Education reimagined: the impact of advanced technologies on learning

Edited by
Mohammad Khalil, Mohammed Saqr, Sonsoles López-Pernas
and Isabel Hilliger

Published in
Frontiers in Psychology
Frontiers in Education

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

Education reimagined: the impact of advanced technologies on learning

Topic editors

Mohammad Khalil — University of Bergen, Norway

Mohammed Saqr — University of Eastern Finland, Finland

Sonsoles López-Pernas — University of Eastern Finland, Finland

Isabel Hilliger — Pontificia Universidad Católica de Chile, Chile

Table of
contents

Predicting students' continued intention to use E-learning platform for college English study: the mediating effect of E-satisfaction and habit

Ping Deng[1], Bing Chen[2]* and Li Wang[3]

[1]Basic teaching department, Hezhou University, Hezhou, China, [2]School of Foreign Languages for Business, Guangxi University of Finance and Economics, Nanning, China, [3]School of Foreign Languages, Hunan International Economics University, Changsha, China

Using technology in education facilitates knowledge dissemination expediently while broadening and deepening learning modes and content diversity. As an information technological innovation, E-learning platform is widely used to learn college English. However, few studies have explored the motivations for students' e-satisfaction and continued intention towards using it for college English study. Based on the extended Unified Theory of Acceptance and Use of Technology (UTAUT2), this study identifies the influencing factors for the continued usage intention and tests the mediating role of e-satisfaction and habit. Six hundred and twenty-six usable responses from Guangxi were analyzed with partial least squares structural equation modelling. Results show that performance expectancy, learning value, hedonic motivation and habit positively affects students' continued usage intention, e-satisfaction positively mediates the relationship between the antecedents and continued usage intention, and habit positively mediates the relationship between e-satisfaction and continued usage intention. The research provides guidelines for the successful implementation of e-learning platform for college English study and key references for improvement of students' engagement and satisfaction experience with using e-learning platform for college English study.

KEYWORDS
UTAUT2, e-learning platform, learning value, e-satisfaction, habit

Introduction

The application of information technology in education has greatly promoted the development of higher education. In particular, it has led to the creation of many efficient and economic information technological tools that are designed to facilitate learning and deliver knowledge without time and place restrictions. E-learning platform (ELP) is such an innovation driven primarily by information technology. ELP provides a versatile and economic method to learn and share knowledge with the help of electronic devices and an internet connection at any time and place (Osei et al., 2022). ELP offers asynchronous and synchronous communication channels for students to undertake online learning according to their own learning time schedule, and it plays a great role in enhancing learning performance in higher education (Ain et al., 2016; Prasetyo et al., 2021). ELP has even replaced traditional face-to-face teaching during

the COVID-19 pandemic (He, 2020; Zou et al., 2021), providing a new way for students to learn college English, which is a vital part of Chinese higher education to develop students' abilities, knowledge, and overall quality.

The use of ELP has revolutionized education systems and changed learning styles all over the world (Raman and Thannimalai, 2021). As a result, educational institutions have invested heavily in accelerating and supporting educational informationization (Ain et al., 2016). However, despite the investment and effort, the results have not always been as beneficial and desirable as expected. In some cases, students have reported low satisfaction, low participation, and low learning efficiency when using ELP (Cai et al., 2020; Li et al., 2020; Qiao et al., 2021; Xu et al., 2022). Ultimately, the success of ELP depends on whether students are satisfied with their experience of using it and are willing to continue using it in future studies (Ismail et al., 2016). This highlights the importance of initial acceptance and continued intention to use the technology in the long term (Bhattacherjee et al., 2008).

While many studies have examined factors influencing university students' initial adoption of ELP (Raman and Thannimalai, 2021; Tandon et al., 2021; Xu et al., 2022; Zacharis and Nikolopoulou, 2022), few have explored the factors contributing to their continued use, particularly in college English study. In order to address this gap in the literature, the present study aims to delve deeper to identify the factors that influence students' continued use of ELP for college English study. Moreover, while previous research has investigated the impact of antecedents on customers' satisfaction and adoption of new technologies (Alalwan, 2020; Santosa et al., 2021; Siyal et al., 2021), few have examined the mediating role of e-satisfaction and habit in the relationship between antecedents and continued use in college English study (Rai, 2020). Therefore, this study will also explore the mediating effect of e-satisfaction and habit in the relationship between predictors and continued usage intention. By identifying the key factors that affect students' continued intention to use ELP and the mediating role of e-satisfaction and habit, this study will provide a more comprehensive understanding of the factors that would increase the successful implementation of ELP and contribute to students' sustained use of ELP for college English study. The findings of this research will offer practical implications for educators, policymakers, and ELP developers, who can better understand the factors and design effective strategies to promote the continued use of ELP for college English learning.

Theoretical background and hypothesis development

Theoretical background

The Unified Theory of Technology Acceptance and Use of Technology (UTAUT) is a widely used model that was proposed to explain and predict the use of new technologies (Venkatesh et al., 2003). The original UTAUT model included four constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions. However, research later found that three additional constructs could improve the model's predictive and explanatory power: hedonic motivation, price value, and habit (Venkatesh et al., 2012). As a result, the UTAUT2 model was created,

which can now explain over 70% of the variance for the intention to adopt and use new technologies (Venkatesh et al., 2012).

In the educational context, a large amount of studies have used this model to investigate the effect of antecedents on students' and teachers' intention to adopt and use new technologies (Cacciamani et al., 2018; Dajani and Abu Hegleh, 2019; Nikolopoulou et al., 2020; Raman and Thannimalai, 2021; Zacharis and Nikolopoulou, 2022). However, few studies have investigated students' continued usage intention. Continued usage intention refers to the user's willingness to continue using a new technology over an extended period of time, even after the initial adoption and use phase (Santosa et al., 2021; Abbasi et al., 2022; Gao, 2023; Maduku and Thusi, 2023). In the context of UTAUT2, continued intention is measured as the intention to continue using a technology for a specific purpose, such as using an ELP for college English study. Understanding the factors that affect continued intention is important as it can impact the long-term success and sustainability of the technology. This research seeks to bridge this gap by adopting the UTAUT2 model to study the factors that impact students' continued use of ELP for college English study.

In order to more thoroughly examine and validate the potency of UTAUT2, Venkatesh et al. (2012) emphasized the importance of incorporating constructs that can reflect the very nature of the targeted subject in specific contexts. In an effort to better investigate the factors that impact students' continued usage intention, the concept of price value is replaced with that of learning value, as students do not need to pay to benefit from ELP; rather, they invest their time and energy to study and obtain skills and knowledge through ELP (Ain et al., 2016). Previous research has shown that learning value has a significant and positive influence on students' intention to use new technology in the learning context (Zacharis and Nikolopoulou, 2022). Furthermore, given that few studies have explored the mediation role of students' e-satisfaction for the relationship between antecedents and continued usage intention and the mediating effect of habit on the relationship between e-satisfaction and continued usage intention, this study attempts to fill this gap by incorporating learning value into the extended UTAUT2 model and investigates the mediating effect of e-satisfaction and habit to better understand the relationship between the antecedents and continued willingness to use ELP for college English study. By doing so, a more comprehensive and nuanced understanding of the factors that affect students' usage intention of ELP can be obtained.

Hypotheses development

Performance expectancy (PE) refers to the perception of how much a new technology can improve a user's performance or how beneficial it is in completing certain activities (Venkatesh et al., 2012). Research has shown that PE strongly influences a user's willingness to continue using new technologies (Santosa et al., 2021; Sasongko et al., 2022; Wu et al., 2022). Users who perceive the benefits of a new technology are more likely to continue using it over time. In the learning context, it is assumed that students are more willing to continuously use ELP if they believe that it is useful and helpful in performing various English study activities, and it enables them to achieve desirable outcomes more efficiently. Therefore, this research posits that:

H1: PE positively impact students' continued intention to use ELP for college English study.

Effort expectancy (EE) refers to individuals' beliefs about the effort or ease of using a particular new technology, or their perception of how easy or difficult it is to use a new technology (Venkatesh et al., 2003). This concept is important because it influences whether users will continue to use the technology in the future. Research has shown that EE has a positive impact on users' willingness to use a new technology in the future (Yan et al., 2021; Abbasi et al., 2022). This study examines whether students perceive the operation instructions of ELP as easy to understand and clear, and whether they find it convenient and effortless to use when completing English tasks and activities. If students feel that the technology is easy to use and does not require much mental or physical effort, they are more likely to continue using it in their future college English studies. Therefore, the following hypothesis can be proposed:

H2: EE positively impacts students' continued intention to use ELP for college English study.

Social influence (SI) refers to the impact of the external environment, or others' beliefs, on an individual's willingness to use a particular new technology (Venkatesh et al., 2003). The influence mainly comes from the social pressure of the external environment surrounding the individual. Several studies have shown the impact of SI on individuals' decision and continued intention to use new technologies. For instance, Gao (2023) found that SI is a crucial factor that affects individuals' decision to adopt new technologies. Similarly, Venkatesh et al. (2003, 2012) and Zacharis and Nikolopoulou (2022) have also found that SI plays an vital role in individuals' intention to use new technologies. In this research, SI mainly refers to the influence of peers, family, friends and teachers' ideas on the use of ELP. If external factors and beliefs are positive regarding students' intention to use ELP for college English study, then their continued usage intention would be enhanced. The following hypothesis investigates this relationship:

H3: SI positively impacts students' continued intention to use ELP for college English study.

Facilitating conditions (FC) refer to the availability of technical support and resources provided by the organization to support the use of a new technology (Venkatesh et al., 2003). This can include things like assistance from IT personnel, access to necessary software and hardware, and a stable internet connection. Without these resources, users may be hindered from using a new technology on a continuous basis (Nanayakkara, 2007). However, if FC are present, users are inclined to continue using the technology (Tandon et al., 2021; Gao, 2023). In this research, FC specifically refers to the accessibility of technical support, tools, facilities, stable internet connection, and other compatible technologies that support the use of ELP for college English study. If students have access to these resources and technical support, as well as other forms of support from their universities, they will be more motivated to continue using ELP for college English study. Therefore, the following hypothesis can be posited:

H4: FC positively impacts students' continued intention to use ELP for college English study.

Hedonic motivation (HM) refers to the perceived happiness and enjoyment that users experience while utilizing a particular technology. According to the experience economy theory, offering unique services and experiences with enjoyment and fun is crucial for winning customers' hearts (Pine and Gilmore, 2011). Gupta and Dogra (2017) and Coves-Martínez et al. (2023) have both confirmed that the higher level of perceived enjoyment from a new technology can increase the continuous use of the technology. When students become interested in using ELP to learn college English, it can stimulate an internal driving force that encourages them to use the platform for their English studies in the future. Therefore, it can be hypothesized that the higher the level of HM that students experience while using ELP to learn college English, the more likely they are to continue using it for their language studies in the future.

H5: HM positively impacts students' continued intention to use ELP for college English study.

Learning value (LV) is used to replace price value in this research. Price value represents the trade-off between the benefits and sacrifices of using a new technology. When using a new information technology, user often weighs the perceived benefits against the costs associated with its use (Venkatesh et al., 2012). From a consumer's perspective, a product has value if it offers some benefits, while from a learning perspective, learning value is defined as the cognitive trade-off students make between the perceived value of using a new technology for study and the time and effort taken for using it (Ain et al., 2016). While students do not typically have to pay to use new technology, they do have to invest their time and energy in order to gain knowledge and learning value (Ain et al., 2016). It is important for any new technology to offer significant benefits that make the investment of time and energy worthwhile. When the students perceive that the investment of time and energy leads to significant improvements in learning outcomes, they are more likely to invest more time and effort into using it, leading to increased continuous intention (Prasetyo et al., 2021; Zacharis and Nikolopoulou, 2022). In this research, if the use of ELP can improve students' language skills and proficiency, they are inclined to continue using it for college English study. Therefore, the hypothesis can be made as follows:

H6: LV positively impacts students' continued intention to use ELP for college English study.

Habit (HB) refers to individuals' automatic or habitual use of new technologies. This behavioral performance is cultivated unconsciously and automatically based on experience gained from a series of previous behaviors (Venkatesh et al., 2012). Moreover, this habitual behavior contributes to the formation of cognitive commitment for specific behavior, which is gradually formed but not easy to change (Murray and Häubl, 2007). After an extended period of time, the automatic behavior will achieve a relatively stable and continuous state (Venkatesh et al., 2012). This means that a person's habitual use of technology will become a consistent and regular behavior that is difficult to change. In the learning context, with regular and repeated use of ELP for course participation, forum discussion, assignment submission, examination, and grade checking, students will develop a habitual positive behavior unconsciously. Habitual positive behavior,

formed through regular and repeated use of ELP, increases students' intention to use it in the long run (Raman and Thannimalai, 2021; Tandon et al., 2021; Xu et al., 2022). Therefore, it can be hypothesized that:

> *H7*: HB positively impacts students' continued intention to use ELP for college English study.

The mediating effect of E-satisfaction (ESA)

Originally, e-satisfaction referred exclusively to customers' satisfaction with their past shopping experience on an electronic commerce company's platform (Anderson and Srinivasan, 2003). However, the current research employs e-satisfaction to measure the degree of satisfaction that students have towards using ELP for their college English studies. Previous studies indicate that the perceived usefulness (PE), ease of use (EE), social approval and support (SI), accessibility of technical support and necessary resources (FC), users' perceived trade-off between profits and cost (PV), and perceived enjoyment and pleasure (HM) from using a new technology have a direct effect on users' satisfaction with their experience of using it (Alalwan, 2020; Siyal et al., 2021; Wu et al., 2022; Mishra et al., 2023). In the context of college English study, with social and technical support from others, the ease of using ELP for English study can not only help students finish English learning tasks more efficiently and improve academic performance (LV) with little effort (EE), but also increase their feelings of happiness and pleasure. This would positively influence their satisfaction with using ELP for college English study. Based on the above information, PE, EE, SI, FC, LV and HM can increase students' e-satisfaction with using ELP for college English study. If ELP can offer real benefits to students and match their expectations, they are inclined to have high e-satisfaction with using it for college English study. Accordingly, they are more motivated to continue using it for college English study in the future(Kim et al., 2019; Alalwan, 2020; Wu et al., 2022; Maduku and Thusi, 2023; Perez-Aranda et al., 2023). Therefore, the following hypotheses can be posited.

> *H8*: ESA mediates the relationship between (a) PE, (b) EE, (c) SI, (d) FC, (e) LV, (f) HM and students' continued intention to use ELP for college English study.

The mediating effect of habit

E-satisfaction refers to the user's evaluation of an information system in terms of whether it can reach their needs and expectations (Zeithaml and Bitner, 2003). Research shows that users' e-satisfaction contributes to habit development and long-term engagement (Tran and Trang, 2018). Favorable user experience plays a key role in developing use habits (Thadani and Cheung, 2011; Wang et al., 2013). Once a habit is formed, users become inclined to automatically and habitually repeat it, making it even more difficult to suppress the habit (Aarts and Dijksterhuis, 2000). The stronger the habit, the more determined the user is to keep using the new technology (Baudier

et al., 2018; Gu et al., 2019). If users are satisfied with their early experience of using a new technology, they are more inclined to develop automatic and spontaneous behavior towards using it, making them more likely to continue using it in the future (Amoroso and Lim, 2017; Alalwan, 2020). In the context of learning, if ELP can offer real benefits to students and meet their expectations, they are more likely to have high e-satisfaction towards using it for college English study. With accumulated learning experience, habit could be developed, which naturally increases their continued intention to use it for college English learning in future. Thus, It can be posited as:

> *H9*: HB mediates the relationship between students' e-satisfaction and their continued intention to use ELP for college English study.

Research design and methodology

Participants

ELP has become an indispensable part of Chinese higher education, enabling students to continue their studies effectively. This research focuses on undergraduate students, as they are significant in the Chinese higher education system. They were learning college English as part of educational programs with ELP at different public universities across Guangxi. As this study explores the factors that impact students' continued intention to use ELP for college English study without revealing any specific personal information or human performance, there is no need for ethical agreement. With the help of college English teachers, a participant information form was presented to all participants to ensure transparency and informed participation during class interval. Following this, a participant consent form was provided to get participants' consent. Only after that, can participants start to answer the questionnaire by scanning the QR code. The students participate voluntarily in this survey and their confidentiality was ensured. They are encouraged to answer the questions according to their real using experience and thoughts about application of ELP for college English study, and all survey data will be used only for research aims.

Instrument development

This research aims to identify the factors that affect students to continue using the ELP for college English study. To achieve this goal, a quantitative approach is adopted, and data is collected online using an online questionnaire. It includes three parts: the first part explains the nature and purpose of the research and includes a filtering question to confirm whether students have used ELP for college English study; the second part is for demographic information, including gender, age, discipline or majors, year currently studying, number of years' experience in using ELP; and the third part contains the measurement items for all constructs in this research. The constructs of UTAUT2 were adapted from Ain et al. (2016) and Venkatesh et al. (2012), while the items for e-satisfaction were adapted from Alalwan (2020) and Wu et al. (2022) and those for continued intention from Dağhan and Akkoyunlu (2016). All items were modified to better fit this study. To ensure that the measured items

accurately reflect the concept of each construct, this study invited information technology personnel and college English teaching experts to evaluate the face and content validity of the questionnaire. Back-translation was done for all questions to maintain the equivalence and consistency of the questionnaire in different languages (English and Chinese). After pilot testing with a group of students and some English teachers, the questionnaire has improved its clarity and quality (see Table 1). A 5-point scale measured the items, with 1 indicating strong disagreement and 5 indicating strong agreement.

Data collection

Data is collected online through Tencent Questionnaire, a survey platform that allows for the creation of customized questionnaires at a relatively lower cost. Eight hundred and sixty-five respondents completed the questionnaire, answers with consistently neutral responses, diagonal lining responses, and alternating extreme pole responses were deleted. This resulted in 626 valid questionnaires (response rate: 72.4%), including 294 males (47%) and 332 females (53%). The sample had a considerable representation of students from different levels of education; 34.5% were freshmen, 39.1% were sophomores, 20.4% were juniors, and 5.9% were seniors. The students were between the ages of 19 and 23 years old. When it comes to online learning experience, most of the students had 2 to 3 years' experience in using ELP. Majors covers engineering, art, history, law, economics, management, education, philosophy, literature, and so on. According to the recommended '10 times rule'(Wu, 2010), which suggests that the sample size should be ten times the number of indicators, 626 samples were sufficient to proceed further with data analysis.

Data analysis

This study uses UTAUT2 to identity the factors that impact students' continued intention to use ELP for college English study and the mediating effect of e-satisfaction and habit. The proposed model, as shown in Figure 1, displays the variable relationships. The partial least square structural equation modelling (PLS-SEM) was used for data analysis. PLS-SEM is a powerful analytical approach to examine structural equation modeling as it is good at analyzing both indirect and direct effects of the mediating effect (Hair et al., 2017) and it can handle complex models with many structural model relations (Shiau et al., 2019). PLS-SEM model assessment focuses on the evaluation of measurement model and structural model. The former describes the relationships between constructs and their indicators, and the latter deals with the relationships between different constructs and how they interact with each other. In this research, data analysis was conducted in three steps with SmartPLS 3.3: evaluation of measurement model, evaluation of structural model, and test of mediating effect.

Research results

Evaluation of measurement model

The evaluation of the measurement model examines the reliability, convergent validity, and discriminant validity of each construct. As shown in Table 2, all Cronbach's alpha values (0.748～0.924) and composite reliability values (0.855～0.952) are higher than 0.7, indicating sufficient internal consistency reliability for all constructs. The factor loading of all constructs (0.784～0.941) is also higher than 0.7, indicating that the measurement items fully captures the concept of each construct and there is acceptable reliability for each construct. Additionally, the average variance extracted (AVE) of all constructs (0.645～0.869) is higher than 0.5, indicating good convergent validity for all constructs.

The Fornell-Larcker criterion and heterotrait-monotrait ratio (HTMT) can be used to assess discriminant validity of the constructs (Hair et al., 2017). As shown in Table 3, the diagonal values represent the square root of the AVE of each construct, and each construct's value is greater than its correlation with other constructs in any column or row, indicating good discriminant validity between the constructs of the measurement model. HTMT assesses the correlation between constructs by comparing the mean of all correlations for measurement items between different constructs and that within the same construct (Henseler et al., 2015). As shown in Table 4, all HTMT correlation values were lower than 0.85, and the 95% confidence intervals do not include the value 1 after bootstrapping, both indicating that there is good discriminant validity for all constructs in the measurement model.

After evaluating the measurement model, the constructs demonstrated satisfactory levels of quality. Prior to evaluating the structural model, variance inflation factor (VIF) was examined to check for collinearity issue. As can be seen in Table 5, all VIF values were below 3.3 (Diamantopoulos and Siguaw, 2006). Therefore, There is no collinearity problem for the structural model in this research.

Evaluation of structural model

In order to predict the variance in the dependent variables, PLS-SEM assessment of the structural model evaluates R^2, Q^2, f^2, and the size and statistical significance of the structural path coefficients (Hair et al., 2017). As shown in Table 5, the R^2 value for e-satisfaction, habit, and continued intention is 0.616, 0.464, and 0.662, respectively. This implies that antecedents can substantially explain the variance in students' e-satisfaction and their continued intention to use ELM for college English study (Chin, 1998). Q^2 value for e-satisfaction, habit, and continued intention is 0.410, 0.371 and 0.569, respectively, all are larger than zero, indicating that the independent variables have predictive relevance for these dependent variables. Based on the f^2 values from Table 6, e-satisfaction($f^2 = 0.156$) exerts the largest impact on college students' continued willingness to use ELM for college English study, and hedonic motivation($f^2 = 0.162$) has the largest effect on students' e-satisfaction with the application of ELM for college English study. To effectively evaluate the structural model, bootstrapping with 5,000 subsamples is set with two tails.

From the path coefficients in Table 6, it is evident that PE, LV, HM, and HB have a positive impact on students' continuous intention to use ELM for college English study, with HB showing the strongest influence, followed by LV, HM, and PE. All t-values are >1.96, all p-values are <0.05, and their 95% confidence intervals do not contain the value zero, indicating that all these relationships are significant. Therefore, PE, LV, HM, and HB have a significant and positive impact on students' willingness to continue using ELM for college English study, and hence, *H1*, *H5*, *H6*, and *H7* are supported. On the other

TABLE 1 Measurement items.

