels. Since the characteristics and requirements of learning differ across subjects and grades, customized design and optimization of schema-based teaching models become necessary. In the development of teaching aids, with the rapid advancement of information technology, visualization techniques and tools have gained widespread application (Yuancheng, 2018). Future research directions may include how to effectively combine visualization techniques with schema-based teaching to create more diverse and personalized teaching resources and tools. In summary, valuable insights can be provided for future research and educational practices by reviewing the limitations of this study, and guidance for further advancement in this field is also available.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by cognitive psychology of Brain Science and Learning Science Committee of Hubei Teachers Education Association. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. The animal study was approved by cognitive psychology of Brain Science and Learning Science Committee of Hubei Teachers Education Association. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

SC: Data curation, Writing – original draft. DL: Methodology, Writing – review & editing. HY: Resources, Writing – review & editing. YM: Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Hubei Provincial Education Science Plan 2022 Special Funding Project (No. 2022ZB23) and Hubei Provincial Education Science Plan 2023 Funding Project (No. 2023ZA036).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1276914/full#supplementary-material

References

Adesope, O. O., Trevisan, D. A., and Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. *Rev. Educ. Res.* 87, 659–701. doi: 10.3102/0034654316689306

Agarwal, P. K., Nunes, L. D., and Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. *Educ. Psychol. Rev.* 33, 1409–1453. doi: 10.1007/s10648-021-09595-9

Ahmadi, M., and Zarei, A. A. (2021). On the effects of linguistic, verbal, and visual mnemonics on idioms learning. *Lang. Rel. Res.* 12, 279–303.

Ashcraft, M. H. (1982). The development of mental arithmetic: a chronometric approach☆. *Dev. Rev.* 2, 213–236. doi: 10.1016/0273-2297(82)90012-0

Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition* 44, 75–106. doi: 10.1016/0010-0277(92)90051-I

Ashcraft, M. H., and Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *J. Exp. Psychol.: Learn. Mem.* 4, 527–538.

Ashcraft, M. H., and Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *J. Exp. Child Psychol.* 33, 216–234. doi: 10.1016/0022-0965(82)90017-0

Ashlock, R. B. (1986). *Error patterns in computation: A semi-programmed approach*. Princeton, NC: Merrill Publishing Company.

Baranyai, T., Egri, E., Molnár, A. É., and Zsoldos-Marchis, I. (2019). "Mental calculation strategies used by pre-service primary school teachers" in *Edulearn19 proceedings* (IATED), 8717–8724.

Bartlett, F. C., and Bartlett, F. C. (1995). *Remembering: a study in experimental and social psychology*. Cambridge: Cambridge University Press.

Brown, J., and Skow, K. (2016). Mathematics: identifying and addressing student errors. in *The Iris Center*. United States of America: The Iris Center, 31

Brown, G., and Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge university press.

Butterworth, B., Zorzi, M., Girelli, L., and Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *Q. J. Exp. Psychol. Section A*, 54, 1005–1029.

Carrell, P. L., and Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Q.* 17, 553–573. doi: 10.2307/3586613

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (*2nd*). Hillsdale, NJ: Lawrence Earlbaum Associates.

Craik, F. I., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

De Rammelaere, S., Stuyven, E., and Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychol. Res.* 62, 72–77. doi: 10.1007/s004260050041

DeStefano, D., and LeFevre, J. (2004). The role of working memory in mental arithmetic. *Eur. J. Cogn. Psychol.* 16, 353–386. doi: 10.1080/09541440244000328

Ding, Y., Liu, R., Liu, H., Wang, J., Zhen, R., and Jiang, R. (2019). Effects of working memory, strategy use, and single-step mental addition on multi-step mental addition in Chinese elementary students. *Front. Psychol.* 10:148. doi: 10.3389/fpsyg.2019.00148

Engelhardt, J. M. (1977). Analysis of CHILDREN'S computational errors: a qualitative approach. *Br. J. Educ. Psychol.* 47, 149–154. doi: 10.1111/j.2044-8279.1977.tb02340.x

Eppler, M. J., and Burkhard, R. A. (2008). "Knowledge visualization" in *Knowledge management: Concepts, methodologies, tools, and applications* (Pennsylvania: IGI Global), 987–999.