Construct	Measurement items	Source
Performance expectancy (PE)	I find e-learning platform(ELP) useful for my English studies.	Venkatesh et al. (2012) and Ain et al. (2016)
	ELP allows me to accomplish English class activities more quickly.	
	ELP increases my English learning productivity.	
Effort expectancy (EE)	ELP is easy to use.	
	Learning how to use ELP is easy for me.	
	My interaction with ELP is clear and understandable.	
Social influence (SI)	My peers who influence my behavior think that I should use ELP.	
	My friends who are important to me think that I should use ELP.	
	My instructors whose opinions that I value prefer that I should use ELP.	
Facilitating conditions (FC)	I have resources to use ELP.	
	I have knowledge to use ELP.	
	A specific person (or group) is available to assist when difficulties arise with ELP.	
Learning value (LV)	ELP is worth more than the time and effort given to it.	
	In less time, ELP allows me to quickly and easily share knowledge with others.	
	ELP gives me the opportunity to decide about the pace of my own learning.	
	ELP gives me opportunity to increase my knowledge and control my success.	
Hedonic motivation (HM)	I feel fun using ELP to learn college English.	
	I enjoy using ELP to learn college English.	
	Using ELP to learn college English is very entertaining.	
Habit (HB)	Using ELP to learn English has become a habit for me.	
	I am addicted to using ELP to accomplish my English study tasks.	
	I must use ELP for my English studies.	
Continued intention (CI)	I intend to continue using ELP for English study in the future.	Dağhan and Akkoyunlu (2016)
	I will continue using ELP for English study in the future.	
	I will regularly use ELP for English study in the future.	
E-satisfaction (ESA)	I am normally willing to adopt ELP for English study.	Alalwan (2020) and Wu et al. (2022)
	I am extremely pleased with ELP for English study.	
	I am joyful to adopt ELP for English study.	
	I am pleased with the way that ELP has carried out for English study.	

hand, EE, SI, and FC do not have a significant effect on students' willingness to continue using ELM for college English study. In the path from EE to CI, the t-value (0.154) is <1.96, the p-value (0.878) is above 0.05, and the 95% confidence intervals [−0.086, 0.071] contains the value zero, indicating that EE does not have a significant effect on CI, and therefor, H2 is not supported. In the path from SI to CI, the t-value (0.151) is <1.96, the p-value (0.880) is >0.05, and the 95% confidence interval [−0.056, 0.048] contains the value zero, indicating that SI does not have a significant effect on CI, and therefore, H3 is not supported. In the path from FC to CI, the t-value (0.096) is <1.96, the p-value (0.923) is >0.05, and the 95% confidence interval [−0.077, 0.073] contains the value zero, indicating that FC does not significantly impact CI, and therefore, H4 is not supported.

Mediating effect analysis

This research follows Zhao et al. (2010) procedures for mediation analysis, using 5,000 bootstrapping samples. A mediation analysis was conducted to assess the impact of e-satisfaction on the relationship between PE, EE, SI, FC, LV, HM, and CI, as well as the effect of habit on the relationship between e-satisfaction and CI. All results are displayed in Table 7.

Regarding the three mediating paths of PE → ESA → CI, LV → ESA → CI, and HM → ESA → CI, the results indicate that their indirect and direct effects are significant. Additionally, all 95% confidence intervals of both indirect and direct effects exclude zero, indicating that e-satisfaction positively and partially mediates these relationships. To determine the type of partial mediation, the product of the direct effect (positive) and the indirect effect (positive) is positive, which supports the idea that e-satisfaction represents complementary partial mediation for these three paths. These findings support H8a, H8e, and H8f.

In the paths EE → ESA → CI, SI → ESA → CI, and FC → ESA → CI, all indirect effects are significant while direct effects are insignificant. Furthermore, the 95% confidence intervals for the indirect effects exclude zero, while the direct effects include zero. Therefore, e-satisfaction fully mediates the relationships along these three paths. As a result, H8b, H8c, and H8d are supported.
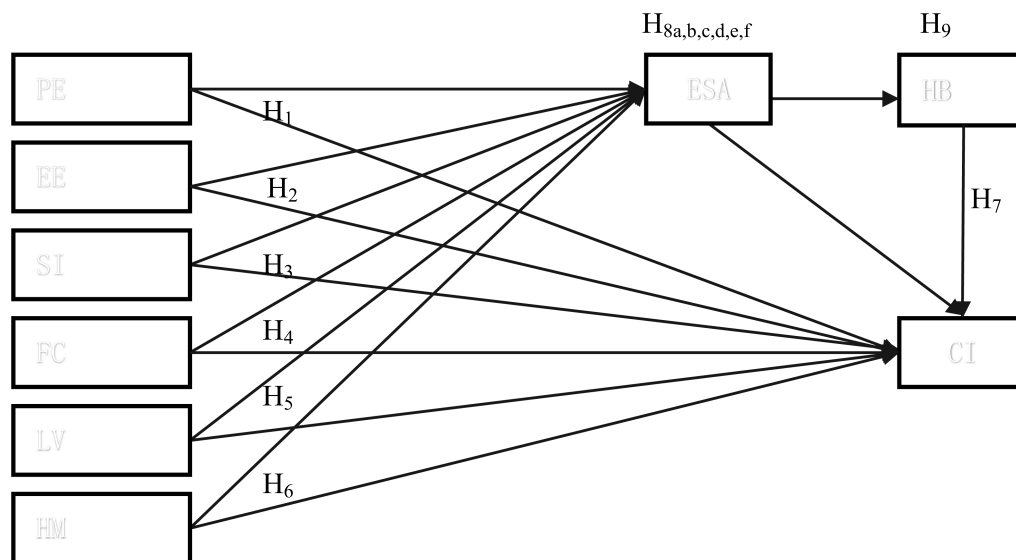
FIGURE 1
Theoretical framework.

In the path *ESA → HB → CI,* the 95% confidence intervals for both the indirect effect [0.056, 0.175] and the direct effect [0.273, 0.505] exclude zero. This indicates that *habit* partially mediates this relationship, since the indirect and direct effects are significant. The fact that the product of the direct effect (positive) and the indirect effect (positive) is positive indicates that habit reveals complementary partial mediation for this path, which supports *H9*.

Discussion

Based on the extended UTAUT2, this research has conducted a thorough analysis of the factors that impact students' e-satisfaction and willingness to continue using ELP for college English study. In addition to identifying these determinants, the study has also tested the mediating role of e-satisfaction and habit. The model has high explanatory power for the variance of students' e-satisfaction and continued intention to use ELP for college English study. The research results show that PE, LV, HM, and HB significantly affects students' willingness to continue using ELP to learn college English, while EE, SI, and FC have no significant impact. E-satisfaction mediates the relationship between PE, EE, SI, FC, LV, HM, and CI, while habit positively mediates the relationship between e-satisfaction and CI.

More specifically, PE significantly influences students' intention to continue using ELP for college English study. This finding is in confirmation with the results of previous research by Chen et al. (2018), Mishra et al. (2023), and Wu et al. (2022), which held that the cognitive and functional usefulness are important factors for saving time and effort. The more efficiently that students perceived they can use ELP to complete class activities and assignments, the higher their intention to continue using it for further study. Therefore, it is crucial for educators to emphasize the importance of using ELP in the classroom to maximize its benefits and encourage its continued use among students. And what's more, the result provides valid reason to

recommend ELP developers to further develop the technologies and improve the functions for students to increase learning efficiency in college English study.

The findings of this study demonstrate that LV positively influences CI. This result corroborates the findings of previous studies, such as those conducted by Ain et al. (2016), Dajani and Abu Hegleh (2019), Prasetyo et al. (2021), and Zacharis and Nikolopoulou (2022), which held that the use of ELP for college English study increases students' perception of academic value, leading to greater willingness to continue using it in the future. This is particularly important given the increasing importance of English proficiency in the global job market and the need for students to have access to effective language learning tools. By taking advantage of ELP, students can improve their English language skills and gain a competitive edge in their future careers. Additionally, the use of ELP has been shown to have broader benefits beyond simply improving language proficiency, such as enhancing critical thinking skills and increasing cultural awareness. Therefore, it is important for educators to recognize the potential of ELP and incorporate it into teaching strategies to increase students' participation in college English learning.

The results of the current research suggest that HM exerts a significant and positive impact on CI, which is consistent with previous research results from Alalwan (2020) and Liu et al. (2022), showing that the students are willing to continue using ELP in the future if they derive pleasure and enjoyment from using it during their college English studies. It is worth noting that the use of ELP contributes to a more immersive and stimulating language learning experience, which can further enhance students' language proficiency. The results of this research can have important implications for educators who are trying to find effective ways to enhance students' motivation and engagement for better learning outcomes.

The research results indicate that HB was the most significant factor in predicting CI to use ELP for college English study. This finding confirms with previous studies by Veeramootoo et al. (2018),

TABLE 2 Reliability and validity results of constructs.

Constructs	Items	Loading	CA	CR	AVE
Performance expectancy (PE)	PE1	0.903	0.893	0.933	0.823
	PE2	0.913			
	PE3	0.907			
Effort expectancy (EE)	EE1	0.875	0.837	0.902	0.753
	EE2	0.858			
	EE3	0.871			
Social influence (SI)	SI1	0.794	0.748	0.855	0.663
	SI2	0.798			
	SI3	0.849			
Facilitating conditions (FC)	FC1	0.860	0.824	0.895	0.740
	FC2	0.877			
	FC3	0.843			
Learning value (LV)	LV1	0.805	0.817	0.879	0.645
	LV2	0.784			
	LV3	0.816			
	LV4	0.806			
Hedonic motivation (HM)	HM1	0.913	0.907	0.942	0.843
	HM2	0.917			
	HM3	0.925			
Habit (HB)	HB1	0.907	0.882	0.927	0.809
	HB2	0.883			
	HB3	0.907			
E-satisfaction (ESA)	ESA1	0.809	0.847	0.896	0.683
	ESA2	0.829			
	ESA3	0.845			
	ESA4	0.823			
Continued intention (CI)	CI1	0.932	0.924	0.952	0.869
	CI2	0.923			
	CI3	0.941			

CA, Cronbach's alpha; CR, composite reliability; AVE, average variance extracted.

Gu et al. (2019), Tandon et al. (2021), and Xu et al. (2022) which have also demonstrated the substantial impact of HB on CI. The importance of fostering positive habitual behavior towards the use of ELP for college English study cannot be overstated. By developing such positive habits, students will be more motivated to use ELP for college English study in the future. Additionally, it might be worthwhile to investigate ways to encourage the development of such habits and to identify factors that might hinder their development. Therefore, it is crucial for universities to offer support and adequate resources to help students develop positive attitudes and create a supportive learning environment that encourages ELP use in college English study.

However, the empirical results of this research failed to confirm the role of EE, SI, and FC in predicting CI. It is worth noting that there may be several reasons for this. With regard to EE, one possible reason is that students have attached more importance to the usefulness of ELP (Prasetyo et al., 2021). In other words, if students perceive a high level of value and benefits, they can overcome any difficulties in using this technology (Davis et al.,

TABLE 3 Discriminant validity with Fornell-Larcker criterion.

	Continued intention	Effort expectancy	Facilitating conditions	Habit	Hedonic motivation	Learning value	Performance expectancy	E-satisfaction	Social influence
Continued intention	0.932								
Effort expectancy	0.518	0.868							
Facilitating conditions	0.525	0.609	0.86						
Habit	0.679	0.47	0.506	0.899					
Hedonic motivation	0.694	0.531	0.561	0.706	0.918				
Learning value	0.693	0.607	0.639	0.655	0.725	0.803			
Performance expectancy	0.617	0.603	0.524	0.568	0.619	0.685	0.907		
E-satisfaction	0.758	0.575	0.575	0.681	0.719	0.699	0.606	0.827	
Social influence	0.322	0.372	0.323	0.318	0.343	0.399	0.522	0.291	0.814

TABLE 4 Discriminant validity with HTMT.

	Continued intention	Effort expectancy	Facilitating conditions	Habit	Hedonic motivation	Learning value	Performance expectancy	E-satisfaction
Effort expectancy	0.584 CI95% [0.505, 0.658]							
Facilitating conditions	0.601 CI95% [0.517, 0.678]	0.733 CI95% [0.673, 0.790]						
Habit	0.749 CI95% [0.688, 0.804]	0.543 CI95% [0.465, 0.613]	0.591 CI95% [0.511, 0.664]					
Hedonic motivation	0.757 CI95% [0.704, 0.805]	0.606 CI95% [0.530, 0.674]	0.648 CI95% [0.572, 0.718]	0.788 CI95% [0.737, 0.834]				
Learning value	0.792 CI95% [0.729, 0.847]	0.732 CI95% [0.655, 0.801]	0.771 CI95% [0.695, 0.839]	0.764 CI95% [0.708, 0.815]	0.837 CI95% [0.791, 0.879]			
Performance expectancy	0.679 CI95% [0.612, 0.741]	0.691 CI95% [0.616, 0.758]	0.61 CI95% [0.527, 0.685]	0.638 CI95% [0.573, 0.695]	0.687 CI95% [0.628, 0.740]	0.805 CI95% [0.752, 0.856]		
E-satisfaction	0.844 CI95% [0.798, 0.885]	0.678 CI95% [0.608, 0.741]	0.685 CI95% [0.613, 0.751]	0.777 CI95% [0.716, 0.831]	0.811 CI95% [0.758, 0.861]	0.824 CI95% [0.763, 0.878]	0.686 CI95% [0.621, 0.744]	
Social influence	0.382 CI95% [0.288, 0.469]	0.466 CI95% [0.379, 0.551]	0.408 CI95% [0.308, 0.502]	0.387 CI95% [0.298, 0.469]	0.412 CI95% [0.323, 0.494]	0.513 CI95% [0.423, 0.596]	0.637 CI95% [0.552, 0.710]	0.348 CI95% [0.255, 0437]

TABLE 5 Results of R^2, Q^2 and VIF.

Construct	R^2	Q^2	VIF							
			PE	EE	SI	FC	LV	HM	HB	ESA
CI	0.662	0.569	2.527	2.047	1.404	2.009	3.181	2.948	2.385	2.788
ESA	0.616	0.410	2.471	1.999	1.388	1.983	2.989	2.299		
HB	0.464	0.371								1.000

1992). Another reason is that students are proficient in using various new information technologies, and ELP is not a complex system for completing English class activities and tasks. This suggests that students may have found ELP to be quite useful and easy to use, which could explain the lack of significance in the relationship between EE and CI. Regarding SI, one reason may be that ELP is the best way to continue studying during the pandemic period, and students can realize the importance of taking responsibility for their own learning without the influence of peers, friends, teachers, and classmates (Tandon et al., 2021). This self-motivated approach could explain why the impact of social influence on ELP adoption is not as significant as initially hypothesized. It is also worth noting that the impact of social influence may diminish over time, as students gain more experience with using new information technologies (Alalwan, 2020). However, it is important to consider that some students may still require social support and encouragement to adopt ELP. Therefore, it is

important to provide a supportive learning environment that encourages students to use ELP for college English study. With regard to FC, it is true that ELP is quite simple from the perspective of knowledge dissemination (Xu et al., 2022). Moreover, with adequate experience using different new technologies, students are less likely to be affected by FC to keep using it continuously (Venkatesh et al., 2012). This suggests that FC may be less important when students have sufficient experience using similar technologies, which could explain the lack of significance in the relationship between FC and IC. However, it is important to consider that some students may require additional resources and support to ensure that they are able to continue using ELP for college English study, for example, providing training sessions and technical support can help students overcome any difficulties they may encounter.

The partial mediating effect of e-satisfaction on the relationship between PE, LV, HM, HB, and CI implies that these relationships are not direct causal ones. Instead, PE, LV, HM, and HB influence

TABLE 6 Direct relationship results and structural model results.

H	Path	Coef.	t value	p value	95% Bias-corrected IC	f²	Result
H₁	PE → CI	0.115	2.742	0.006	[0.031, 0.198]	0.016	Yes
H₂	EE → CI	−0.006	0.154	0.878	[−0.086, 0.071]	0.000	No
H₃	SI → CI	−0.004	0.151	0.880	[−0.056, 0.048]	0.000	No
H₄	FC → CI	−0.004	0.096	0.923	[−0.077, 0.073]	0.000	No
H₅	LV → CI	0.152	3.176	0.002	[0.057, 0.243]	0.022	Yes
H₆	HM → CI	0.123	2.627	0.009	[0.036, 0.221]	0.015	Yes
H₇	HB → CI	0.172	3.953	0.000	[0.083, 0.252]	0.037	Yes
	PE → ESA	0.127	3.029	0.002	[0.047, 0.209]	0.017	
	EE → ESA	0.129	3.613	0.000	[0.059, 0.199]	0.022	
	SI → ESA	−0.075	2.354	0.019	[−0.139, −0.014]	0.010	
	FC → ESA	0.096	2.416	0.016	[0.022, 0.179]	0.012	
	LV → ESA	0.227	4.551	0.000	[0.123, 0.321]	0.045	
	HM → ESA	0.378	8.419	0.000	[0.288, 0.462]	0.162	
	ESA → CI	0.383	6.437	0.000	[0.273, 0.505]	0.156	
	ESA → HB	0.681	25.651	0.000	[0.622, 0.728]	0.867	

*p < 0.05; ** < 0.01; ***p < 0.001 (two-tails).

TABLE 7 Mediating analysis.

H	Path	Total indirect effects/P1xP2				Direct effects/P3				Mediation type	Result
		Coef.	t value	p value	95% Bias-corrected IC	Coef.	t value	p value	95% bias-corrected IC		
H8a	PE → ESA → CI	0.049	2.815	0.005	[0.019, 0.088]	0.115	2.742	0.006	[0.031, 0.198]	Complementary partial mediation	Yes
H8b	EE → ESA → CI	0.05	3.243	0.001	[0.023, 0.084]	−0.006	0.154	0.878	[−0.086, 0.071]	Full mediation	Yes
H8c	SI → ESA → CI	−0.029	2.091	0.037	[−0.061, −0.006]	−0.004	0.151	0.88	[−0.056, 0.048]	Full mediation	Yes
H8d	FC → ESA → CI	0.037	2.426	0.015	[0.010, 0.071]	−0.004	0.096	0.923	[−0.077, 0.073]	Full mediation	Yes
H8e	LV → ESA → CI	0.087	3.800	0.000	[0.048, 0.139]	0.152	3.176	0.002	[0.057, 0.243]	Complementary partial mediation	Yes
H8f	HM → ESA → CI	0.145	4.648	0.000	[0.093, 0.216]	0.123	2.627	0.009	[0.036, 0.221]	Complementary partial mediation	Yes
H9	ESA → HB → CI	0.117	3.845	0.000	[0.056, 0.175]	0.383	6.437	0.000	[0.273, 0.505]	Complementary partial mediation	Yes

e-satisfaction, which in turn influences CI. The fact that the partial mediation of e-satisfaction implies that it explains some, but not all of these relationships. This raises the possibility that other potential mediating variables, which were not examined in this study, may further clarify the nature of these relationships. In contrast, e-satisfaction fully accounts for the relationship between EE, SI, FC, and CI, as shown by the full mediating effect of e-satisfaction. This finding supports the hypothesized theoretical framework for these relationships. The partial mediating effect of habit on the relationship between e-satisfaction and CI implies that the relationship is not a direct causal one. Rather, e-satisfaction influences habit, which in turn influences CI. The partial mediation of habit implies that it explains some, but not all for this relationship. Other potential mediating variables may further clarify the nature of this relationship.

Conclusion

This research, based on the extended UTAUT2, explores the determinants for continued intention to use ELP and tests the mediating effect of e-satisfaction on the relationships between antecedents and CI, as well as the mediating effect of habit on the relationship between e-satisfaction and CI. The model has high explanatory and predicative power in this research. Results show that PE, LV, HM, and HB significantly affect students' continued intention to use ELP to learn college English. E-satisfaction and habit play a mediating role in the proposed relationships. This study is helpful for education administrators, particularly online education policy makers, to better understand the determinants that would increase students' learning efficiency, participation and satisfaction with using ELP for college English study.

The research results can guide online education policy makers and ELP designers to increase the effectiveness of ELP and improve students' usage experience, leading to greater satisfaction, higher learning efficiency and participation with continued use of ELP in future college English study.

Theoretically, this research aimed to broaden the scope and theoretical depth of UTAUT2 by examining the role of e-satisfaction and habit as mediating factors, rather than simply investigating the determinants of students' continued intention to use ELP for college English study. To achieve this, this research introduces e-satisfaction and learning value as additional components of the UTAUT2 framework in the context of college English study. The results show that e-satisfaction not only significantly affect students' willingness to continue using new technology, but also plays a partial and full mediating effect in the relationship between antecedents and CI. Furthermore, the research confirms the mediating effect of habit on the relationship between e-satisfaction and CI. By identifying the mediating effect of e-satisfaction and habit, this study contributes to a better understanding of the factors that would increase students' learning efficiency, participation, and satisfaction with using ELP, which can also increase the successful implementation of ELP for college English study.

Practically, this research provides suggestions for online education policy makers to increase students' learning efficiency, participation and satisfaction with using ELP for college English study. In order to increase students' learning efficiency, universities must take steps to raise students' awareness of ELP's usefulness and functions. This can be accomplished by emphasizing how ELP can help students complete class activities and tasks more efficiently. Students' participation can be increased by enhancing the usability of ELP. Designer should consider simplifying its operation procedures and categorizing learning resources, which would help students operate the platform more easily and efficiently, saving them time and energy. And what's more, it would be better for designers to integrate a game incentive mechanism into ELP. Gamifying the learning activities can offer more happiness and pleasure to students, making them more likely to participate the class activities. By using game incentive mechanisms, the learning experience can be made more exciting and interesting for the students. If students are satisfied with the learning experience, they are more likely to repeat the behavior. Once habitual behavior is developed, students will continuously use ELP for their daily study.

Although the research findings provide new insights for the successful implementation of ELP for college English study, they still have some limitations. Firstly, this cross-sectional research only reflects the students' perception and attitude towards using ELP for college English study at a certain point in time. Conducting longitudinal research could tract students' perception and experience with using a new technology over time. Secondly, the study has only examined the mediating effect of e-satisfaction and habit within the UTAUT2 model. There might be other potential mediators or moderators that could further explain the relationship between the antecedents and the continued willingness to use new technologies in the context of education. Thirdly, in the context of foreign language learning with information technology, students' language proficiency and interaction with instructors and peers have a great impact on their willingness to continuously use a certain new technology (Deng et al., 2019). Further research could incorporate language competence and interactive quality into the model to study its impact on students' satisfaction and continued intention to use ELP for foreign language study. Additionally, the research has not considered the influence of students' engagement and self-regulation on their continued intention to use ELP.

Data availability statement

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

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

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

Aarts, H., and Dijksterhuis, A. (2000). The automatic activation of goal-directed behaviour: the case of travel habit. *J. Environ. Psychol.* 20, 75–82. doi: 10.1006/jevp.1999.0156

Abbasi, G. A., Sandran, T., Ganesan, Y., and Iranmanesh, M. (2022). Go cashless! Determinants of continuance intention to use E-wallet apps: a hybrid approach using PLS-SEM and fsQCA. *Technol. Soc.* 68:101937. doi: 10.1016/j.techsoc.2022.101937

Ain, N., Kaur, K., and Waheed, M. (2016). The influence of learning value on learning management system use: an extension of UTAUT2. *Inf. Dev.* 32, 1306–1321. doi: 10.1177/0266666915597546

Alalwan, A. A. (2020). Mobile food ordering apps: an empirical study of the factors affecting customer e-satisfaction and continued intention to reuse. *Int. J. Inf. Manag.* 50, 28–44. doi: 10.1016/j.ijinfomgt.2019.04.008

Amoroso, D., and Lim, R. (2017). The mediating effects of habit on continuance intention. *Int. J. Inf. Manag.* 37, 693–702. doi: 10.1016/j.ijinfomgt.2017.05.003

Anderson, R. E., and Srinivasan, S. S. (2003). E-satisfaction and e-loyalty: a contingency framework. *Psychol. Mark.* 20, 123–138. doi: 10.1002/mar.10063

Baudier, P., Ammi, C., and Deboeuf-Rouchon, M. (2018). Smart home: highly-educated students' acceptance. *Technol. Forecast. Soc. Chang.* 153:119355. doi: 10.1016/j.techfore.2018.06.043

Bhattacherjee, A., Perols, J., and Sanford, C. (2008). Information technology continuance: a theoretic extension and empirical test. *J. Comput. Inf. Syst.* 49, 17–26. doi: 10.1080/08874417.2008.11645302

Cacciamani, S., Villani, D., Bonanomi, A., Carissoli, C., Olivari, M. G., Morganti, L., et al. (2018). Factors affecting students' acceptance of tablet PCs: a study in Italian high schools. *J. Res. Technol. Educ.* 50, 120–133. doi: 10.1080/15391523.2017.1409672

Cai, B. L., Lin, Q., Liang, G., Shi, Y. L., and Sun, H. C. (2020). Study on the transformation of college students' learning style in online teaching environment: problems, trends and development paths. *Heilongjiang Res High Educ* 320, 140–144. doi: 10.19903/j.cnki.cn23-1074/g.2020.12.026

Chen, Y., Yang, L. L., Zhang, M., and Yang, J. (2018). Central or peripheral? Cognition elaboration cues' effect on users' continuance intention of mobile health applications in the developing markets. *Int. J. Med. Inform.* 116, 33–45. doi: 10.1016/j.ijmedinf.2018.04.008

Chin, W. W. (1998). "The partial least squares approach to structural equation modeling," in *Modern methods for business research*. G. A. Marcoulides (Mahwah, NJ: Lawrence Erlbaum Associates) 295–336.

Coves-Martínez, A. L., Sabiote-Ortiz, C. M., and Frías-Jamilena, D. M. (2023). How to improve travel-app use continuance: the moderating role of culture. *Tour. Manag. Perspect.* 45:101070. doi: 10.1016/j.tmp.2022.101070

Dağhan, G., and Akkoyunlu, B. (2016). Modeling the continuance usage intention of online learning environments. *Comput. Hum. Behav.* 60, 198–211. doi: 10.1016/j.chb.2016.02.066

Dajani, D., and Abu Hegleh, A. S. (2019). Behavior intention of animation usage among university students. *Heliyon* 5:e02536. doi: 10.1016/j.heliyon.2019.e02536

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace. *J. Appl. Soc. Psychol.* 22, 1111–1132. doi: 10.1111/j.1559-1816.1992.tb00945.x

Deng, R., Benckendorf, P., and Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Comput. Educ.* 129, 48–60. doi: 10.1016/j.compedu.2018.10.019

Diamantopoulos, A., and Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration. *Br. J. Manag.* 17, 263–282. doi: 10.1111/j.1467-8551.2006.00500.x

Gao, B. (2023). Understanding smart education continuance intention in a delayed benefit context: an integration of sensory stimuli, UTAUT, and flow theory. *Acta Psychol.* 234:103856. doi: 10.1016/j.actpsy.2023.103856

Gu, W., Bao, P., Hao, W., and Kim, J. (2019). Empirical examination of intention to continue to use smart home services. *Sustainability* 11:5213. doi: 10.3390/su11195213

Gupta, A., and Dogra, N. (2017). Tourist adoption of mapping apps: a UTAUT2 perspective of smart travelers. *Tour. Hosp. Manag.* 23, 145–161. doi: 10.1108/GKMC-01-2022-0001

Hair, J. F., Hult, G. T. M., Ringle, C., and Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd). Thousand Oakes, CA: Sage.

He, L. Z. (2020). New requirements for college English teaching in the new era--- basis and key points for revision of college English teaching guidelines. *Foreign Lang World* 199, 13–18.

Henseler, J., Ringle, C. M., and Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* 43, 115–135. doi: 10.1007/s11747-014-0403-8

Ismail, I., Gunasegaran, T., and Azizan, S. N. (2016). Mobile learning in Malaysian universities: are students ready? *Int J. Interact. Mob. Technol.* 10, 17–23. doi: 10.3991/ijim.v10i3.5316

Kim, K. H., Kim, K. J., Lee, D. H., and Kim, M. G. (2019). Identification of critical quality dimensions for continuance intention in mHealth services: case study of onecare service. *Int. J. Inf. Manag.* 46, 187–197. doi: 10.1016/j.ijinfomgt.2018.12.008

Li, Y. Y., Zhang, H. M., and Zhang, H. Z. (2020). Model construction and empirical test of college students' satisfaction with online learning during epidemic prevention and control period: based on the survey of 15 universities in Shanghai. *Open Educ. Res.* 4, 102–111. doi: 10.13966/j.cnki.kfjyyj.2020.04.012

Liu, X. H., He, X. Y., Wang, M. M., and Shen, H. Z. (2022). What influences patients' continuance intention to use AI-powered service robots at hospitals? The role of individual characteristics. *Technol. Soc.* 70:101996. doi: 10.1016/j.techsoc.2022.101996

Maduku, D. K., and Thusi, P. (2023). Understanding consumers' mobile shopping continuance intention: new perspectives from South Africa. *J. Retail. Consum. Serv.* 70:103185. doi: 10.1016/j.jretconser.2022.103185

Mishra, A., Shukla, A., Rana, N. P., Currie, W. L., and Dwivedi, Y. K. (2023). Re-examining post-acceptance model of information systems continuance: a revised theoretical model using MASEM approach. *Int. J. Inf. Manag.* 68:102571. doi: 10.1016/j.ijinfomgt.2022.102571

Murray, K. B., and Häubl, G. (2007). Explaining cognitive lock-in: the role of skill-based habits of use in consumer choice. *J. Consum. Res.* 34, 77–88. doi: 10.1086/513048

Nanayakkara, C. (2007). A model of user acceptance of learning management systems: a study within tertiary institutions in New Zealand. *Int. J. Learn* 12, 223–232. doi: 10.18848/1447-9494/CGP/v13i12/45146

Nikolopoulou, K., Gialamas, V., and Lavidas, K. (2020). Acceptance of mobile phone by university students for their studies: an investigation applying UTAUT2 model. *Educ. Inf. Technol.* 25, 4139–4155. doi: 10.1007/s10639-020-10157-9

Osei, H. V., Kwateng, K. O., and Boateng, K. A. (2022). Integration of personality trait, motivation and UTAUT 2 to understand e-learning adoption in the era of COVID-19 pandemic. *Educ. Inf. Technol.* 27, 10705–10730. doi: 10.1007/s10639-022-11047-y

Perez-Aranda, J., Robles, E. M. G., and Urbistondo, P. A. (2023). Understanding antecedents of continuance and revisit intentions: the case of sport apps. *J. Retail. Consum. Serv.* 72:103288. doi: 10.1016/j.jretconser.2023.103288

Pine, B. J., and Gilmore, J. H. (2011), *The Experience Economy*, Harvard Business Press and McGill, Boston, MA.

Prasetyo, Y. T., Roque, R. A. C., Chuenyindee, T., Young, M. N., Diaz, J. F. T., Persada, S. F., et al. (2021). Determining factors affecting the acceptance of medical education e-learning platforms during the COVID-19 pandemic: UTAUT2 approach. *Healthcare* 9:780. doi: 10.3390/healthcare9070780

Qiao, W. F., Liu, W. T., and Li, M. L. (2021). Online teaching in the perspective of students: behaviors, effectiveness, and challenges---based on an online learning survey at Tsinghua University during COVID-19. *Tsinghua J. Educ.* 42, 57–66. doi: 10.14138/j.1001-4519.2021.01.005710

Rai, A. (2020). Editor's comments: the COVID-19 pandemic: building resilience with IS research. *Manag. Inf. Syst. Q.* 44, iii–vii.

Raman, A., and Thannimalai, R. (2021). Factors impacting the behavioral intention to use e- learning at higher education amid the Covid-19 pandemic: UTAUT2 model. *Psychol. Sci. Educ.* 26, 82–93. doi: 10.17759/pse.2021260305

Santosa, A. D., Taufk, N., Prabowo, F. H. E., and Rahmawati, M. (2021). Continuance intention of baby boomer and X generation as new users of digital payment during COVID-19 pandemic using UTAUT2. *J. Financ. Serv. Mark.* 26, 259–273. doi: 10.1057/s41264-021-00104-1

Sasongko, D. T., Handayani, P. W., and Satria, R. (2022). Analysis of factors affecting continuance use intention of the electronic money application in Indonesia. *Procedia Comput. Sci.* 197, 42–50. doi: 10.1016/j.procs.2021.12.116

Shiau, W. L., Sarstedt, M., and Hair, J. F. (2019). Internet research using partial least squares structural equation modeling (Pls-Sem). *Internet Res.* 29, 398–406. doi: 10.1108/IntR-10-2018-0447ff

Siyal, A. W., Hongzhuan, C., and Gang, C. (2021). From consumer satisfaction to recommendation of mobile app-based services: an overview of mobile taxi booking apps. *SAGE Open* 11:215824402110041. doi: 10.1177/21582440211004179

Tandon, U., Mittal, A., Bhandari, H., and Bansal, K. (2021). E-learning adoption by undergraduate architecture students: facilitators and inhibitors. *Eng. Constr. Archit. Manag.* 29, 4287–4312. doi: 10.1108/ECAM-05-2021-0376

Thadani, D. R., and Cheung, C. M. K. (2011). "Exploring the role of online social network dependency in habit formation," in *International Conference on Information Systems*. 4, 3446–3461.

Tran, V. D., and Trang, H. N. D. (2018). Exploring the relationships among social benefits, online social networks dependency, satisfaction and youth's habit formation. *Int. J. Inform Inf. Syst.* 1, 68–77. doi: 10.47738/IJIIS.V1I2.101

Veeramootoo, N., Nunkoo, R., and Dwivedi, Y. K. (2018). What determines success of an e-government service? Validation of an integrative model of e-filing continuance usage. *Gov. Inf. Q.* 35, 161–174. doi: 10.1016/j.giq.2018.03.004

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: toward a unified view. *MIS Q.* 27, 425–478. doi: 10.2307/30036540

Venkatesh, V., Thong, J. Y., and Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q.* 36, 157–178. doi: 10.2307/41410412

Wang, C., Harris, J., and Patterson, P. (2013). The roles of habit, self-efficacy, and satisfaction in driving continued use of self-service technologies: a longitudinal study. *J. Serv. Res.* 16, 400–414. doi: 10.1177/1094670512473200

Wu, M. L. (2010). *Practice of Questionnaire Statistical Analysis*. Chongqing: Chongqing University Press.

Wu, P., Zhang, R., Zhu, X., and Liu, M. (2022). Factors influencing continued usage behavior on mobile health applications. *Healthcare* 10:208. doi: 10.3390/healthcare10020208

Xu, W., Shen, Z. Y., Lin, S. J., and Chen, J. C. (2022). Improving the behavioral intention of continuous online learning among learners in higher education during COVID-19. *Front. Psychol.* 13:857709. doi: 10.3389/fpsyg.2022.857709

Yan, M., Filieri, R., Raguseo, E., and Gorton, M. (2021). Mobile apps for healthy living: factors influencing continuance intention for health apps. *Technol. Forecasting Soc. Change* 166:120644. doi: 10.1016/j.techfore.2021.120644

Zacharis, G., and Nikolopoulou, K. (2022). Factors predicting university students' behavioral intention to use e-learning platforms in the post-pandemic normal: an UTAUT2 approach with 'learning value'. *Educ. Inf. Technol.* 27, 12065–12082. doi: 10.1007/s10639-022-11116-2

Zeithaml, V. A., and Bitner, M. J. (2003). *Services Marketing Integrating Customer Focus Across the Firm* (*3rd*). New York, NY: Irwin McGraw-Hill.

Zhao, X., Lynch, J. G., and Chen, Q. (2010). Reconsidering baron and Kenny: myths and truths about mediation analysis. *J. Consum. Res.* 37, 197–206. doi: 10.1086/651257

Zou, C., Li, P., and Jin, L. (2021). Online college English education in Wuhan against the COVID-19 pandemic: student and teacher readiness, challenges and implications. *PLoS One* 16:e0258137. doi: 10.1371/journal.pone.0258137

Virtual reality: a promising instrument to promote sail education

Fa Ji[1,2], Xingjian Zhang[1], Shan Zhao[1] and Qun Fang[1]*

[1]School of Physical Education, Qingdao University, Qingdao, China, [2]Development Center for Water Sports, Qingdao University, Qingdao, China

Sailing has gained an increasing attention among children and adolescents in China, which raised a strong need for sail courses through physical education (PE). However, challenges in teaching practice arise with rapid development of the sport. In the current study, we proposed a perspective that virtual reality (VR) technology makes high-quality sail education accessible for students. Critical analysis summarized the prominent features that enhance sail education, including immersive experience, interactive learning, the first-person view, and practice under well-controlled conditions. Further, research on VR sport training indicated successful transfer from virtual environment to real situation. Specifically, significant improvement in skill performance and tactical behaviors were identified, which was attributed to the enhanced perception-action coupling after VR training. Additionally, VR-based coding programs provide students with affordances of designing the virtual environment. The content design education promotes comprehension and application of knowledge and theories when students develop the simulated environment with a high level of presence. Therefore, VR technology is a promising instrument to meet the increasing demand on sail education. While VR enriches educational resources for a large class size, the interdisciplinary feature of VR-based sail course can attract students with different study interests and backgrounds to the class.

KEYWORDS

virtual reality, sailing, skill transfer, physical education, pedagogy

1. Introduction: opportunities and challenges in sail education

Sailing has become a popular sport in China since 2008 Beijing Olympic Games. In coastal cities such as Qingdao, sailing has been embedded in the school curriculum. Educational resources nowadays are available for students to learn knowledge and skills via sail classes at school. In the meantime, sailing clubs provide children and adolescents with a variety of off-school services from sail experience for recreation to regular training for competition.

The increasing popularity of sailing can be attributed to awareness of the benefits in physical and psychological development. As a primary form of outdoor and adventure education, sailing is more of a powerful educational approach than of adventurous recreation (Mcculloch et al., 2010). Qualitative research by means of focus group summarized benefits of sailing for children in self-confidence and competence, key personal and interpersonal skills, specific life skills, academic performance, physical fitness, and mental health (Cotterill and Brown, 2018). Additionally, a pre-post study involving 147 adolescents was conducted to examine influence of sail education programs on adolescents. Significant improvement was found in self-concept of

competence and social skills after the sail experience (Capurso and Borsci, 2013). Further evidence has shown that water-based environment is favorable to natural settings without water (Depledge et al., 2011). A meta-analysis identified positive effects of green environment (e.g., urban green, countryside/farmland, forest and woodland, waterside, and wild habitats) on self-esteem and mood, but the presence of water generated the largest effect size (Barton and Pretty, 2010).

Physical education (PE) is an important approach to promoting sailing among the young generations in China. However, challenges also arise in teaching practice during rapid development of the sport. The increased class size raises a major challenge on availability of resources for students to gain adequate sailing experience. Opportunities of operation and practice can be reduced in a larger class size, which demands supplementary teaching strategies to ensure quality of the course. In addition, traditional sail course begins with teaching knowledge out of water, and then progresses to application and practice (Morales-Belando and Arias-Estero, 2017). The course integrates mathematics, physics, geometry, and other disciplines, which increases difficulty in learning. Weeks of lecture-based study can be mentally exhausting and boring for children and adolescents, thus reducing teaching and learning effects. Another main challenge lies in transfer of knowledge and skills to practice. It is common to see that high uncertainties of sailing on the sea make novices struggling, including those who showed outstanding performance in theoretical learning. For the considerations of the above-mentioned challenges, innovative strategies are needed to enrich educational affordances, enhance motivation of learning, and facilitate transition from theory to practice.

Virtual reality (VR) technology using head-mounted display (HMD) is portable, convenient, and affordable (Schiza et al., 2019), which has been widely applied to medical and language education (Chen et al., 2020; Parmaxi, 2023). A simulated environment providing learners with authentic learning experience can enhance both motivations and engagement in task performance, thus offering potential solutions to the challenges in teaching practice (Rey et al., 2022). The current study aims to investigate feasibility and efficacy of applying VR technology to sail education. Critical analysis focused on the prominent features of VR which facilitate acquisition of sailing knowledge and skills. In addition, efficacy of VR in sail education was evaluated by evidence regarding transfer of learning from virtual environment to real task. Further discussions provided insights into VR course design, leading to the perspective that integration of VR into sail education is a promising approach to diverse PE content and effective pedagogical strategies.

2. Features of VR technology to facilitate sail education

VR technology generates immersive experience with a high level of representation to the real world. The extent of immersion is measured by presence which refers to subjective feeling of "being there" (Sanchez-Vives and Slater, 2005). Simulating the real experience is important to sail education given that perception of wind speed, wind direction, current, and wave can hardly be developed by knowledge learning in a classroom. VR presents theoretical knowledge in the form of 3D practical scenarios (Plotzky et al., 2021). Sailing

simulators based on mathematic models have accounted for a variety of variables (Mulder and Verlinden, 2013; Angelou and Spyrou, 2021) which enable novices to gain valuable hands-on experience in VR at the early stage of learning. Research has shown that virtual environment with the higher immersive quality induces the greater psychological presence and similar responses to the real situations (Bowman and Mcmahan, 2007; Miles et al., 2012). The immersive experience can familiarize novices with the environment and lower the impact of nervousness and anxiety on performance. Indeed, mental readiness is essential for novices due to the high pressure and tight temporal constraints in sailing. Researchers used VR technology to prepare athletes for high-pressure situations such as soccer penalty kick (Stinson and Bowman, 2014). The significant reduction in sport-induced anxiety implies promising applications of VR to facilitate adaptation to the real sailing.

In addition to the immersive experience which helps learners to be prepared with the real task, another remarkable feature of VR is to provide interactive experience between the learners and the virtual environment (Neumann et al., 2018). In a scoping review regarding VR training on team ball sports performance, researchers summarized three types of interactions, including interaction with ball, interaction with other players, and interaction with both ball and players, which showed positive effects on both perception and performance (Faure et al., 2020). In the research on rowing simulation, interactions between oar and water were essential factors which induced comparable skill gains to the real training (Rauter et al., 2013). Accordingly, it is reasonable to expect positive effects of VR sailing should interactions between actions and corresponding outcomes be represented in the sailing simulation.

The first-person view in HMD enhances immersion and interactive experience, which plays an essential role in training effects and sport performance (Covaci et al., 2015). Petit and Ripoll (2008) found that experienced soccer players made faster and more accurate decisions in the first-person view than in the broadcast point of view. In the case of sail education, the first-person view makes the learning process more active given the fact that learners can freely observe surroundings and collect perceptual information from the virtual environment. In this sense, learning in VR becomes an exploratory process. Taking course keeping in the upwind sailing as an example, theories only give learners the optimal angle to sail in a desired direction. It is also important for learners to gain insights into various outcomes of adjusting the sail at different angles. The exploratory learning by interacting with virtual environment can be greatly helpful for novices to enrich sailing experience from the simulated operation. Therefore, VR technology enables the principle of learning by doing to be implemented in sail education, which is a prominent advantage over traditional approaches.

The high level of control over the designed task and environment is of great value in VR-based training (Bideau et al., 2010; Zaal and Bootsma, 2011). It is suitable to use VR as a supplementary approach to the on-site training which is expensive, dangerous, or difficult to replicate in real life (Oagaz et al., 2022). For sail education, standardized procedures can be simulated to provide affordances for repetitions on specific skills and performance (Mulder and Verlinden, 2013). VR training is not only for the realistic situations, but also for extreme, unrealistic conditions to "over-prepare" trainees for the real situation (Oagaz et al., 2022). In addition to the environmental design, the customized sail simulator setup also allows a full control

over feedback in both quantity and quality to guide learners during training (Faure et al., 2020). For example, a checklist for safety check before departure can be transferred to visual and verbal instructions. The instructions can be provided in concurrent with each step to assist students in completing the task. Therefore, the above-mentioned characteristics of VR technology, including immersive experience, interactive learning, the first-person view, and affordances for adequate repetitions under well-controlled conditions, substantiate the feasibility of integrating VR into sail education.