Erdem, E. D. (2017). Mental computation: evidence from fifth graders. *Int. J. Sci. Math. Educ.* 15, 1157–1174. doi: 10.1007/s10763-016-9722-1

Fuchs, L. S., Fuchs, D., Prentice, K., Hamlett, C. L., Finelli, R., and Courey, S. J. (2004). Enhancing mathematical problem solving among third-grade students with schema-based instruction. *J. Educ. Psychol.* 96, 635–647. doi: 10.1037/0022-0663.96.4.635

FÜrst, A., and Hitch, G. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Mem. Cogn.* 28, 774–782. doi: 10.3758/BF03198412

Geary, D. C., and Brown, S. C. (1991). Cognitive addition: strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Dev. Psychol.* 27, 398–406. doi: 10.1037/0012-1649.27.3.398

Geary, D. C., Hoard, M. K., Byrd-Craven, J., and Desoto, M. C. (2004). Strategy choices in simple and complex addition: contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002

Gembong, S., and Suwarsono, S. T. (2018). Schema building profiles among elementary school students in solving problems related to operations of addition to fractions on the basis of mathematic abilities. *J. Phys. Conf. Ser.* 983:012155. doi: 10.1088/1742-6596/983/1/012155

Gilboa, A., and Marlatte, H. (2017). Neurobiology of schemas and Schema-mediated memory. *Trends Cogn. Sci.* 21, 618–631. doi: 10.1016/j.tics.2017.04.013

Groen, G. J., and Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychol. Rev.* 79, 329–343. doi: 10.1037/h0032950

Gulson, R. (2021). Using schema theory to reduce cognitive load in stage 4 equation solving. *Teach. Math.* 46, 27–33.

Herholdt, R., and Sapire, I. (2014). An error analysis in the early grades mathematics - a learning opportunity? *South Afr. J. Childh. Educ.* 4:19. doi: 10.4102/sajce.v4i1.46

Husserl, E. (1891). *Philosophie der Arithmetik: Psychologische und logische Untersuchungen [philosophy of arithmetic: Psychological and logical investigations]*. Halle: Verlag Pfeffer.

Jensen, A. R., and Whang, P. A. (1994). Speed of accessing arithmetic facts in long-term memory: a comparison of Chinese-American and Anglo-American children. *Contemp. Educ. Psychol.* 19, 1–12. doi: 10.1006/ceps.1994.1001

Jonides, J. (1995). Working memory and thinking. *Think. Inv. Cogn. Sci.* 3, 215–265.

Joung, E. (2018). *A study of preservice Teachers' mental computation attitudes, knowledge, and flexibility in thinking for teaching mathematics* Doctoral dissertation Carbondale: Southern Illinois University.

Jurić, J., and Pjanić, K. (2023). An overview of mental calculation strategies and the frequency of their application. *J. Element. Educ.* 16, 277–294. doi: 10.18690/rei.16.3.3087

Kobayashi, S. (1986). Theoretical issues concerning superiority of pictures over words and sentences in memory. *Percept. Mot. Skills* 63, 783–792. doi: 10.2466/pms.1986.63.2.783

Lemonidis, C. (2015). *Mental computation and estimation: Implications for mathematics education research, teaching and learning*. Routledge.

Lestiana, H. T., Rejeki, S., and Setyawan, F. (2017). Identifying students' errors on fractions. *J. Res. Adv. Math. Educ.* 1, 131–139. doi: 10.23917/jramathedu.v1i2.3396

Lucangeli, D., Tressoldi, P. E., Bendotti, M., Bonanomi, M., and Siegel, L. S. (2003). Effective strategies for mental and written arithmetic calculation from the third to the fifth grade. *Educ. Psychol.* 23, 507–520. doi: 10.1080/0144341032000123769

Maclellan, E. (2001). Mental calculation: its place in the development of numeracy. *Westminst. Stud. Educ.* 24, 145–154. doi: 10.1080/0140672010240205

Månsson, A. (2023). Preservice elementary teachers' mental computation strategy use in multiplication on one-and two-digit natural numbers. *Acta Didactica Napocensia* 16, 99–108. doi: 10.24193/adn.16.1.7

Morales, Z.A. (2014). *Analysis of students' misconceptions and error patterns in mathematics: The case of fractions*. Education Commons. Available at: http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1350&context=sferc

Mudrikah, A., and Hakim, L. L. (2016). Problem-based learning associated by action process object schema theory in mathematics instruction. *Proc. AICS-Soc. Sci.* 6, 367–374.

Muthukrishnan, P., Kee, M. S., and Sidhu, G. K. (2019). Addition error patterns among the preschool children. *Int. J. Instr.* 12, 115–132. doi: 10.29333/iji.2019.1228a

Myers, G. C. (1924). Persistence of errors in arithmetic. *J. Educ. Res.* 10, 19–28. doi: 10.1080/00220671.1924.10879475

National Council of Teachers of Mathematics, Commission on Standards for School Mathematics. (2000).*Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics, Commission on Standards for School Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics*. Reston, VA: Author.

Ngcobo, Z. (2021). Error analysis: supporting the development of pre-service mathematics teachers' subject matter knowledge of school mathematics topics. *Int. J. Learn. High. Educ.* 28, 151–168. doi: 10.18848/2327-7955/CGP/v28i01/151-168

Noël, M., Désert, M., Aubrun, A., and Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Mem. Cogn.* 29, 34–42. doi: 10.3758/BF03195738

Office for Economic Co-operation and Development. (2014). *In PISA 2012 results: What students know and can do: Student performance in mathematics, reading and science* (*I*). Paris, France: OECD.