3. Efficacy of skill development: transfer from virtual to real environment

Transfer of learning from virtual environment to real task is a determinant for the efficacy of VR-based practice. A wide application of VR to sail education should be justified by empirical evidence for successful transfer. Harris et al. (2021) examined validity of a VR golf putting simulator. Participants reported a high degree of presence in the virtual environment, which validated using VR technology to simulate real golf putting. Additionally, in a study which compared movement patterns in VR with those in real situations, similar stance patterns were identified when tennis players performed groundstrokes, suggesting adequate representativeness of a virtual environment to the actual tennis performance (Le Noury et al., 2021). Cumulative evidence suggests a greater role of functional fidelity (e.g., tracking level, stereoscopy, and field of view) than physical fidelity (e.g., image and sound quality) in contributing to a high-quality virtual environment (Cummings and Bailenson, 2016). To increase functional fidelity, it is important to preserve the perception-action loop that is normally experienced in a real environment (Craig, 2013).

The extent of learning transfer to real situation is a critical consideration regarding the effectiveness of sport training via VR technology. Research has shown that VR facilitates acquisition of table tennis skills measured by quality of returned balls (Oagaz et al., 2022). In this study, novices were randomly assigned to either VR group or conventional training group. After five table tennis sessions, the VR group outperformed the control group in terms of speed and height of the returned ball, indicating improved returning technique. The findings led to the conclusion that VR is effective in improving complex sport skills and the learned skills can be transferred to enhance actual performance.

It is also an interest of study in the influence of VR on cognitive perception which is critical for sport performance. A recent study examined effects of using a VR system to assist basketball players in understanding and executing offensive tactics (Tsai et al., 2022). Participants learned and practiced four tactics which were provided by VR system, video clips, and instructions on tactic board. VR training induced greater improvement in accuracy of running path in the field tests than the other groups. The advantage of VR in tactic training is more evident when tactics involve with multiple steps and more players. For simple tactics, the conventional training approach by tactic board could provide adequate information. As tactical complexity increases, the immersive experience facilitates strategic imagery and comprehension of tactics which help the players with execution in the real situations.

A possible explanation to the positive transfer of VR training to actual performance is the enhanced perception-action coupling. In a baseball batting study, high school baseball players were allocated to one of the groups which received adaptive VR practice, VR practice, and real batting practice in addition to the regular baseball training (Gray, 2017). Adaptive VR practice was implemented by manipulating pitch parameters (e.g., spin, speed, pitch crossing height, and pitcher's handedness) to match performance in each trial. Favorable batting performance was identified in the adaptive VR over the other groups, suggesting the positive effect of pitch recognition on actual performance. Therefore, evidence for the transfer of learning has been found in both behavioral (Oagaz et al., 2022) and cognitive performance (Tsai et al., 2022), which further implies an essential role of the enhanced perception-action coupling in the successful transfer from virtual environment to real situation (Gray, 2017).

Therefore, the main findings from the existing research substantiated the following statements: (1) VR technology is valid in simulating the real situation; (2) training in the simulated environment benefited actual performance in skill and tactical behaviors; and (3) enhanced perception-action coupling may contribute to the positive transfer from practice in the virtual environment to performance in the real task.

4. VR sailing design: a novel approach to enhance understandings

While sailing skills can be practiced in a simulated environment, understandings of knowledge can be enhanced via VR-based design education. This innovative approach allows teachers and students to collaboratively create course content, which stimulates new ways of thinking during the design activity (Denner et al., 2012). Such an idea was inspired by a VR-based coding education program for middle and high school students in South Korea (Shim and Lee, 2022). The researchers implemented a six-week coding course with an emphasis on energy education. By asking the students to design a VR theme park using future energy, the teachers helped the students to raise awareness of energy issues in the communities and improve problem solving abilities in energy saving. Significant benefits were reported in the students' competence and educational satisfaction, thus providing evidence for the VR-based design education in promoting effectiveness of both teaching and learning.

Adequate simulation of sailing demands comprehensive understandings of the sport. Considerations regarding stereoscopy and principles of physics should be counted when designing virtual environment of sailing on the sea. The sail course may give particular attention to VR content design by encouraging students to develop code and algorithm. More importantly, interactions between the environmental factors (e.g., wind, wave, current) and outcomes (e.g., vessel direction and velocity) are essential for a high level of presence. Students can easily identify potential flaws in the coding if the outcome is not in line with expectation. Further corrections then should be conducted to make the virtual environment close to the real one.

In summary, a framework of VR-based sail course was developed in correspondence to the increasing need among the students in China (Figure 1). Sail education can be provided through VR training and coding program. While VR training facilitates the learning process in which knowledge, skills, and experience of sailing can be improved in the simulated environment, coding program enhances
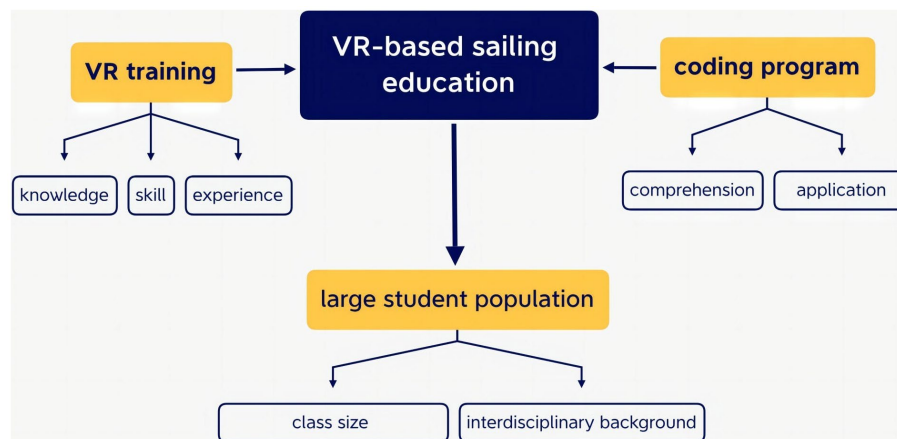
FIGURE 1
Framework of VR-based sail course.

comprehension and application of sailing knowledge by designing the simulated environment. Therefore, the course framework allows sail education to accommodate a large student population. On the one hand, VR technology enriches educational affordances for a larger class size. On the other hand, the interdisciplinary feature allows students with a wide variety of study interests and backgrounds to take the course.

5. Conclusion

The increasing popularity of sailing in China results in the rapidly growing need for sail education at school. However, practical challenges also arise in considerations of the increasing class size, study motivation, and transition from theory to practice, which highlight the necessity of innovative pedagogical strategies to enhance sail education. VR is characterized by immersive experience, interactive learning, the first-person view, and affordances for adequate repetitions under well-controlled conditions. Additionally, empirical evidence from the existing research indicates successful transfer of learning from virtual environment to real task, which implies the efficacy of using VR for knowledge learning and skill development in sail education. VR technology not only provides opportunities for practice, but also offers a platform for VR content design. Coding programs help students to enhance comprehension and application while designing a simulated sailing environment. Therefore, VR is a promising instrument to meet the increasing demand on sail education by providing adequate educational resources for a large class size and making the class accessible for students with different study interests and backgrounds.

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.

Author contributions

FJ and XZ prepared the draft. FJ, SZ, and QF worked on revision and approved the submitted version. FJ, XZ, SZ, and QF collaborated in preparing 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

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

Angelou, M., and Spyrou, K. J. (2021). Course-keeping assessment of yachts' upwind sailing in waves. *J. Fluids Struct.* 101:103227. doi: 10.1016/j.jfluidstructs.2021.103227

Barton, J., and Pretty, J. (2010). What is the best dose of nature and green exercise for improving mental health? A multi-study analysis. *Environ. Sci. Technol.* 44, 3947–3955. doi: 10.1021/es903183r

Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., and Craig, C. (2010). Using virtual reality to analyze sports performance. *IEEE Comput. Graph. Appl.* 30, 14–21. doi: 10.1109/MCG.2009.134

Bowman, D. A., and Mcmahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer* 40:36. doi: 10.1109/MC.2007.257

Capurso, M., and Borsci, S. (2013). Effects of a tall ship sail training experience on adolescents' self-concept. *Int. J. Educ. Res.* 58, 15–24. doi: 10.1016/j.ijer.2013.01.004

Chen, F. Q., Leng, Y. F., Ge, J. F., Wang, D. W., Li, C., Chen, B., et al. (2020). Effectiveness of virtual reality in nursing education: meta-analysis. *J. Med. Internet Res.* 22:e18290. doi: 10.2196/18290

Cotterill, S. T., and Brown, H. (2018). An exploration of the perceived health, life skill and academic benefits of dinghy sailing for 9–13-year-old school children. *J. Advent. Educ. Outdoor Learn.* 18, 227–241. doi: 10.1080/14729679.2018.1424001

Covaci, A., Olivier, A. H., and Multon, F. (2015). Visual perspective and feedback guidance for VR free-throw training. *IEEE Comput. Graph. Appl.* 35, 55–65. doi: 10.1109/MCG.2015.95

Craig, C. (2013). Understanding perception and action in sport: how can virtual reality technology help? *Sports Technol.* 6, 161–169. doi: 10.1080/19346182.2013.855224

Cummings, J. J., and Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychol.* 19, 272–309. doi: 10.1080/15213269.2015.1015740

Denner, J., Werner, L., and Ortiz, E. (2012). Computer games created by middle school girls: can they be used to measure understanding of computer science concepts? *Comput. Educ.* 58, 240–249. doi: 10.1016/j.compedu.2011.08.006

Depledge, M. H., Stone, R. J., and Bird, W. J. (2011). Can natural and virtual environments be used to promote improved human health and wellbeing? *Environ. Sci. Technol.* 45, 4660–4665. doi: 10.1021/es103907m

Faure, C., Limballe, A., Bideau, B., and Kulpa, R. (2020). Virtual reality to assess and train team ball sports performance: a scoping review. *J. Sports Sci.* 38, 192–205. doi: 10.1080/02640414.2019.1689807

Gray, R. (2017). Transfer of training from virtual to real baseball batting. *Front. Psychol.* 8:2183. doi: 10.3389/fpsyg.2017.02183

Harris, D. J., Buckingham, G., Wilson, M. R., Brookes, J., Mushtaq, F., Mon-Williams, M., et al. (2021). Exploring sensorimotor performance and user experience within a virtual reality golf putting simulator. *Virtual Reality* 25, 647–654. doi: 10.1007/s10055-020-00480-4

Le Noury, P., Buszard, T., Reid, M., and Farrow, D. (2021). Examining the representativeness of a virtual reality environment for simulation of tennis performance. *J. Sports Sci.* 39, 412–420. doi: 10.1080/02640414.2020.1823618

Mcculloch, K., Mclaughlin, P., Allison, P., Edwards, V., and Tett, L. (2010). Sail training as education: more than mere adventure. *Oxf. Rev. Educ.* 36, 661–676. doi: 10.1080/03054985.2010.495466

Miles, H. C., Pop, S. R., Watt, S. J., Lawrence, G. P., and John, N. W. (2012). A review of virtual environments for training in ball sports. *Comput. Graph. UK* 36, 714–726. doi: 10.1016/j.cag.2012.04.007

Morales-Belando, M. T., and Arias-Estero, J. L. (2017). Effect of teaching races for understanding in youth sailing on performance, knowledge, and adherence. *Res. Q. Exerc. Sport* 88, 513–523. doi: 10.1080/02701367.2017.1376032

Mulder, F. A., and Verlinden, J. C. (2013). Development of a motion system for an advanced sailing simulator. *Proc. Eng.* 60, 428–434. doi: 10.1016/j.proeng.2013.07.030

Neumann, D. L., Moffitt, R. L., Thomas, P. R., Loveday, K., Watling, D. P., Lombard, C. L., et al. (2018). A systematic review of the application of interactive virtual reality to sport. *Virtual Reality* 22, 183–198. doi: 10.1007/s10055-017-0320-5

Oagaz, H., Schoun, B., and Choi, M. H. (2022). Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Trans. Vis. Comput. Graph.* 28, 4332–4343. doi: 10.1109/TVCG.2021.3086403

Parmaxi, A. (2023). Virtual reality in language learning: a systematic review and implications for research and practice. *Interact. Learn. Environ.* 31, 172–184. doi: 10.1080/10494820.2020.1765392

Petit, J. P., and Ripoll, H. (2008). Scene perception and decision making in sport simulation: a masked priming investigation. *Int. J. Sport Psychol.* 39, 1–19.

Plotzky, C., Lindwedel, U., Sorber, M., Loessl, B., König, P., Kunze, C., et al. (2021). Virtual reality simulations in nurse education: a systematic mapping review. *Nurse Educ. Today* 101:104868. doi: 10.1016/j.nedt.2021.104868

Rauter, G., Sigrist, R., Koch, C., Crivelli, F., Van Raai, M., Riener, R., et al. (2013). Transfer of complex skill learning from virtual to real rowing. *PLoS One* 8:e82145. doi: 10.1371/journal.pone.0082145

Rey, B., Oliver, A., Monzo, J. M., and Riquelme, I. (2022). Development and testing of a portable virtual reality-based Mirror visual feedback system with behavioral measures monitoring. *Int. J. Environ. Res. Public Health* 19:2276. doi: 10.3390/ijerph19042276

Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 332–339. doi: 10.1038/nrn1651

Schiza, E., Matsangidou, M., Neokleous, K., and Pattichis, C. S. (2019). Virtual reality applications for neurological disease: a review. *Front. Robot. AI* 6:100. doi: 10.3389/frobt.2019.00100

Shim, H., and Lee, H. (2022). The effect of design education using virtual reality-based coding on student competence and educational satisfaction. *Educ. Inf. Technol.* 27, 4577–4597. doi: 10.1007/s10639-021-10730-w

Stinson, C., and Bowman, D. A. (2014). Feasibility of training athletes for high-pressure situations using virtual reality. *IEEE Trans. Vis. Comput. Graph.* 20, 606–615. doi: 10.1109/TVCG.2014.23

Tsai, W. L., Pan, T. Y., and Hu, M. C. (2022). Feasibility study on virtual reality based basketball tactic training. *IEEE Trans. Vis. Comput. Graph.* 28, 2970–2982. doi: 10.1109/TVCG.2020.3046326

Zaal, F., and Bootsma, R. J. (2011). Virtual reality as a tool for the study of perception-action: the case of running to catch Fly balls. *Presence Teleoperators Virtual Environ.* 20, 93–103. doi: 10.1162/pres_a_00037

Utility-value change and the role of emotional cost in video-based learning: a matter of student teachers' interpretation of experience

Mathias Dehne ⓘ* and Alexander Gröschner ⓘ

Faculty of Social and Behavioural Sciences, Friedrich Schiller University Jena, Jena, Germany

Motivational and emotional characteristics influence teachers' reflections on video clips from their own teaching. However, utility values and the role of emotional cost have not been considered yet in video-based research in teacher education. In the present study, 102 student teachers were randomly assigned to an intervention group (IG) with video-based documentation of a lesson and systematic writing assignment or to a control group (CG) with protocol-based documentation of a lesson and writing assignment. Multigroup latent change score analysis indicated that IG participants, on average, showed a 0.52 SDs higher increase in utility values than the CG three months after the teaching practicum. Emotional cost was negatively related to baseline utility-value scores but predicted latent change scores positively after the writing assignment. The study adds to the current repertoire of video-based learning opportunities by providing a systematic writing assignment targeting student teachers' interpretation of experiences to leverage motivation.

KEYWORDS

utility-value intervention, emotional cost, video-based learning, latent change analysis, teaching practicum

1. Introduction

Utility values are typically addressed by a scaled-up wise intervention approach known as utility-value interventions (Gaspard et al., 2021). When students struggle in courses, psychological processes might underlie that "interfere with optimal academic functioning" (Harackiewicz and Priniski, 2018, p. 410) and "often act as key levers … that give rise to social problems" (Walton, 2014, p. 80). Drawing on situated expectancy-value theory (SEVT; Eccles and Wigfield, 2020), this study explored the utility-value change of student teachers reflecting by video vs. student teachers reflecting by a protocol on a lesson taught during their teaching practicum. Ample research has demonstrated the effectiveness of video-based interventions in terms of enhancing student teachers' classroom-related outcomes, such as "noticing" or "reflecting" on teaching and learning (Korthagen, 2010; Gaudin and Chaliès, 2015; Kleinknecht and Gröschner, 2016; König et al., 2022). Although previous research has stressed that higher utility values can be expected for video-based compared to text-based reflections (cf. Brouwer et al., 2017), studies concluded that video does not automatically activate emotional and motivational processes (Kleinknecht and Schneider, 2013). Neither does simply viewing video guarantee teacher learning (Gaudin and Chaliès, 2015). In contrast, research has not yet

implemented interventions targeting student teachers' motivation in complex video-based learning environments with a higher cognitive load than protocol-based methods induce (Weber et al., 2023). Consequently, the role of utility values remains unexplored (Dehne et al., 2018; Nickl et al., 2023). Drawing on this body of research, the design of augmented learning environments regarding student teachers' motivational and emotional characteristics is an important prerequisite to enhancing teacher learning with video.

In this study, we tested a new intervention approach in line with SEVT and previous utility-value-like interventions in order to promote student teachers' utility values for reflecting on teaching and learning (cf. Harackiewicz et al., 2012). Student teachers had to relate their video-based or text-based reflection to their achievement-related experiences and interpretations thereof (Brouwer et al., 2017). Additionally, we investigated the influence of emotional cost on changes in student teachers' utility values in video-based learning. Control-value theory, as a well-known framework on achievement emotions, has pointed out that negative outcome emotions such as anxiety and hopelessness are aroused by a lack of control and doubts about the certainty of success and negatively relate to achievement (Pekrun, 2006). Regarding video-based interventions, research has referred to the negative emotional arousal for video-based interventions when they make use of participants' video clips of their teaching (e.g., Kleinknecht and Schneider, 2013; Chan et al., 2018), whereas the learning outcome will be influenced negatively (Charalambous et al., 2022). By investigating motivational and emotional characteristics with the conceptual lens of SEVT, our study will provide new knowledge to promote student teachers' video-based learning.

1.1. Situated expectancy-value theory

SEVT constitutes a comprehensive lens for investigating motivation and its relationship to various academic outcomes, such as effort and course grades, in various academic settings. In being ample regarding the background characteristics and the ontogeny of achievement motivation, the model provides a broad range of paths to be harnessed in interventions (see Figure 1). However, when the model is traditionally considered in psychological research, expectancy beliefs ("Can I do this task?" with its self-efficacy-like implications) and values ("Why should I do this?" with personality, intrinsic, extrinsic, and cost implications) are differentiated as predictors of achievement-related choices among students (Eccles and Wigfield, 2002, 2020; Eccles, 2005). In this regard, utility value reflects the relevance to future goals (Eccles, 2005). Cost marks the fear of failure or perceived psychological threat associated with a task (Song et al., in press). When students value a task and believe they can succeed, they are more likely to take on a challenging task, as negative emotions will be lowered (Pekrun, 2006; Harackiewicz et al., 2012). In this regard, the critical role of cost perceptions as part of SEVT has been widely remarked upon, although it has received relatively little research attention (Rosenzweig et al., 2020). Cost perceptions negatively affect students' learning outcomes. For instance, university students with higher perceptions of cost show stronger intentions to leave their STEM major (Perez et al., 2014). Flake et al. (2015) focused on a more sophisticated differentiation of the cost component in task effort, outside effort, loss of valued alternatives, and emotional cost.

Recently, Song et al. (in press) have critically investigated emotional cost and anxiety regarding their mutual relevance in students' experiences. Both constructs provided insufficient evidence for structural and discriminant validity (i.e., high cross-loadings and high factor correlations). Furthermore, both constructs showed an overlap when predicting final exam scores in an introductory biology course at the university with almost the same magnitude. These findings support our assumption that emotional cost could be related to video-based learning during a teaching practicum, as previous research has pointed out anxiety as one of the most prevalent achievement emotions when using videos of one's teaching (e.g., Chan et al., 2018).

The exclusive focus on expectancies and values as one part of the model in the past decades has prompted Eccles and Wigfield (2020) to refine the model and refer more specifically to its situated components. Consequently, they referred to facets included in the social and experiential background as well as surrounding interpretations of experiences as the "often-neglected" left and middle parts. Although the components have not been the focus, the experiential background and achievement-related interpretations in the SEVT model frame personal characteristics, the enculturation, and socialization of a person's motivational beliefs and illustrate how cognitive processes mediate the effects of social and experiential background with regard to expectancies, values, and the self-concept of ability (Eccles and Wigfield, 2020; cf. Tschannen-Moran et al., 1998).

1.2. The psychology of utility-value interventions

Reasons for learning more abstract contents and higher-order processes are typically not obvious (Wentzel and Brophy, 2014). Likewise, this circumstance applies to more immersive video-based reflections in a teaching practicum (cf. Beck and Kosnik, 2002). Inter-individual variability regarding the learning outcomes and difficulties student teachers experience will result from video-based learning activities, as a study on ambitious teaching pointed out (Charalambous et al., 2022). Consequently, teacher educators are requested to design learning environments thoroughly and consider student teachers' heterogeneous backgrounds regarding their emotions and motivation as they matter for their learning outcomes and professional competence (Vermunt and Endedijk, 2011; Kunter, 2013; Klassen et al., 2020). Kleinknecht and Schneider's (2013) emphasis that video-based reflections do not automatically motivate (student) teachers and will thus not automatically lead to learning gains without regarding their learner characteristics captures a major requirement for teacher learning in any learning environment (e.g., Vandewaetere et al., 2011; Bardach et al., 2022). To date, video-based research lacks interventions on student teachers' psychological characteristics (e.g., motivation or personality) that would contribute knowledge about ways to scaffold them in video-based tasks.

During the last decade, research employing utility-value interventions provided substantial evidence for their effectiveness in fostering student outcomes, such as interest, utility values, retention, course grades, or continuation to the next course, mainly in STEM fields (Hulleman and Harackiewicz, 2009; Lazowski and Hulleman, 2016; Harackiewicz and Priniski, 2018; Hulleman and Harackiewicz, 2021). They have been applied in such diverse contexts like mathematics classrooms in school (Gaspard et al., 2015a), introductory biology courses in universities (Canning et al., 2018), and for parents to foster
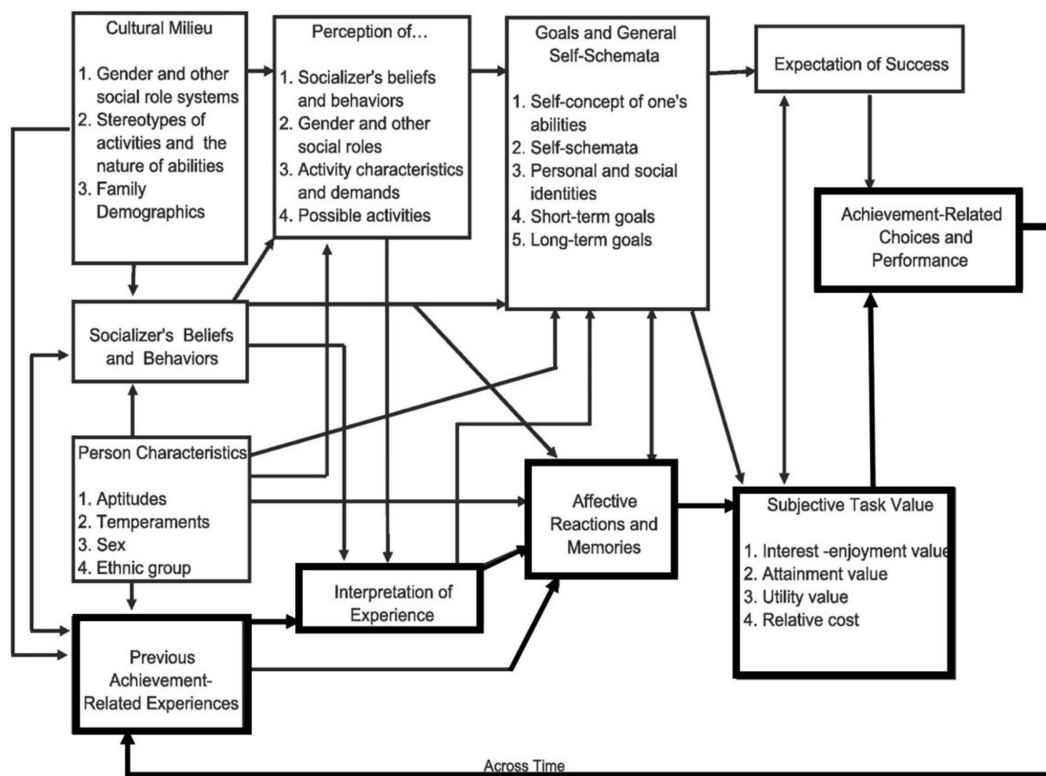
FIGURE 1
Situated expectancy-value model of achievement-related choices and performance. Bold = Paths underlying the intervention. Reprinted from Contemporary Educational Psychology, Vol 61, Jacquelynne S. Eccles & Allan Wigfield, From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation, Page 2, Copyright (2020), with permission from Elsevier.

adolescents' career orientation, career support, and the importance of career support (Piesch et al., 2019). Typically, social-psychological utility-value interventions use essay-based approaches with only little effort to implement. Here, students are asked to relate course content in a personal (e.g., by pronoun use), specific (reg. Course content and its relevance for an activity, hobby, etc.), and context-relevant manner (reg. Course content and everyday life or future career goals). According to Hulleman and Harackiewicz' (2021) logic model, these aspects add up to an individual's intervention fidelity; they are part of a process of raising students' perceived utility and success expectations and lowering their perceived cost. In turn, this process contributes to co-occurring psychological (e.g., developing an enduring interest or becoming more actively involved in a task) and behavioral (e.g., engagement or performance) mechanisms. Finally, outcomes such as grades, persistence, or career intentions are fostered.

By harnessing utility values in this set of interventions, STEM students in an introductory undergraduate biology course were more likely to enroll in the second course (Canning et al., 2018). Concerning the timing, students with a history of poor achievement benefitted most from a utility-value intervention at the beginning of the semester. In contrast, higher-performing students benefitted most by completing a utility-value task at the end of the semester to enroll in the second course. However, Canning et al. (2019) found some unintended consequences: struggling university students lost interest and perceived less utility. In a similar vein, the essay's length and quality (i.e., using simpler words) decreased for struggling students in one

intervention assignment. Another strand of research has focused on whether an essay-based intervention design works best, showing equal or more favorable results for quotation-based approaches where students are asked to evaluate quotations from former students regarding their personal relevance (Gaspard et al., 2015a, 2021). Although the results are promising in terms of transferring experiences into students' mindsets (cf. Eccles and Wigfield, 2020), these findings are limited to high school students.

In an intervention approach to career orientation with parents, utility-value quotes from former students and parents were presented on a website. It was intended to address students' interpretations of experiences more directly by socializers' beliefs (Piesch et al., 2019). However, the study by Piesch et al. found no positive intervention effects on either students' or parents' utility values or career support and orientation. Thus, the promising results from Harackiewicz et al. (2012) harnessing parental behaviors (and thus also adolescents' perceptions of their behaviors) for students' interpretations of experiences were not replicated. Harackiewicz's team mailed brochures to parents and used a website over a 15-month period that led students to take significantly more mathematics and science classes during their last two years of high school compared to a control group. Eccles and Wigfield (2020) acknowledged this as a step in the right direction (i.e., focusing on the "often-neglected middle part" of the SEVT model) "by introducing new experiences into students' school and family settings, thus acknowledging the role of experience" (p. 8). However, most utility-value interventions do not fully cover

experiential backgrounds. Especially for the teaching practicum, it is argued that student teachers value it most when experiences are being analyzed, for instance, as part of the course (Kleinknecht and Gröschner, 2016). This raises the question: What happens if student teachers' interpretations are addressed directly in interventions (linked to affective memories and indirectly to utility values) after an achievement-related experience?

Similar to utility-value interventions, it could be assumed that a cost-effective essay-based approach implemented during the semester might be utilized to address student teachers' interpretations of experiences more specifically. In this sense, the interpretation-of-experience box (Figure 1, in bold) could be understood with Tschannen-Moran et al.'s (1998) cyclical model of teacher efficacy. Cognitive (interpretative) processes mediate the influence of sources of teacher self-efficacy (e.g., mastery experience or verbal persuasion) on the task and ability analysis and, as a result, indirectly shape teachers' sense of efficacy. Thus, simply introducing the relationship of a certain topic to one's daily life or career goals is not to be seen as promising in enhancing student teachers' motivation. Rather, how they react and attempt to overcome obstacles in the future and whether they are capable of doing so is the key to maintaining motivation for certain topics (cf. Kunter, 2013). In the following section, we will liken common motivational and emotional conceptualizations for student teachers' video-based learning with utility values and emotional cost.

1.3. Video-based learning and emotional-motivational processes

Ample research has focused on the motivational and emotional facets of (student) teachers' video-based learning (Goldman, 2007; Seidel et al., 2011; Kleinknecht and Schneider, 2013; Gaudin and Chaliès, 2015; Kleinknecht and Gröschner, 2016). However, the conceptualizations of different facets remain heterogeneous (Dehne et al., 2018). Video, in this context, is regarded as a powerful tool as it provides performance in action and can be repeatedly viewed and analyzed from different perspectives (Brouwer et al., 2017). During a teaching practicum, video-based tools act as a "third space" where practitioner and academic knowledge are brought together in new hybrid ways to enhance student teachers' learning (Zeichner, 2010). Research on teachers' video-based learning (e.g., Seidel et al., 2011) often characterizes emotional and motivational processes by the terms "resonance" and "immersion."

1.3.1. The concepts of resonance and immersion

Resonance shows a direct relation to the concept of utility value. It captures how a video-based learning tool is connected to the viewer's situation (Goldman, 2007). More specifically, resonance can be conceptualized in terms of the relevance for student teachers' future work in the teaching profession (Dehne et al., 2018). Hence, the resonance one perceives when watching a video points to the facet utility for job (cf. Gaspard et al., 2015b), if the video is guiding one's focus on his/her instructional practices (Seidel et al., 2011). Consequently, research has stressed the importance of utility values for (student) teachers' intention to use technology in the classroom or their frequency of technology integration (Teo, 2009; Backfisch et al., 2021). However, research

has yet largely missed introducing the concept of utility value to student teachers' video-based learning.

In contrast, Goldman (2007) refers to *immersion* as a form of deep-level engagement or emotional involvement in the topic. While it contains positive facets such as excitement (Seidel et al., 2011), the other side of the coin might be a feeling of shame as a direct result of videotaping, having to select, and later watching own—potentially unpleasant—classroom sequences (Chan et al., 2018). Richards et al. (2021) have pointed to this complex interplay when teachers generate video clips for teacher professional development programs. While generating videos already guides teachers' focus to student thinking, the complexity of the whole process makes less favorable outcomes possible, especially in diverse settings. Dehne et al. (2018) have shown that sustaining these efforts requires increased support in teacher education. Emotional cost is a task value facet described by SEVT (cf. Gaspard et al., 2015b; Part et al., 2020) and is directly related to immersion. The concept is helpful in refining the understanding of the process of immersion, which has been regarded positively as a motivational reference for teachers' involvement in the video-based task (Seidel et al., 2011). As emotional cost refers to the broader motivational value of a task (cf. Part et al., 2020), it might enrich conclusions for value-based interventions in heterogeneous learning communities.

Reviews on (student) teachers' video-based learning (Gaudin and Chaliès, 2015; Santagata et al., 2021) referred to studies where teachers watched videos of their own teaching versus videos of peers or unknown teachers (Seidel et al., 2011; Kleinknecht and Schneider, 2013; Chan et al., 2018) and where video-based conditions were compared to protocol-based conditions (Moreno and Valdez, 2007; Kleinknecht and Gröschner, 2016; Prilop et al., 2019) as significant research strands. A large body of research has sought interrelations between the degree of emotional-motivational activation when (student) teachers watch videos of their own versus videos of others teaching. Following, we refer to exemplary findings that provide insight into individual learning with video and emotional-motivational processes.

1.3.2. Literature review

Seidel et al. (2011) found that teachers who watched videos of their own teaching showed higher levels of immersion and resonance compared to teachers watching videos of others' teaching. In addition, when teachers were asked whether they found using video motivating, those using their own videos again showed higher scores. Against this backdrop, Kleinknecht and Schneider's (2013) study investigated the activation of emotional-motivational processes for teachers reflecting on videos of their own compared to others' teaching. The study showed no significant differences in immersion or resonance in teachers' comments on the respective videos. However, a tendency was found that teachers in the "own video group" showed more resonance with their practice. Counterintuitively, they found that participants in the "other video group" showed significantly more emotional expressions and a tendency toward more negative emotions. This finding is in stark contrast with Chan et al. (2018). They found that student teachers experienced more frequent negative emotions such as anxiety or shame when watching videos of their own teaching compared to fewer negative and more positive emotions when watching videos of their peers. The authors concluded that although reflecting own practices retrospectively and apart from cognitive and

emotional involvement in the teaching situations in which they occur, a variety of negative emotions is elicited. Whether the difference compared to Kleinknecht and Schneider's (2013) study dates back to different stages of experience (in- vs. pre-service) has not been reflected. One assumption was that certain emotions when watching own teaching for the first time are more likely or that these emotions result from unrealistic and unmet expectations.

Regarding research that compared video-based with protocol-based reflections, Prilop et al. (2019) used the setting of a teaching practicum in teacher education to investigate changes in teacher self-efficacy and constructivist beliefs. However, self-efficacy was fostered in either group, and constructivist belief did not decrease for the video-based reflections but for protocol-based conditions. Dehne et al. (2018) showed a similar finding regarding student teachers' expectancy for success and related it to overall mastery experiences in the teaching practicum. Additionally, the study by Prilop et al. (2019) showed that protocol-based reflections led to increases in more traditional teaching beliefs. In another study, Koehler et al. (2005) found some beneficial effects of video on college students' emotional engagement, positive affect, or interest engagement compared to text-based approaches. The authors linked this finding to information typically unavailable in texts (e.g., information about mood or tone). However, they noted that the effects of videos are not ubiquitous and largely depend on the context for which researchers have developed structured viewing guides in the field of teacher education (Brouwer et al., 2017). In essence, protocol-based approaches offer better possibilities for comparing the merits of video-based tools since video as a medium shows a stark contrast with additional sources of information (concrete contexts, audio, etc.) not available in a protocol format (cf. Prilop et al., 2019).

A few limitations arise from the reviewed literature. First, more robust analyses and larger samples must be considered in educational research to verify findings on (student) teachers' motivation and emotions, as case study designs were oftentimes adopted. Second, the context of the studies must be considered. All studies attempted to harness teachers' learning (e.g., noticing or reflecting) with videos. However, virtually none of them directly addressed what Kleinknecht and Schneider (2013) called for—to make use of more prearrangements or scaffolding, which is especially necessary when reflecting on their own videos (see also Brouwer et al., 2017). As resonance has been described as the extent to which teachers are able to relate a video to their own experience (Goldman, 2007) and has yielded mixed results with a tendency to be more activated when reflecting on their own videos (e.g., Seidel et al., 2011), our intervention targeting student teachers' interpretations of experiences shows a promising and easy-to-implement approach to foster student teachers' utility beliefs in the context of video-based studies.

1.4. The present study

In the present intervention study, we focused on the extent to which a video-based learning tool in combination with a structured essay-based writing assignment leads to changes in student teachers' utility values (cf. Kleinknecht and Schneider, 2013; Brouwer et al., 2017). Additionally, we focused on student teachers' emotional cost as a predictor of utility-value change to capture difficulties that student teachers perceive when using videos of their own teaching (cf. Chan et al., 2018) with quite plenty of simultaneous tasks to execute when

videos are generated (Richards et al., 2021). The combination of video-based learning in teaching practicums has been described as a promising conducive context as student teachers have contact with real students in authentic classroom situations, whereas the video setting enables them to collaboratively, supportively, and with discipline reflect on examples of their own teaching (Charalambous et al., 2018, 2022). Finally, by contrasting the effects of this intervention with a group using a protocol-based approach, we referred to findings emphasizing benefits concerning affective processes and transfer of learning outcomes (Koehler et al., 2005; Moreno and Valdez, 2007; Kleinknecht and Gröschner, 2016; Prilop et al., 2019; Weber et al., 2023). This study will contribute new evidence to the role of utility values and emotional cost for, but not exclusively, video-based learning in teacher education.

1.5. Research questions and hypotheses

The following research questions were investigated:

(RQ 1) How do utility values change in a group with video-based documentation of a lesson and systematic writing assignment (IG) compared to a group with protocol-based documentation of a lesson and unsystematic writing assignment?

We expect that harnessing student teachers' interpretation of experiences in the IG will increase latent difference scores compared to CG. (Hypothesis 1.1)

Furthermore, we expect significant interindividual variability in IG participants' latent difference scores and negative bivariate associations between their initial utility values and their latent difference scores since the writing assignment targets student teachers perceiving video-based learning as challenging. (Hypothesis 1.2)

(RQ 2) What effects arise concerning the influence of emotional cost on changes in utility values?

We hypothesize that latent difference scores in the IG are positively affected by emotional cost. More relevant, we assume, in contrast, negative associations regarding their initial utility values. This hypothesis was formulated with respect to the specifics arising if experiences are addressed systematically so that struggling student teachers can benefit. (Hypothesis 2)

2. Materials and methods

2.1. Sample

A full cohort of 102 student teachers ($M_{age} = 22.73$, $SD = 3.23$; 63.2% female) was enrolled in the study and gave their informed consent to participate. The Ministry of Education, Youth, and Sport granted ethical approval for the study. In their third academic year, student teachers attended a five-month teaching practicum in a German teacher education program for secondary education in all subjects. To answer our research questions, we implemented a pre-test-post-test control group design. Participants were randomly

assigned to an intervention group receiving a systematic SEVT-based writing assignment and video-based documentation intervention ($n_{IG} = 51$) or an unsystematic writing assignment and protocol-based documentation ($n_{CG} = 51$). Student teachers spent four days a week in different placement schools to which $n = 2$–3 preservice teachers were randomly assigned and one day on campus during the practicum.

2.2. Intervention

2.2.1. Lesson documentation and writing assignments

A schematic representation of the study is depicted in Figure 2. Student teachers had a passive run-in period after the first course in which they filled in a background survey. During this period, student teachers observe teaching and make their first experiences under mentorial observation. As part of the accompanying courses in educational sciences that started two months later, student teachers in the IG and CG were advised to plan a lesson with the scope of academically productive talk (see O'Connor et al., 2017). Student teachers uploaded a lesson plan that was further revised based on coursework. In the mid of the practicum, student teachers in both groups held a lesson based on their lesson plan in the placement school. IG participants videotaped the lesson and were provided technical support by a fellow student at the same placement school. In turn, CG participants' lesson was protocoled in written form by a fellow student who sat in. Finally, student teachers were instructed in the last course session (end of the teaching practicum) to reflect on the lesson as the major objective for their final papers that were credited with 10 ECTS. For this purpose, they chose a research question and consequently had to base their reflections on the lesson on a standardized approach comprising a three-step analysis of description, explanation, and suggesting alternatives (Seidel and Stürmer, 2014; Kleinknecht and Gröschner, 2016).

IG participants had to provide an additional ½ of a page in which they reflected on their achievement-related experiences (Eccles and Wigfield, 2020). Student teachers in both groups were requested to reflect on a self-chosen topic covered in the courses using course materials and notes. In line with Hulleman and Harackiewicz (2021), the essays of IG participants had to comprise

1. Personalized narratives on strengths, challenges, and how they overcame obstacles;
2. Strongly contextualized information on findings in the video analysis and theoretical claims made in their papers;
3. Implications for their future professional practice with information on what they would like to improve based on the video excerpts

The intervention approach, therefore, highlights aspects of Rosenzweig et al's (2020) cost-reduction intervention that aimed students by quotations of former physics students to reflect challenges positively to make them seem less costly. The control group, in contrast, was requested to provide ½ of a page with unstructured information regarding a topic covered in the accompanying courses. Information should be structured in a meaningful way by using own words without further requirements. The essay-based writing assignments were to be fulfilled within one month (washout period) after the pre-test.

2.2.2. Randomization test and manipulation check

Participants in the two groups did not significantly differ in gender ($\chi^2[1] = 3.50$, $p = 0.061$), age ($t[81] = 0.32$, $p = 0.747$), lessons taught in the placement school ($t[41.85] = 1.14$, $p = 0.261$) or the time (in minutes) per week that mentor teachers supported student teachers ($t[75] = -1.26$, $p = 0.212$). We conducted a manipulation check when the teacher educators graded the final papers. Therefore, teacher educators checked whether the essay comprised the three above-mentioned aspects (see Chapter 2.2.1). The quality of the essays was not coded. All participants in both groups complied with the lesson documentation and submitted their final papers. Only one participant in the IG did not provide the additional writing assignment.

FIGURE 2
Study design.
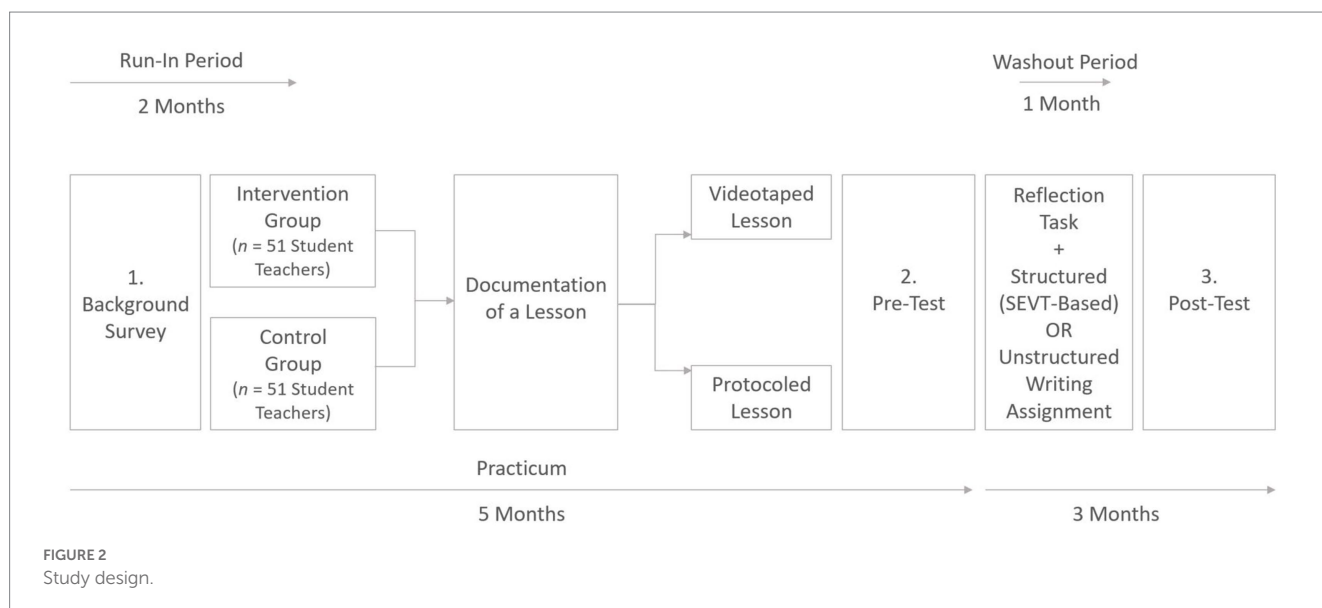
2.3. Data collection and operationalization

Utility values were assessed with a questionnaire at the end of the teaching practicum in the closing session of the accompanying course (pre-test) and three months later (post-test). Emotional cost was assessed only at the pre-test. The pre-test, hence, assessed student teachers' emotional-motivational beliefs in terms of the video-based vs. protocol-based reflection, whereas the post-test highlights how utility values changed after the writing assignments. In addition, a background survey captured student teachers' characteristics before the teaching practicum while also comprising the utility-value scale. Utility values, reflected by the facet utility for job, and emotional cost (see Appendix A Table A1 for all questionnaire items), were assessed on an instrument developed by Gaspard et al. (2015b). This instrument aimed for a finer-grained utility value and cost differentiation and has been applied in other interventions (e.g., Rosenzweig et al., 2020). We slightly modified the items in a content-specific way (see Dehne et al., 2018, for the validation of the instrument) and evaluated the scale reliability ρ for utility values and emotional cost (see Table 1).