O'malley, J. M., and Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press

Ophuis-Cox, F. H., Catrysse, L., and Camp, G. (2023). The effect of retrieval practice on fluently retrieving multiplication facts in an authentic elementary school setting. *Appl. Cogn. Psychol.* 37, 1463–1469. doi: 10.1002/acp.4141

Paivio, A. (2013). *Imagery and verbal processes*. New York: Psychology Press.

Pankow, L., Kaiser, G., König, J. A., and Blömeke, S. (2018). Perception of student errors under time limitation: are teachers faster than mathematicians or students? *ZDM* 50, 631–642. doi: 10.1007/s11858-018-0945-1

Paul, J. M., and Reeve, R. A. (2016). Relationship between single digit addition strategies and working memory reflects general reasoning sophistication. *Learn. Instr.* 42, 113–122. doi: 10.1016/j.learninstruc.2016.01.011

Peng, P., Yang, X., and Meng, X. (2017). The relation between approximate number system and early arithmetic: the mediation role of numerical knowledge. *J. Exp. Child Psychol.* 157, 111–124. doi: 10.1016/j.jecp.2016.12.011

Piaget, J. (2002). *Judgement and reasoning in the child*. London: Routledge.

Radatz, H. (1979). Error analysis in mathematics education. *J. Res. Math. Educ.* 10, 163–172. doi: 10.2307/748804

Reason, J. T. (1990). *Human error*. Cambridge: Cambridge University Press.

Reisman, F. K. (1972). *A guide to the diagnostic teaching of arithmetic*. Columbus, OH: Charles E. Merrill Publishing Company.

Ritchey, G. H. (1980). Picture superiority in free recall: the effects of organization and elaboration. *J. Exp. Child Psychol.* 29, 460–474. doi: 10.1016/0022-0965(80)90107-1

Rozak, A.Nurwiani (2020). Statistical reasoning levels and error analysis of prospective mathematics teacher students. *J. Phys.*:1464.

Sanina, S., and Sokolov, V. (2021). *Approaches to the typology of the common mistakes of younger schoolchildren in the development of mathematical concepts*. Современная зарубежная психология.

Sarwadi, H. R., and Shahrill, M. (2014). Understanding Students' mathematical errors and misconceptions: the case of year 11 repeating students. *Math. Educ. Trends Res.* 2014, 1–10. doi: 10.5899/2014/metr-00051

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *J. Exp. Psychol. Gen.* 117, 258–275. doi: 10.1037/0096-3445.117.3.258

Simonton, D. K. (2000). Creativity: cognitive, personal, developmental, and social aspects. *Am. Psychol.* 55, 151–158. doi: 10.1037/0003-066X.55.1.151

Sokolov, V. (2023). *Psychological and pedagogical methods of working on the mistakes of younger schoolchildren in the development of mathematical concepts*. Вестник практической психологии образования.

Steel, S., and Funnell, E. (2001). Learning multiplication facts: a study of children taught by discovery methods in England. *J. Exp. Child Psychol.* 79, 37–55. doi: 10.1006/jecp.2000.2579

Steinborn, M. B., Flehmig, H. C., Bratzke, D., and Schrö ter, H. (2012). Error reactivity in self-paced performance: highly-accurate individuals exhibit largest post-error slowing. *Q. J. Exp. Psychol.* 65, 624–631. doi: 10.1080/17470218.2012.660962

Swanson, H. L., Jerman, O., and Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *J. Educ. Psychol.* 100, 343–379. doi: 10.1037/0022-0663.100.2.343

Thorndike, E. L. (1922). *The psychology of arithmetic*. New York: Macmillan.

Varol, F., and Farran, D. (2007). Elementary school students' mental computation proficiencies. *Early Childhood Educ. J.* 35, 89–94. doi: 10.1007/s10643-007-0173-8

Verguts, T., and Fias, W. (2005). Interacting neighbors: a connectionist model of retrieval in single-digit multiplication. *Mem. Cogn.* 33, 1–16. doi: 10.3758/BF03195293

Xin, Y. P. (2008). The effect of schema-based instruction in solving mathematics word problems: an emphasis on prealgebraic conceptualization of multiplicative relations. *J. Res. Math. Educ.* 39, 526–551. doi: 10.5951/jresematheduc.39.5.0526

Yuancheng, W.. *Research on the knowledge visualization design of Chinese language*. Guangzhou: Jinan University (2018).

Zbrodoff, N. J. (1995). Why is 9+ 7 harder than 2+ 3? Strength and interference as explanations of the problem-size effect. *Mem. Cogn.* 23, 689–700. doi: 10.3758/BF03200922