2.4. Data analysis

Effects of the intervention were investigated using a multigroup latent change score model (LCSM) that is an extension of the traditional LCSM for between-group comparisons (i.e., non-invariance) of the model parameters (McArdle and Nesselroade, 1994; Kievit et al., 2018). An LCSM is a latent structural equation model for investigating interindividual differences in intraindividual change (Geiser, 2013). These models (see Figure 3) mimic Equation (1),

$$\text{State } 2 = 1 \times \text{State } 1 + 1 \times (\text{State } 2 - \text{State } 1) \qquad (1)$$

where State 2 (i.e., the post-test) is treated as a function of State 1 (i.e., the pre-test) and the difference between pre-test and post-test. Compared to autoregressive models, where the change is modeled only indirectly via residuals, LCSMs are part of the latent-state-model class and allow obtaining mean and variance estimates for the latent pre-test variable and latent difference variable (State 2 – State 1) directly. Different hypotheses can be tested: (1) the hypothesis of no mean differences over time ($\mu_\Delta = 0$); (2) the hypothesis of no interindividual differences in the

rates of change ($\sigma_\Delta^2 = 0$); (3) the hypothesis of no covariance between the latent baseline variable and latent difference variable (COV[State 1, State 2 – State 1] = 0) (McArdle and Nesselroade, 1994; McArdle, 2009); (4) the explanation of interindividual differences in the initial level (State1) or change scores (State2 – State1) by covariates (Geiser, 2013).

Across-time and between-group assumptions regarding the equality of the factor structure (configural invariance), the same meaning of indicators (metric invariance), and additionally, the intercepts (scalar invariance) need to be ensured for meaningful comparisons in latent mean scores (Steinmetz, 2018). Based on the sample size, we used item parcels as indicator variables for the longitudinally assessed utility-value construct (Little et al., 2002). As shown in Tables B1–B3 (Appendix B), scalar invariance between groups and across time was established.

To identify the mean of the latent pre-test variable and the mean of the latent difference variable, we used the "marker method" (Steinmetz, 2018). We consider Equation (2),

$$\overline{X}_i = \tau_i + \lambda_i \times \kappa + \varepsilon_i \qquad (2)$$

where τ_i defines the intercept, λ_i the factor loading, and κ the mean of the latent variable. As $E(\varepsilon_i) = 0$, ε_i can be omitted. By fixing the intercept of the reference indicator (or "marker variable") to "0," it follows that the mean of the latent variable is identified by the mean ($\overline{X}_i$) of the reference indicator as the factor loading is set to "1" for identification (Geiser, 2013). The mean of the latent difference variable indicates the change from the pre-test to the post-test, and the *value of p* indicates whether the estimate is significantly different from zero. Cohen's (1988) guidelines for effect sizes of the *d* family were considered, with values of 0.20 indicating small effects, 0.50 indicating moderate effects, and 0.80 indicating large effects. We computed an effects size (d_{ppc2}), taking pre-test scores into account (Morris, 2008). To test for associations between the latent baseline and difference variable and emotional cost (RQ 2), we computed a conditional multigroup LCSM.

2.5. Model estimation and fit evaluation

We evaluated the model fit using the χ^2 statistic, the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Hu and Bentler's (1999) cutoff criteria for model fit indices were applied.

TABLE 1 Descriptive statistics, sample items, intercorrelations, and composite reliability of study variables.

Variable	Sample item	IG			CG			Correlations		
		M	*SD*	ρ	*M*	*SD*	ρ	1.	2.	3.
Utility value	Good knowledge of reflecting on teaching and learning will help me in my future job.									
1. Pre-test		3.13	0.56	0.882	3.33	0.51	0.855	–	0.116	−0.570***
2. Post-test		3.46	0.43	0.834	3.33	0.49	0.827	0.392*	–	0.086
3. Emotional cost (pre-test)	When I deal with reflections, I get annoyed.	1.64	0.61	0.893	1.51	0.45	0.779	−0.550***	−0.393**	–

All variables were measured on a four-point Likert scale. Correlation coefficients listed above the diagonal are for the IG; below the diagonal for the CG. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$. Reliability estimates are based on previously conducted confirmatory factor analyses.
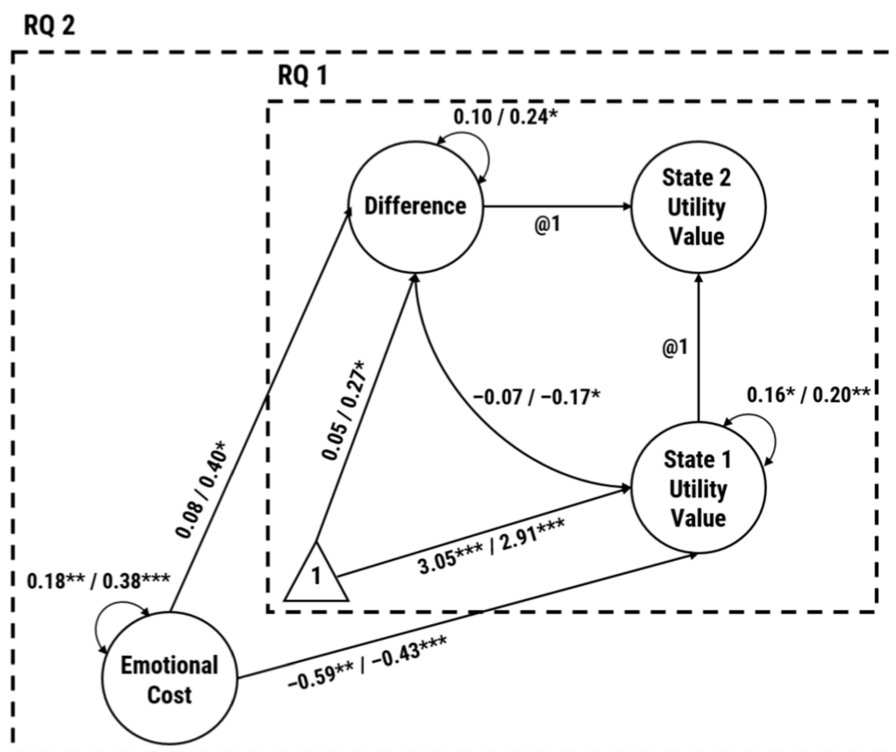
FIGURE 3
Multigroup latent change score model for testing unconditional and conditional utility-value change. Unstandardized estimates before the slash are for the CG and following the slash for the IG. The parameter estimates are corrected for measurement error. Indicator variables are intentionally left out. *p < 0.05, **p < 0.01, ***p < 0.001 (two-tailed p-value).

Therefore, CFI values ≥0.90 and ≥0.95 were considered acceptable or excellent regarding the model fit. An RMSEA of less than 0.06 was considered a good fit, whereas Little (2013), in turn, argues that values of 0.05–0.08 reflect an acceptable fit in longitudinal structural equation modeling. SRMR values of less than 0.08 are considered a good fit, although less is known regarding their performance in longitudinal studies. All models were estimated using M*plus* 8.6 (Muthén and Muthén, 1998–2021) using the maximum likelihood estimator. We used Mardia's skewness and kurtosis test implemented in M*plus* to investigate whether the maximum likelihood estimator was suitable for estimating the model parameters despite our small sample. We could retain the null hypothesis either for skew (p =0.080) or Kurtosis (p =0.480), indicating that our dependent variables were multivariate normally distributed. Missing data were addressed by full information maximum likelihood (FIML), as appearance was completely random (Little's MCAR test: $\chi^2[10]$ =12.63, p =0.245). The FIML approach has been shown to work well even with 50% of missing data and is preferable over listwise deletion with potential higher bias (Enders, 2010).

3. Results

3.1. *Post-hoc* power analysis

We tested the probability of the implemented hypothesis test to find an effect, given that differences in the population exist, using G*Power 3.1 (Faul et al., 2007). According to Cohen (1992), a power of at least 80% has to be achieved to draw meaningful inferences. We set input parameters for a one-tailed independent samples t-test (α = 0.05) and investigated power for medium effect size (Cohen's d = 0.5). The input parameters reflect a notable meta-analysis that found an average effect size of Cohen's d = 0.39 for interventions based on SEVT and Cohen's d = 0.49 for motivational interventions in general (Lazowski and Hulleman, 2016). The *post-hoc* power analysis revealed a power of 80.6%.

3.2. Changes in utility values

Descriptive statistics, correlations, and reliabilities are shown in Table 1.[1] We estimated models where we identified latent mean scores for both groups (marker method). The multigroup LCSM had eight degrees of freedom due to the invariance constraints on factor structure, factor loadings, and item intercepts and showed an excellent fit to the data

---

1   In pre-analysis, we tested for changes from the beginning (background survey) to the end of the teaching practicum (pre-test). The multigroup LCSM fitted the data well ($\chi^2[8]$ = 10.51, p = 231, CFI = 0.977, RMSEA = 0.071, 90% CI (0.000, 0.173), SRMR = 0.057]. Neither in the IG (M = 2.96, variance = 0.19, M = −0.06, variance = 0.28) nor in the CG (M = 3.03, variance = 0.20, M = 0.06, variance = 0.18), significant changes in utility values occurred across the teaching practicum.

($\chi^2[8] = 7.71$, $p = 0.462$, CFI = 1.00, RMSEA = 0.000, 90% CI [0.000, 0.160], SRMR = 0.076). Figure 3 shows the models with coefficients respective for groups. We found significant increases after the writing assignment for the IG ($M = 0.27$, $SE = 0.13$, $p = 0.030$) following a latent baseline score of 2.91 ($SE = 0.08$, $p < 0.001$). In contrast, a pre-test score of 3.05 ($SE = 0.08$, $p < 0.001$) in the CG was followed by a non-significant increase ($M = 0.05$, $SE = 0.08$, $p = 0.544$). The change for the IG was significantly higher than for the CG ($t[100] = 2.69$, $p = 0.008$, Cohen's $d_{ppc2} = 0.52$).[2] Differences in pre-test scores turned out to be non-significant ($t[100] = -1.68$, $p = 0.096$). The combination of a positive mean for the latent difference variable and a negative covariance in the IG (covariance = -0.17, $SE = 0.08$, $p = 0.024$) indicates that individuals with lower pre-test utility values showed higher increases after the intervention, whereas participants with higher pre-test scores showed lower increases. In the CG, the covariance of -0.07 turned out to be non-significant ($SE = 0.05$, $p = 0.136$). For both groups, we found significant interindividual differences in the pre-test scores (IG: variance = 0.20, $SE = 0.08$, $p = 0.009$; CG: variance = 0.16, $SE = 0.07$, $p = 0.022$). In contrast, only the variance of the difference variable in the IG (variance = 0.24, $SE = 0.12$, $p = 0.037$) indicated significant interindividual differences in student teachers' intraindividual change rates, whereas CG participants showed a rather low variability in the rate of change (variance = 0.10, $SE = 0.06$, $p = 0.104$). None of the between-group comparisons in the parameter estimates yielded statistically significant differences. Constraining the latent-variable variances ($\Delta\chi^2[2] = 1.45$, $p = 0.485$, ΔCFI = 0.000, ΔRMSEA = 0.000) or covariances ($\Delta\chi^2[3] = 1.87$, $p = 0.601$, ΔCFI = 0.000, ΔRMSEA = 0.000), therefore, did not indicate a worse model fit, whereas the SRMR continuously increased (model with equal variances + covariances: 0.164).

3.3. Predicting utility-value change

To answer RQ 2, we regressed the latent baseline and latent difference variable on emotional cost at the pre-test. The conditional multigroup LCSM showed an overall acceptable fit to the data [$\chi^2(46) = 62.10$, $p = 0.057$, CFI = 0.936, RMSEA = 0.083, 90% CI (0.000, 0.132), SRMR = 0.122]. As depicted in Figure 3, we found that emotional cost had a significant negative effect on baseline levels (IG: $\beta = -0.43$, $SE = 0.12$, $p < 0.001$; CG: $\beta = -0.59$, $SE = 0.18$, $p = 0.001$). The effect on the latent difference variable, however, was positive and only significant for the IG ($\beta = 0.40$, $SE = 0.19$, $p = 0.040$) but not for the CG ($\beta = 0.08$, $SE = 0.19$, $p = 0.666$). The conditional latent change score in the IG remained positive and significant ($M = 0.21$, $SE = 0.10$, $p = 0.031$; variance = 0.15, $SE = 0.07$, $p = 0.025$). Equality constraints on the unstandardized regression parameters did not lead to a substantially worse model fit ($\Delta\chi^2[3] = 4.65$, $p = 0.199$, ΔCFI = -0.006, ΔRMSEA = 0.001). However, the SRMR (ΔSRMR = 0.069) provided evidence for misfit. Scripts and outputs for all analyses described in the results section can be obtained from the Open Science Framework.[3]

---

2   We refer detailed to problems associated with the reference-group method for the identification of mean structures in multigroup LCSM under the following link on the Open Science Framework: https://osf.io/8fm53/?view_only=483 70a3ef7624402a1c38b1564f36402.

3   https://osf.io/8fm53/?view_only=48370a3ef7624402a1c38b1564f36402

4. Discussion

Student teachers' utility values have not been the target of systematic interventions to date. However, their implications could influence a wide range of studies. Against the backdrop of the claim that video-based learning "does not automatically activate emotional-motivational processes" (Kleinknecht and Schneider, 2013, p. 21), this study had two aims: First, to investigate utility-value change if student teachers' interpretations of experiences are systematically harnessed. Second, to consider the extent to which negatively considered task values accompanying video-based tasks, in which own teaching sequences are analyzed, influence motivational developments within the intervention. Crafted on SEVT (Eccles and Wigfield, 2020) and the seldom-regarded component of learners' interpretations of experiences, we investigated our research questions using a pre-test-post-test control group design after the end of a teaching practicum, in which student teachers' lessons were either videotaped (IG) or protocoled (CG).

4.1. Utility-value change

Regarding the first aim of our study, the results were consistent with our hypotheses. Utility values were promoted with a medium effect size compared to CG (confirming *Hypothesis 1.1*) while mirroring previous interventions and their effect sizes with a similar magnitude (see Lazowski and Hulleman, 2016). Reasons for learning might not be obvious (Wentzel and Brophy, 2014), even though videos of one's teaching might enhance student teachers' learning during a teaching practicum (Charalambous et al., 2022). Our study helped student teachers identify the utility of reflections on teaching and learning better than a protocol-based reflection with an unstructured writing assignment. Again, it must be referred to the promises and pitfalls of video application in teacher education. A vital review has pointed out that video-based reflections necessitate scaffolding to make them an effective learning experience for student teachers (Gaudin and Chaliès, 2015). Therefore, instructional guidance must appear highly structured to make video truly an opportunity to overcome the theory-practice gap (Korthagen, 2010). Plenty of studies (e.g., Brouwer et al., 2017) delivered structured viewing guides to their participants and ultimately led to enhanced targeted teaching behaviors such as dimensions of teaching quality (e.g., cognitive activation) for either student or in-service teachers. Research has agreed that facilitators (e.g., cooperating teachers, teacher educators, etc.) play a crucial role in guiding student teachers' focus on the events in the classroom and topics that matter (Seidel and Stürmer, 2014; Goldberg et al., 2021). If this aspect is not ensured, student teachers will feel overwhelmed by the many aspects to simultaneously reflect on and perceive less utility (Tripp and Rich, 2012). Even though this regards video-based tools, it should also enrich our understanding of learners' characteristics in a wide range of interventions. To this end and next to implications arising from motivational literature, the writing assignment in our study can be considered as working on student teachers' theory-practice relations in a way that a motivationally relevant experience resulted (*cf.* Hulleman and Harackiewicz, 2021). Courses in teacher education that make use of video provide *representations of practice* (Grossman et al., 2009). The degree of authenticity or extent to which an authentic representation is given is not merely the video. A video of one's teaching is more or less proximal to the teaching practices, although it attempts to offer a more

authentic learning opportunity for reflections on teaching and learning (McDonald et al., 2013). One way of moving forward to more intense and authentic experiences is to encourage student teachers to take a critical stance toward their achievement-related experiences with videos, to discover how they want to improve their teaching based on their experiences, and to let them seek relationships with the professional practice that teacher educators struggle to implement by task design (Kleinknecht and Schneider, 2013; Brouwer et al., 2017; Kang and van Es, 2019). The findings of this intervention study support the attempt that student teachers may be in the position to learn from video-based reflection of their own teaching.

Another aspect is related to the substantial variability in student teachers' latent difference scores. Lower scores before the intervention covaried with higher change scores or vice versa (confirming *Hypothesis 1.2*). In the context of utility-value-like interventions, it is less surprising that especially those with a history of poor achievement, higher perceptions of cost, and related lower utility values benefit. Therefore, the study by Canning et al. (2018) is noteworthy as they critically investigated the role of timing. Lower-performing students benefitted more regarding course grades and showed a tendency concerning enrollment in the second course when the writing assignment was implemented in the first unit compared to the last. The opposite was true for higher-performing students. In our intervention, the writing assignments on student teachers' achievement-related experiences were implemented at the end of the semester. Therefore, they did not mirror this aspect of Canning et al.'s findings. Altogether, these findings beg implications for video-based interventions. Thus, identifying predictors for changes in utility values, as a study by Canning et al. (2019) did, will help to verify the findings and aid subsequent attempts in considering student teachers' psychological characteristics.

4.2. The role of emotional cost in utility-value change

Regarding the second aim of our study, the findings were again consistent with the hypothesis. Implementing a writing assignment covering student teachers' perceptions of their achievement-related experiences influenced structural relations in the way that emotional cost predicted student teachers' utility-value change positively after the writing assignment (confirming *Hypothesis 2*). Apart from past explanations—studies comparing (student) teachers watching videos of their own vs. others' teaching (e.g., Kleinknecht and Schneider, 2013; Chan et al., 2018)—a more processual view could be held. Outcome-focused explanations highlighted discrepancies between expectations and the videotaped result as a possible trigger for negative emotions (Chan et al., 2018). However, it could be assumed that own teaching sequences will also activate mental representations of teacher-student relationships. According to a well-known relational model (Spilt et al., 2011), the association between classroom events (e.g., student misbehavior) and teacher well-being is mediated by mental representations of the teacher-student relationship (i.e., the degree of closeness and conflict). Video as an immersive experience shows merits in enabling student teachers to reflect retrospectively without situational pressure to act at home or in a video club with peers (van Es, 2012; Brouwer et al., 2017; Charalambous et al., 2018, 2022). However, there are also pitfalls. Repeatedly viewed adverse classroom events could trigger negative emotions, at least partially

mediated by teacher-student relationships. This perspective could guide future research.

After adjusting for baseline emotional cost, the latent change score in the IG remained significant and positive. In sum, the findings of the conditional LCSM underpin the efficacy of the writing assignment for the often-mentioned difficulties in the implementation of video-based tools (Chan et al., 2018). This study helped incapacitate troublesome characteristics of video-based tasks and reflected parts of Rosenzweig et al.'s (2020) intervention materials. It should be considered that much research has pointed to the negative influence of cost on the learning outcome as well as other subjective task values and expectancies (Eccles, 2005; Pekrun, 2006; Flake et al., 2015). For the latent pre-test utility values, this was the case as expected. For the latent difference scores, the results point out how the intervention helped student teachers with higher perceptions of cost after the teaching practicum, in turn stressing the need for researchers and teacher educators to reflect carefully on costly characteristics in video-based learning.

Our intervention provided a wise approach to harnessing and fostering emotional-motivational processes in a video-based task. Thus, video-based interventions can be supportive of situating student teachers' reflections; however, doing such tasks may also be a troublesome experience with increasing negative emotions (Kleinknecht and Schneider, 2013). Therefore, interventions during a teaching practicum need to be carefully designed to be a successful and high-leverage learning opportunity.

4.3. Limitations

Although the present study overcame several issues related to important previous exploratory research (e.g., Chan et al., 2018), some limitations should be taken into account. First, the study made it possible to compare the effects of video and a motivational intervention with a larger sample size than previous studies were able to accomplish. However, the study is—as many video-based studies and interventions in teaching practicums—largely depended on the context (e.g., countries and characteristics such as the duration). So-called "Many Labs"-projects would improve the confidence concerning the existence of effects from video-based studies and thus aid in overcoming their assumed situated character and dependence on aspects of the data collection (see Klein et al., 2014). A second restriction is that we considered item parcels as indicator variables due to the limited sample size (Little et al., 2002). As this is an established method to make structural equation models less complex and has been applied in similar research in teacher education (e.g., Gold et al., 2017), we achieved satisfactory model fit indices. However, risks such as estimation bias and model misspecification must be considered carefully (Matsunaga, 2008). Therefore, confirmatory factor analyses were conducted by Dehne et al. (2018) in a previous study, ensuring the assumed dimensionality with the original number of items. Furthermore, none of the between-group comparisons for the structural parameters (i.e., variances, covariances, and regression parameters) yielded significantly higher/lower overall estimates for the IG. As a result, comparisons between both groups have to be taken as tendencies. However, it has to be remarked that a couple of non-significant estimates (variance, covariance, and regression of the latent difference variable on emotional cost) contrasted statistically

significant estimates in the IG. Finally, one limitation of the study is that student teachers' essays were not examined using qualitative analyses. In most utility-value interventions, the written essays were not systematically investigated. However, exploratory research by Canning et al. (2019) has pointed to the importance of considering the quality of students' essay-based writing assignments. As this was beyond the scope of the present research, this perspective could guide future avenues of research.

4.4. Implications and future directions for research

This study has practical implications for the use of video in teacher education. Student teachers appreciate the systematic and contextualized use of video, and it helps them to identify the utility of reflecting on teaching and learning for their future careers. However, the use of video also has emotional costs, and not all student teachers benefit from it. That means teacher educators should carefully select video-based tasks or other digital tools and be aware of their consequences for student teachers' learning (cf. Kang and van Es, 2019). More immersive tools, such as virtual reality (VR), have become increasingly available and could enable a better opportunity regarding emotional-motivational processes. Two studies (Huang et al., 2022; Richter et al., 2022) compared VR with video in teacher education. Richter et al. (2022) concluded from their quasi-experimental study that reflective learning processes were equally triggered in both conditions. In contrast, reflection-related self-efficacy was only fostered in the VR group. Huang et al. (2022) conducted a pre-registered experiment and found favorable results for the VR condition in terms of interest and self-efficacy in classroom management. However, the extraneous cognitive load was also higher for the VR task, which would again require approaching negative task-related aspects. As a consequence, future research should take utility values and the role of emotional cost into consideration. These important predictors of student teachers' learning have yet to be investigated in the context of VR or other immersive technologies.

5. Conclusion

Video-based tools in teacher education provide a complex learning opportunity, and student teachers may need additional support to achieve the learning objectives. This support includes material artifacts as instructional tools to augment or prearrange videos for student teachers' learning. The study describes the successful implementation of a learning opportunity that orchestrated student teachers' interpretations of experiences in order to foster emotional-motivational beliefs. By critically investigating the role of psychological characteristics in a video-based intervention, this study adds value to the fields of educational psychology and teacher education as it addresses the less comprehensively studied parts of the SEVT model and the cost component. In future research, both utility values and emotional cost need consideration because they are critical psychological characteristics of participants accomplishing a video-based task. Promoting student teachers' motivation to reflect routinely is a professional task for teacher education (Kunter, 2013). Continuous use of learned content is at the heart of the teaching profession and is relevant in any educational setting. For this purpose, student teachers need to identify its utility value.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation and based on German data protection law.

Ethics statement

The study was reviewed and approved by the Thuringian Ministry of Education, Youth and Sport, Erfurt, Germany. Written informed consent to participate in this study was provided by the participants.

Author contributions

MD and AG conceptualized the idea, contributed to the study design, and collected data. MD was responsible for data curation, executing the formal analysis, writing, revising, and editing the original draft, and visualizing research findings. AG administered the project, acquired funding, and reviewed and approved the submitted manuscript. All authors contributed to the article and approved the submitted version.

Funding

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.2023.1166921/full#supplementary-material

References

Backfisch, I., Lachner, A., Stürmer, K., and Scheiter, K. (2021). Variability of teachers' technology integration in the classroom: a matter of utility! *Comp. Educ.* 166:104159. doi: 10.1016/j.compedu.2021.104159

Bardach, L., Klassen, R. M., and Perry, N. E. (2022). Teachers' psychological characteristics: do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative review. *Educ. Psychol. Rev.* 34, 259–300. doi: 10.1007/s10648-021-09614-9

Beck, C., and Kosnik, C. (2002). Components of a good practicum placement: pre-service teacher perceptions. *Teach. Educ. Quart.* 29, 81–98.

Brouwer, N., Besselink, E., and Oosterheert, I. (2017). The power of video feedback with structured viewing guides. *Teach. Teach.* 66, 60–73. doi: 10.1016/j.tate.2017.03.013

Canning, E. A., Harackiewicz, J. M., Priniski, S. J., Hecht, C. A., Tibbetts, Y., and Hyde, J. S. (2018). Improving performance and retention in introductory biology with a utility-value intervention. *J. Educ. Psychol.* 110, 834–849. doi: 10.1037/edu0000244

Canning, E. A., Priniski, S. J., and Harackiewicz, J. M. (2019). Unintended consequences of framing a utility-value intervention in two-year colleges. *Learn. Instr.* 62, 37–48. doi: 10.1016/j.learninstruc.2019.05.001

Chan, K. K. H., He, C., Ng, R. C. K., and Leung, J. S. C. (2018). Student teachers' emotions when watching their own videos and those of their peers. *J. Prof. Capit. Community* 3, 192–211. doi: 10.1108/JPCC-12-2017-0031

Charalambous, C. Y., Philippou, S., and Olympiou, G. (2018). Reconsidering the use of video clubs for student-teachers learning during field placement: lessons drawn from a longitudinal multiple case study. *Teach. Teach.* 74, 49–61. doi: 10.1016/j.tate.2018.04.002

Charalambous, C. Y., Philippou, S., Olympiou, G., and Georgiou, K. (2022). Experimenting with enablers and extenders to support ambitious teaching in mathematics: a video-club case study of student teachers during their field placement. *Teach. Teach.* 119:103874. doi: 10.1016/j.tate.2022.103874

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Dehne, M., Klaß, S., and Gröschner, A. (2018). Veränderung motivationaler Orientierungen im Praxissemester: Eine videobasierte Studie auf basis der Erwartungs-wert-Theorie [changes in motivational orientations during pre-service teacher practicum: a video-based study based on expectancy-value theory]. *Lehrerbildung auf dem Prüfstand* 11, 109–130.

Eccles, J. S. (2005). "Subjective task value and the Eccles et al. model of achievement-related choices" in *Handbook of competence and motivation*. eds. A. J. Elliot and C. S. Dweck (New York: Guilford Press), 105–121.

Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Ann. Rev. Psych.* 53, 109–132. doi: 10.1146/annurev.psych.53.100901.135153

Eccles, J. S., and Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemp. Educ. Psychol.* 61:101859. doi: 10.1016/j.cedpsych.2020.101859

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

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. Meth.* 39, 175–191. doi: 10.3758/BF03193146

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., and Welsh, M. E. (2015). Measuring cost: the forgotten component of expectancy-value theory. *Contemp. Educ. Psychol.* 41, 232–244. doi: 10.1016/j.cedpsych.2015.03.002

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., et al. (2015a). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Dev. Psychol.* 51, 1226–1240. doi: 10.1037/dev0000028

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., et al. (2015b). More value through greater differentiation: gender differences in value beliefs about math. *J. Educ. Psychol.* 107, 663–677. doi: 10.1037/edu0000003

Gaspard, H., Parrisius, C., Piesch, H., Kleinhansl, M., Wille, E., Nagengast, B., et al. (2021). The potential of relevance interventions for scaling up: a cluster-randomized trial testing the effectiveness of a relevance intervention in math classrooms. *J. Educ. Psychol.* 113, 1507–1528. doi: 10.1037/edu0000663

Gaudin, C., and Chaliès, S. (2015). Video viewing in teacher education and professional development: a literature review. *Educ. Res. Rev.* 16, 41–67. doi: 10.1016/j.edurev.2015.06.001

Geiser, C. (2013). *Data analysis with Mplus*. New York: Guilford Press.

Gold, B., Hellermann, C., and Holodynski, M. (2017). Effekte videobasierter Trainings zur Förderung der Selbstwirksamkeitsüberzeugungen über Klassenführung im Grundschulunterricht [Effects of video-based trainings for promoting self-efficacy in elementary classroom management]. *Z. Erzieh.* 20, 115–136. doi: 10.1007/s11618-017-0727-5

Goldberg, P., Schwerter, J., Seidel, T., Müller, K., and Stürmer, K. (2021). How does learners' behavior attract preservice teachers' attention during teaching? *Teach. Teach.* 97:103213. doi: 10.1016/j.tate.2020.103213

Goldman, R. (2007). "Video representations and the perspectivity framework: epistemology, ethnography, evaluation, and ethics" in *Video research in the learning sciences*. eds. R. Goldman, R. Pea, B. Barron and S. J. Derry (Mahwah, NJ: Lawrence Erlbaum), 383–395.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., and Williamson, P. W. (2009). Teaching practice: a cross-professional perspective. *Teach. Coll. Rec.* 111, 2055–2100. doi: 10.1177/016146810911100905

Harackiewicz, J. M., and Priniski, S. J. (2018). Improving student outcomes in higher education: the science of targeted intervention. *Ann. Rev. Psychol.* 69, 409–435. doi: 10.1146/annurev-psych-122216-011725

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., and Hyde, J. M. (2012). Helping parents to motivate adolescents in mathematics and science: an experimental test of a utility value intervention. *Psychol. Sci.* 23, 899–906. doi: 10.1177/0956797611435530

Huang, Y., Richter, E., Kleickmann, T., and Richter, D. (2022). Comparing video and virtual reality as tools for fostering interest and self-efficacy in classroom management: results of a pre-registered experiment. *Brit. J. Educ. Tech.* 54, 467–488. doi: 10.1111/bjet.13254

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

Hulleman, C. S., and Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science* 326, 1410–1412. doi: 10.1126/science.1177067

Hulleman, C. S., and Harackiewicz, J. M. (2021). "The utility-value intervention" in *Handbook of wise interventions: How social psychology can help people change*. eds. G. M. Walton and A. J. Crum (New York: Guilford Press), 100–125.

Kang, H., and van Es, E. A. (2019). Articulating design principles for productive use of video in preservice education. *J. Teach. Educ.* 70, 237–250. doi: 10.1177/0022487118778549

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., et al. (2018). Developmental cognitive neuroscience using latent change score models: a tutorial and applications. *Dev. Cogn. Neurosci.* 33, 99–117. doi: 10.1016/j.dcn.2017.11.007

Klassen, R. M., Kim, L. E., Rushby, J. V., and Bardach, L. (2020). Can we improve how we screen applicants for initial teacher education? *Teach. Teach.* 87:102949. doi: 10.1016/j.tate.2019.102949

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B. Jr., Bahník, Š., Bernstein, M. J., et al. (2014). Investigating variation in replicability: a "many labs" replication project. *Soc. Psychol.* 45, 142–152. doi: 10.1027/1864-9335/a000178

Kleinknecht, M., and Gröschner, A. (2016). Fostering preservice teachers' noticing with structured video feedback: results of an online– and video-based intervention study. *Teach. Teach.* 59, 45–56. doi: 10.1016/j.tate.2016.05.020

Kleinknecht, M., and Schneider, J. (2013). What do teachers think and how do they feel when they analyze videos of themselves teaching and of other teachers teaching? *Teach. Teach.* 33, 13–23. doi: 10.1016/j.tate.2013.02.002

Koehler, M. J., Yadav, A., Phillips, M. M., and Cavazos-Kottke, S. C. (2005). What is video good for? Examining how media and story genre interact. *J. Educ. Multim. Hyperm.* 14, 249–272.

König, J., Santagata, R., Scheiner, T., Adleff, A., Yang, X., and Kaiser, G. (2022). Teacher noticing: a systematic literature review of conceptualizations, research designs, and findings on learning to notice. *Educ. Res. Rev.* 36:100453. doi: 10.1016/j.edurev.2022.100453

Korthagen, F. A. J. (2010). Situated learning theory and the pedagogy of teacher education: towards an integrative view of teacher behavior and teacher learning. *Teach. Teach.* 26, 98–106. doi: 10.1016/j.tate.2009.05.001

Kunter, M. (2013). "Motivation as an aspect of professional competence: research findings on teacher enthusiasm" in *Cognitive activation in the mathematics classroom and professional competence of teachers*. eds. M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss and M. Neubrand (New York: Springer), 273–289. doi: 10.1007/978-1-4614-5149-5_13

Lazowski, R. A., and Hulleman, C. S. (2016). Motivation interventions in education: a Meta-analytic review. *Rev. Educ. Res.* 86, 602–640. doi: 10.3102/0034654315617832

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford.

Little, T. D., Cunningham, W. A., Shahar, G., and Widaman, K. F. (2002). To parcel or not to parcel: exploring the question, weighing the merits. *Struct. Equ. Model.* 9, 151–173. doi: 10.1207/S15328007SEM0902_1

Matsunaga, M. (2008). Item parceling in structural equation modeling: a primer. *Communic. Meth. Meas.* 2, 260–293. doi: 10.1080/19312450802458935

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Ann. Rev. Psychol.* 60, 577–605. doi: 10.1146/annurev.psych.60.110707.163612

McArdle, J. J., and Nesselroade, J. R. (1994). "Using multivariate data to structure developmental change" in *Life-span developmental psychology: Methodological contributions*. eds. S. H. Cohen and H. W. Reese (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), 223–267.

McDonald, M., Kazemi, E., and Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: a call for a common language and collective activity. *J. Teach. Educ.* 64, 378–386. doi: 10.1177/0022487113493807

Moreno, R., and Valdez, A. (2007). Immediate and delayed effects of using a classroom case exemplar in teacher education: the role of presentation format. *J. Educ. Psychol.* 99, 194–206. doi: 10.1037/0022-0663.99.1.194

Morris, S. B. (2008). Estimating effect sizes from pre-test-post-test-control group designs. *Org. Res. Meth.* 11, 364–386. doi: 10.1177/1094428106291059

Muthén, L.K., and Muthén, B.O. (1998–2021). *Mplus user's guide* (*8th ed.*). Los Angeles, CA: Muthén and Muthén.

Nickl, M., Sommerhoff, D., Böheim, R., Ufer, S., and Seidel, T. (2023). Fostering pre-service teachers' assessment skills in a video simulation. *Zeitschrift für Pädagogische Psychologie*. doi: 10.1024/1010-0652/a000362

O'Connor, C., Michaels, S., Chapin, S., and Harbaugh, A. G. (2017). The silent and the vocal: participation and learning in whole-class discussion. *Learn. Instr.* 48, 5–13. doi: 10.1016/j.learninstruc.2016.11.003

Part, R., Perera, H. N., Marchand, G. C., and Bernacki, M. L. (2020). Revisiting the dimensionality of subjective task value: towards clarification of competing perspectives. *Contemp. Educ. Psychol.* 62:101875. doi: 10.1016/j.cedpsych.2020.101875

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

Perez, T., Cromley, J. G., and Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *J. Educ. Psychol.* 106, 315–329. doi: 10.1037/a0034027

Piesch, H., Häfner, I., Gaspard, H., Flunger, B., Nagengast, B., and Harackiewicz, J. M. (2019). Helping parents support adolescents' career orientation: effects of a parent-based utility-value intervention. *Unterrichtswissenschaft* 47, 271–293. doi: 10.1007/s42010-018-0024-x

Prilop, C. N., Weber, K. E., and Kleinknecht, M. (2019). How digital reflection and feedback environments contribute to pre-service teachers' beliefs during a teaching practicum. *Stud. Educ. Eval.* 62, 158–170. doi: 10.1016/j.stueduc.2019.06.005

Richards, J., Altshuler, M., Sherin, B. L., Sherin, M. G., and Leatherwood, C. J. (2021). Complexities and opportunities in teachers' generation of videos from their own classrooms. *Learn. Cult. Soc. Inter.* 28:100490. doi: 10.1016/j.lcsi.2021.100490

Richter, E., Hußner, I., Huang, Y., Richter, D., and Lazarides, R. (2022). Video-based reflection in teacher education: comparing virtual reality and real classroom videos. *Comp. Educ.* 190:104601. doi: 10.1016/j.compedu.2022.104601

Rosenzweig, E. Q., Wigfield, A., and Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *J. Educ. Psychol.* 112, 166–182. doi: 10.1037/edu0000370

Santagata, R., König, J., Scheiner, T., Nguyen, H., Adleff, A.-K., Yang, X., et al. (2021). Mathematics teacher learning to notice: a systematic review of studies of video-based programs. *ZDM Math. Educ.* 53, 119–134. doi: 10.1007/s11858-020-01216-z

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., and Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: does it make a difference whether teachers observe their own teaching or that of others? *Teach. Teach.* 27, 259–267. doi: 10.1016/j.tate.2010.08.009

Seidel, T., and Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *Am. Educ. Res. J.* 51, 739–771. doi: 10.3102/0002831214531321

Song, Y., Rosenzweig, E. Q., and Barger, M. M. (in press). Disentangling emotional cost, psychological cost, and anxiety in motivation research. *Mot. Emo.*

Spilt, J. L., Koomen, H. M. Y., and Thijs, J. T. (2011). Teacher wellbeing: the importance of teacher–student relationships. *Educ. Psychol. Rev.* 23, 457–477. doi: 10.1007/s10648-011-9170-y

Steinmetz, H. (2018). "Estimation and comparison of latent means across cultures" in *Cross-cultural analysis: Methods and applications*. eds. E. Davidov, P. Schmidt, J. Billiet and B. Meuleman (New York: Routledge), 95–126.

Teo, T. (2009). Modelling technology acceptance in education: a study of pre-service teachers. *Comp. Educ.* 52, 302–312. doi: 10.1016/j.compedu.2008.08.006

Tripp, R. T., and Rich, J. R. (2012). The influence of video analysis on the process of teacher change. *Teach. Teach.* 28, 728–739. doi: 10.1016/J.TATE.2012.01.011

Tschannen-Moran, M., Hoy, A. W., and Hoy, W. K. (1998). Teacher efficacy: its meaning and measure. *Rev. Educ. Res.* 68, 202–248. doi: 10.3102/00346543068002202

Vandewaetere, M., Desmet, P., and Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Comp. Hum. Behav.* 27, 118–130. doi: 10.1016/j.chb.2010.07.038

van Es, E. A. (2012). Examining the development of a teacher learning community: the case of a video club. *Teach. Teach.* 28, 182–192. doi: 10.1016/j.tate.2011.09.005

Vermunt, J. D., and Endedijk, M. D. (2011). Patterns in teacher learning in different phases of the professional career. *Learn. Individ. Differ.* 21, 294–302. doi: 10.1016/j.lindif.2010.11.019

Walton, G. M. (2014). The new science of wise psychological interventions. *Curr. Direct. Psychol. Sci.* 23, 73–82. doi: 10.1177/0963721413512856

Weber, K. E., Prilop, C. N., and Kleinknecht, M. (2023). Effects of different video– or text-based reflection stimuli on pre-service teachers' emotions, immersion, cognitive load and knowledge-based reasoning. *Stud. Educ. Eval.* 77:101256. doi: 10.1016/j.stueduc.2023.101256

Wentzel, K. R., and Brophy, J. E. (2014). *Motivating students to learn*. New York: Routledge

Zeichner, K. (2010). Rethinking the connections between campus courses and field experiences in college– and university–based teacher education. *J. Teach. Educ.* 61, 89–99. doi: 10.1177/0022487109347671

Check for updates

The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial

Ping Wei*, Xiaosai Wang and Hui Dong

School of Foreign Languages, Tangshan Normal University, Tangshan, Hebei, China

Introduction: In the context of the burgeoning field of second language (L2) education, where proficient writing plays an integral role in effective language acquisition and communication, the ever-increasing technology development has influenced the trajectory of L2 writing development.

Methods: To address the need for enhanced writing skills among English as a Foreign Language (EFL) learners, this study investigates the efficacy of Automated Writing Evaluation (AWE) training. A randomized controlled trial employing repeated measures was conducted, involving a participant pool of 190 Chinese EFL students. The study comprehensively assessed the effects of AWE training, utilizing the Grammarly platform—an AI-driven program—on various dimensions of writing skills, encompassing task achievement, coherence and cohesion, lexicon, and grammatical accuracy. Control variables included writing self-efficacy and global English proficiency. Writing skills were evaluated through the administration of an International English Language Testing System (IELTS) writing sample test.

Results: The results unequivocally demonstrate that the experimental group consistently exhibited superior performance across all facets of writing skills compared to the control group. Furthermore, the predictive influence of pre-test scores was pronounced in task achievement, coherence and cohesion, and lexicon, highlighting the pivotal role of learners' initial proficiency levels in shaping subsequent writing outcomes. Notably, the emergence of writing self-efficacy as a significant predictor of task achievement and coherence and cohesion underscores the role of learners' beliefs and confidence in shaping their writing abilities.

Discussion: These findings conclusively suggest that Artificial Intelligence-based instructional programs, specifically AWE, hold the potential to effectively enhance second language writing skills, especially among learners with lower proficiency levels. This study carries crucial implications for EFL educators and researchers, advocating for the seamless integration of AWE into pedagogical strategies to foster a marked improvement in writing competence.

KEYWORDS

automated writing evaluation, L2 writing skills, EFL learners, randomized controlled trial, writing self-efficacy, complex script

1. Introduction

The integration of technology in language learning has gained increasing attention in recent years, with automated writing evaluation (AWE) tools being at the forefront of this development (Grimes and Warschauer, 2010; Liu et al., 2022). AWE is an AI-powered technology that leverages Natural Language Processing (NLP) to evaluate and provide feedback on written texts. These tools are capable of identifying a wide range of linguistic features, such as grammar, vocabulary, coherence, and organization (Link et al., 2022). The immediate and personalized feedback provided by AWE can be useful for students who may not have regular access to writing tutors or instructors (Saricaoglu and Bilki, 2021). AWE can provide feedback on different types of writing tasks, including essays, research papers, and business reports, making it a versatile tool for a variety of educational contexts, such as language learning, academic writing, and workplace training. AWE tools are commonly integrated into Learning Management Systems (LMS) or writing platforms, and can be used as part of online writing courses or as standalone tools (Zhai and Ma, 2022).

The use of AWE in language learning has been found to improve writing skills, increase writing fluency, and enhance writing accuracy (Ranalli et al., 2017; Zhang and Hyland, 2018; Zhang, 2020; Ngo et al., 2022; Nunes et al., 2022). However, English as a foreign language (EFL) learners might face challenges in acquiring proficient writing skills due to limited opportunities for practicing writing and receiving feedback from experts (Hyland, 2007; Storch, 2011). In China, for instance, students often have limited opportunities to practice writing and receive individual feedback from teachers, who may have a large number of students and limited time. To address these challenges, AWE-based instructional programs have been occasionally employed in EFL classrooms in China (Tang and Rich, 2017). Despite this trend, there is little empirical research investigating the effectiveness of AWE on EFL writing skills in China. This study aims to address this gap by conducting a randomized controlled trial to examine the impact of AWE on the second language writing skills of Chinese EFL learners.

Anchored in the Social Cognitive Theory (SCT; Bandura, 2001), which emphasizes the roles of self-efficacy, observational learning, and self-regulation in learning and behavior change, this research investigates the effectiveness of AWE-based instructional programs on the development of writing skills among Chinese EFL learners. SCT underscores the intricate interplay between personal factors, environmental influences, and behavior within learning contexts. This theoretical framework posits that individuals learn through observing others and their interactions with the environment, enabling the development of self-efficacy beliefs that significantly influence motivation and performance. In the context of our study, the SCT offers a lens through which to comprehend how the AWE intervention, by delivering immediate and personalized feedback, could enhance learners' self-efficacy in writing tasks, thus potentially impacting their overall writing skills.

Against this backdrop, the purpose of this study is to investigate the effectiveness of AWE as an AI-powered system on second language writing skills of Chinese EFL learners. Specifically, this research aims to address the following research questions:

1. What is the impact of AWE on L2 writing skills of Chinese EFL learners?
2. Is the effect of AWE on L2 writing skills of Chinese EFL learners mediated by writing self-efficacy?
3. Does the effect of AWE on L2 writing skills of Chinese EFL learners differ across proficiency levels?

This study contributes to the existing literature by providing empirical evidence on the effectiveness of AWE-based instructional programs on the development of writing skills of Chinese EFL learners. The outcomes of this research can inform educators and researchers on the potential benefits of AWE-based instructional programs in EFL writing instruction, particularly in non-English speaking contexts.

2. Literature review

2.1. Theoretical framework

2.1.1. The technology acceptance model

Adoption theories aim to provide logic and explanation for people's intention to whether utilize an activity for the first time (Wallace and Sheetz, 2014). Concerning the technology use, the technology acceptance model (TAM) proposed by Davis (1989) has been a valid theoretical model to measure a one's degree of technological acceptance and evaluating the quality of e-learning. Evolved from the Theory of Reasoned Action (TRA), TAM tries to explain why users are willing or not to adopt technologies when performing a task (Wu and Chen, 2017). Technically speaking, TAM attempts to delve into the impact of technology on individuals' behavior (Moon and Kim, 2001). In fact, by focusing on two major factors, namely perceived usefulness and perceived ease of use, TAM explains user willingness to integrate a particular kind of technology (Abdullah et al., 2016). More specifically, Venkatesh and Davis (1996) proposed the final version of TAM composed of four underlying sub-constructs: perceived usefulness (i.e., how much a user believes that her/his job performance would be boosted while using a specific technology), perceived ease of use (i.e., how much a user believes that utilizing a specific technology would be unchallenging and effortless), external variables (i.e., factors which are at play when accepting a particular technology, such as user training, user engagement in design, technology characteristics, and the process of incorporating the technology), and behavioral intention (i.e., user's behavior towards utilizing a specific technology determined by her/his perceived usefulness and perceived ease of use) (Davis, 1989; Marangunić and Granić, 2015; Dizon, 2016). Due to its significant contribution, the Technology Acceptance Model (TAM) has frequently been acknowledged as the most influential and widely used theory for explaining an individual's adoption of information systems (Lee et al., 2003). Despite the wide investigation of the effectiveness and acceptance of computer based-technologies using TAM (e.g., Al-Azawei et al., 2017; Li et al., 2019; Fathi and Ebadi, 2020; Al-Azawei and Al-Azawi, 2021), fewer studies have employed TAM in the context of EFL learning and teaching. Also, the TAM model is still sorely underappreciated and insufficiently understood when applied to EFL field of study, and a comprehensive TAM model still needs to be investigated instantly.

2.2. Integrating AWE into L2 writing

In the 1960s the pioneer work regarding the automated scoring application was first developed trying to save teachers' time when scoring written texts and allowing teachers to provide feedback on learners' manuscripts (Parra and Calero, 2019). More importantly, given the improvements in artificial intelligence technology which has significantly contributed to the process of natural language and intelligent language system, the programs for automated grading have been upgraded and promoted since the1990s (Liu et al., 2016). Consequently, numerous researchers have tried to develop computer-based applications and tools that can promote the writing skill and add value to scoring and feedback of it. As such, supported by the computer-mediated feedback technology, AWE is an ingenious technological tool that is implemented in various educational settings to provide evaluative feedback on learners' writing (Warschauer and Ware, 2006; Grimes and Warschauer, 2010). In fact, AWE is equipped with the kind of capacity that can constantly give qualitative and quantitative feedback on writing process by automatically scoring the text, analyzing the structure and creating a comprehensive evaluation of the text (Cotos, 2011; Li et al., 2014). The use of this technological tool is becoming increasingly common as a learning affordance in the learning process in various educational settings (Chen et al., 2009). Moreover, AWE is not only utilized for summative assessment in high-stakes writing tests but is also being effectively incorporated into classroom writing instruction.

According to Hassanzadeh and Fotoohnejad (2021), AWE plays a central role in the writing process, as it allows diagnostic and summative feedback to the learners. Furthermore, as Roscoe et al. (2017) asserted, AWE is a critical technological tool that saves teachers time when it comes to assessing writing, allows for more writing practice, and boost writing instruction. It is worth mentioning that one of the significant features of AWE tools is that they are interactive learning platforms. AWE tools often provide both build-in and customizable prompt for instructors to assign, as well as affording a diverse range of forms for the teacher to give comments on writing tasks (Palermo and Wilson, 2020). In addition, by using AWE, students are able to revise their manuscripts regarding the feedback they received from the source of the AWE tool, instructor, and peers (Geng and Razali, 2022).

Writing is often acknowledged as a demanding and intricate skill, particularly when it involves composing in a second language (Hashemian and Heidari, 2013; Marzban and Jalali, 2016; Hyland, 2019). This task becomes even more challenging for EFL learners, as acquiring writing proficiency in English poses difficulties not only for students but also for instructors (Cheung, 2016). As highlighted by Yu (2021), teaching writing skills, especially providing effective feedback on students' written work, can be a daunting endeavor for L2 teachers. However, writing is a skill that can be nurtured through consistent practice and timely feedback (Burstein et al., 2004; Fathi et al., 2020). When it comes to writing assessment, four distinct metrics that illuminate various facets of proficient written communication are usually employed (Polio, 1997; Uysal, 2010). The concept of task achievement, which gauges the extent to which a written piece fulfills given prompts or objectives, underscores the alignment between a writer's content and the prescribed context (Cumming, 2001). Coherence and cohesion, on the other hand, delve into the logical organization and seamless connection of ideas within a text, ensuring its fluidity and accessibility to readers (Hyland, 2019). The lexical dimension, encompassing vocabulary selection and precision, significantly contributes to the depth and richness of expression (Nation and Nation, 2001). Finally, the aspect of grammatical accuracy, a pivotal component of effective communication, involves the meticulous application of language rules to convey meaning with clarity and precision (Bitchener and Ferris, 2012).

With respect to the field of language education (i.e., EFL), computer-based technologies have offered innovative trends of language instruction and language assessment which can be used for writing development and writing evaluation (Yousefifard and Fathi, 2021; Hsu and Lin, 2022; Parmaxi, 2023). As an appropriate technology to meet these needs, AWE automated serving can aid teachers with evaluating the texts and act as supporter which allows language learners to experience a sense of freedom and plan their own time to promote their motivation. Moreover, AWE is a technological tool which can generate timely and supportive feedback for EFL students in order to promote their writing process (Wang et al., 2013; Li et al., 2019; Ngo et al., 2022). As Jiang et al. (2020) demonstrated, AWE is an integral software that can significantly exert influence on L2 learners' writing skills. AWE computer-based programs can act as tools to evaluate EFL students' writing output and generate unique and individualized feedback (Jingxin and Razali, 2020; Fu et al., 2022). It is well-documented that the automated feedback provided by AWE in Second Language Acquisition (SLA) classrooms can offer significant benefits, such as writing longer texts, acquiring promoted machine scores, making fewer errors in essays, and boosting the rhetorical quality of writing texts (Li et al., 2015; Parra and Calero, 2019; Xu and Zhang, 2022). As put forward by Jingxin and Razali (2020), in L2 classrooms, AWE tools can offer authentic synchronous scores (i.e., holistic and analytic scores), as well as providing automated personalized diagnostic feedback on L2 students' manuscripts in various features of writing traits.

L2 teachers can integrate a variety of automatic feedback programs in classrooms to help them while teaching writing skills like wikis, MS Word computer software, and Grammar software among others (Zhang and Hyland, 2018; Stevenson and Phakiti, 2019). As one of the efficient automated feedback tools, Grammarly can be incorporated in L2 instruction to help learners and instructors in promoting EFL writing skills (Ebadi et al., 2023). Grammarly is an example of AWE that has gained particular attention as a practical tool in EFL classrooms. It can be integrated in L2 writing instruction to recognize structure deviations of texts, review spelling, punctuation, and check the originality to ensure that the text is mistake-free, clear, and polished (Ghufron and Rosyida, 2018; Barrot, 2022). This program, which is connected to the Internet, provides alternative words that are relevant if there are wrong words in the English language. Furthermore, Grammarly is incorporated into the Microsoft Word application which makes it a less demanding tool for learners to use to review deviations in English grammar with computer and suggests clarifications or samples of well-formed sentences and/or words. More importantly, the real-time writing feedback of Grammarly can assist EFL teachers to prevent writing deviations (Qassemzadeh and Soleimani, 2016). Grammarly contains an AI method which puts together deep learning and some approaches to natural language analysis in order to review grammatical constructs, phrases, paragraphs and written texts.

Previous research evidence indicates that AWE can greatly affect L2 students' writing skills (e.g., Liao, 2016a,b; Roscoe et al., 2017; Khoshnevisan, 2019; Jingxin and Razali, 2020; Lee, 2020; Tambunan et al., 2022; Waer, 2023). For instance, Liao (2016b) investigated the influence of the AWE-based approach in improving the writing accuracy of EFL students. To this end, 63 EFL learners took part in the study. Developing a15-item questionnaire and a 12-question interview protocol, the findings indicated that AWE enhanced the writing accuracy of learners. In another study, Lee (2020) conducted a longitudinal study to explore the effects of AWE on Korean university learners' English writing competence. The perceptions towards their writing development which was acquired via interviews and journal entries were explored as well. Using a mixed-methods research design, the authors pointed to potential benefit of AWE, as it increased writing development and writing fluency of EFL students. In the context of Egypt, Waer (2023) explored the potential role of AWE in affecting EFL learners' writing process and grammatical competence. The findings revealed that AWE reduced writing apprehension and promoted the grammatical knowledge of participants. Also, Liao (2016b) examined the impact of AWE applications in shaping the writing improvement in an EFL context. The findings revealed that AWE facilitated the writing accuracy and writing development of EFL students. In their study, Saricaoglu and Bilki (2021) investigated EFL students' written language under the influence of AWE. The findings indicated that EFL students' engagement with AWE significantly reduced their errors in writing and promoted their writing accuracy. Employing a mixed methods design, Wang et al. (2013) investigated the effect of integrating AWE in EFL university students' writing. Their outcomes revealed that AWE substantially enhanced EFL students' writing accuracy and promoted their autonomy awareness. With respect to other AWE tools, namely Grammarly, there are few studies which have examined the role of this tool in L2 writing (e.g., Khoshnevisan, 2019; Parra and Calero, 2019). Integrating Grammarly as an AWE tool, Khoshnevisan (2019) investigated the role of this software in developing and honing learners' writing skills. Gathering data from a sample of 12 students, the findings demonstrated that Grammarly contributed to participants' writing by motivating learners to develop their English writing competencies and produce more accurate essays. Moreover, it was found that Grammarly promoted English writing development by offering practical tips about grammar, vocabulary, and punctuation. Similarly, Parra and Calero (2019) found in their study that Grammarly was greatly conducive to EFL students' writing accuracy.

Despite the contributions of AWE tools (i.e., Grammarly) to EFL writing competencies, previous studies have mentioned some limitations to these technological programs. For instance, Stevenson and Phakiti (2014) demonstrated in their study that there is not much certainty regarding the positive effects of AWE on writing process, as it may not generate improvements in writing proficiency. The reason behind this may be attributed to the fact that computers-based technologies do not have the required judgement to evaluate those elements that are often associated with adequate writing, such as logic, clarity, accuracy, fluency, and relevancy. As Liao (2016a) demonstrated, AWE tools cannot perform imperfectly while addressing written language concerns (i.e., meaning, idea development, humor or irony, features of writing in which higher-order thinking is needed, quality of evidence, to name just a few). Therefore, due to its limitations and

drawbacks, AWE needs to be employed as a supplemental instrument rather than a substitute for instructor feedback.

Taken together, while many researches have mainly focused on the processes and perspectives of L2 teachers and learners, few researches have examined the role of AWE tools in affecting the L2 writing skills and competencies. Furthermore, most of the previous studies have investigated commercial AWE tools while neglecting others, namely Grammarly. More importantly, so far, the research regarding the integration of AWE in SLA domain is in nascent stages and little is known about this computer-based tool. In addition, to the best knowledge of the researcher, so far, few (if any) studies have surveyed the effects of AWE tools, namely Grammarly on EFL students' writing development. Hence, as an attempt to fill this research lacuna, the present study delved into the role of AWE in affecting L2 writing development and accuracy of EFL students, with a focus on the use of Grammarly.

2.3. The present study

The present study aimed to investigate the effectiveness of an Automated Writing Evaluation (AWE) tool on the second language writing skills of Chinese EFL learners. A randomized controlled trial (RCT) design was used, with participants being randomly assigned to either the experimental group or the control group. Both groups received 12 weeks of instruction, during which the treatment group underwent AWE-based instruction. In this instruction, participants utilized an AWE tool, Grammarly, to submit their written essays each week. The AWE tool provided immediate feedback on various aspects of writing, including grammar, spelling, vocabulary, and organization. Additionally, the treatment group attended weekly one-hour writing workshops designed to enhance their writing skills and incorporate the feedback from the AWE tool. In contrast, the control group received traditional writing instruction without the integration of AWE or the additional writing workshops. Based on the literature, it was hypothesized that the AWE-based writing instruction would lead to improvements in students' writing skills, as reflected by the four measures used in this study.

3. Methods

3.1. Participants

The AWE-based writing evaluation intervention was administered as an extracurricular program targeting intermediate EFL students in Mainland China. Informed written consent was obtained from all participants prior to their involvement in the intervention. The study cohort comprised 190 intermediate EFL students (60% female), all of whom were enrolled in one of four distinct writing courses hosted by different institutes offering the writing intervention. The participants' mean age was 21.5 years ($SD = 2.8$; range: 18–28 years).

To rigorously assess the efficacy of the writing intervention, we adopted a randomized controlled trial (RCT) design with repeated measures (Friedman et al., 2010). Initial measurements were conducted as pretests, seamlessly integrated into the first two sessions of the respective course. Subsequent posttest measurements were conducted during the final session of the course.

Within each institute, a control group was established, participating in a conventional writing course. Importantly, both the intervention and control courses were conducted simultaneously. The implementation of the AWE-based writing evaluation was overseen by a team of researchers collaborating with two proficient English teachers. It is essential to note that the AWE-based writing intervention remained consistent across all groups.

The randomization process was facilitated by bundling the two course options (AWE-based and conventional) under a single course-tandem, aptly named the *English Writing Course*. Enrollment into the course-tandem was exclusive, thus ensuring a controlled environment for the study. Post-enrollment, a blocked randomization technique was employed, utilizing computer-generated random numbers to allocate students to either the control or experimental groups. Through this approach, an equitable distribution of students was achieved across all participating institutes.

In total, 95 students were randomly assigned to the AWE-based writing intervention group (average age: $M = 21.6$, $SD = 2.9$; 60% female), while another 95 students were assigned to the traditional writing course group (average age: $M = 21.4$, $SD = 2.7$; 40% female). Following the study's completion, all students were invited to engage in the alternate course as a continuation of their learning process.

3.2. Instruments

3.2.1. Writing skills

In this study, two sample tasks from the International English Language Testing System (IELTS) were used to measure the writing skills of the participants. A pre-test task was administered to all participants before the intervention, serving as a baseline measure. Subsequently, a post-test task was given to both the experimental and control groups after the completion of the 12-week instructional period, which included the AWE-based instruction and traditional writing instruction without AWE, respectively. A pre-test task was administered before the intervention, while a post-test task was given after the AWE-based instruction. The writing performance of the participants was assessed using an analytic essay scoring scale based on the IELTS rubric.

The IELTS rubric, renowned for its reliability, is extensively employed for assessing writing abilities within second language contexts. This rubric employs a range of scores, typically from 1 to 9, to evaluate distinct descriptors across various dimensions of writing, such as task accomplishment, coherence and cohesion, lexical richness, and grammatical precision. Each criterion encompasses specific descriptors that correspond to different levels of proficiency, and these descriptors are scored individually within the established score range.

The final score derived from the IELTS rubric is calculated as the mean score of the descriptors. In this method, each descriptor's score is assigned a weight based on its significance within the overall writing competence. The individual scores for task achievement, coherence and cohesion, lexical resource, and grammatical accuracy are averaged to determine the participant's final writing proficiency score. This approach provides a comprehensive and nuanced assessment of the participants' writing skills, accounting for their performance across a spectrum of criteria.

The selection of the IELTS rubric for the analytic essay scoring scale was based on its comprehensive nature and established reliability and validity in assessing writing skills. By employing the IELTS rubric, this study ensured a standardized and consistent evaluation of participants' writing performance, enabling a reliable comparison of their progress and the impact of the AWE intervention. To ensure the consistency of the scoring process, two independent raters were recruited, and inter-rater reliability was calculated using Cohen's Kappa, which was reported to be 0.82.

3.2.2. Global English proficiency

To evaluate the participants' general English language proficiency and ensure their comparability, the Oxford Placement Test (OPT) developed by Allan (2004) was employed. The OPT is a versatile assessment tool that accurately determines the appropriate proficiency level for English learners, evaluating dimensions such as vocabulary, grammar, listening comprehension, and reading skills. The internal consistency of the OPT, assessed using Cronbach's alpha, yielded a reliability coefficient of 0.83 in this study, indicating a high level of internal reliability.

To enhance the comparability of the OPT scores with the IELTS rubric, the total scores obtained from the OPT were transformed onto a 0–9 scale. This conversion was undertaken to align the OPT scores with the scoring scale familiarly associated with the IELTS rubric. This approach allowed for a consistent interpretation of participants' language proficiency across both assessments, providing a unified framework for evaluating their language skills.

3.2.3. Writing self-efficacy

To measure the writing self-efficacy of L2 students, the scale developed by Han and Hiver (2018) was utilized. This scale consisted of seven items designed to assess students' beliefs and assurance in their writing abilities. The questionnaire adopted a 5-point Likert scale format, ranging from 1 (strongly disagree) to 5 (strongly agree). The internal consistency of the scale, assessed using Cronbach's Alpha, was found to be 0.78 in the present investigation.

3.2.4. Procedure

The experimental intervention in this study aimed to enhance the L2 writing competencies of Chinese EFL learners through the use of an AWE tool. The AWE tool was provided by Grammarly and was used by the students to submit a written essay in English every week for a period of 12 weeks. The tool provided immediate feedback on various aspects of writing, including grammar, spelling, vocabulary, and organization. The feedback was given in the form of suggested corrections and explanations, which the students were encouraged to review and incorporate into their subsequent writing.

In addition to the AWE tool, the students in the experimental group received a weekly one-hour writing workshop that focused on developing their writing skills and providing additional opportunities for practice. The writing workshop was designed to complement the AWE tool by giving learners the individualized feedback on their writing, as well as guidance on how to improve their writing skills. The workshop covered various aspects of writing, including grammar, vocabulary, sentence structure, and organization.

On the other hand, the control group in this study received traditional writing instruction without the use of an AWE tool. The students in the control group were asked to write an essay in English

every week for a period of 8 weeks, which were graded by the instructor based on a rubric that evaluated various aspects of writing, including grammar, spelling, vocabulary, and organization. The students in the control group also received a weekly one-hour writing workshop that was similar in content and structure to the workshops provided to the experimental group. However, the writing workshops in the control group did not include the use of an AWE tool.

Overall, the experimental intervention in this study aimed to improve the second language writing skills of Chinese EFL learners by providing them with immediate feedback on their writing using an AWE tool and additional opportunities for practice through weekly writing workshops. The control group, on the other hand, aimed to improve the second language writing skills of Chinese EFL learners through traditional writing instruction without the use of an AWE tool. The effectiveness of these two approaches was compared to determine the impact of AWE on the second language writing skills of Chinese EFL learners.

3.3. Ensuring treatment fidelity

In order to ensure the validity of the results, treatment fidelity was closely monitored across all groups. To achieve this, a guideline was developed to provide the instruction to two pilot groups of 10 and 12 EFL students prior to the actual study. The teaching materials and course content were standardized for all groups and given in the same order. In addition, the pretest and all trainings were conducted by the research team to ensure consistency and fidelity to the experimental design. By implementing these measures, the study ensured that the intervention was delivered as intended and that any differences observed between the experimental and control groups could be confidently attributed to the use of the AWE tool. This approach is consistent with prior research on treatment fidelity (Graham and Harris, 2014) and strengthens the internal validity of the study.

3.4. Analysis

To evaluate the effectiveness of the AWE-based instruction, four measures including task achievement, coherence and cohesion, lexicon, and grammatical accuracy were used. To enhance the accuracy of the regression coefficients and mitigate potential biases resulting from between-group differences at the study's outset (Cohen et al., 2003), control variables were incorporated. The first control variable was global English proficiency, which was measured using the OPT. The second control variable was writing self-efficacy, which was included due to its potential impact on writing performance. To further explore the impact of the AWE-based instruction, the interaction term of the course and pretest score was included as an additional predictor variable. This allowed us to assess the differential effects of the intervention for EFL students with low versus high pretest scores on the dependent variable.

To ensure that the training conditions did not differ significantly at the outset of the research, two-tailed t-tests were performed to examine the pretest measures for all dependent and control variables. The baseline equivalence was examined for key characteristics. The dependent variables included posttest measurements for each of the four sub-scales (i.e., Task Achievement, Coherence and cohesion,

Lexicon, Grammatical accuracy) of writing skills. To evaluate the intervention's efficacy, multiple linear regression analyses were employed using Mplus Version 7 (Muthén and Muthén, 1998–2012) with maximum likelihood robust estimation (MLR). The predictors were entered simultaneously into the multiple linear regression model. This approach allowed us to examine the collective impact of all predictors on the dependent variable, writing skills. By including all predictors together, we aimed to understand how their combined effects contribute to explaining the variance in writing skills among the participants. The percentage of missing data ranged from 2 to 6%, and there was no significant differential drop-out between the treatment and control groups [$\chi^2(1, 190) = 1.08$; $p = 0.299$]. Significance tests were one-tailed, with a significance level (α) set at 0.05. Hypotheses were formulated in a directed manner to examine the training effects.

The full-information maximum likelihood (FIML) estimator was used to handle missing data, assuming that the missing data were missing at random (Enders, 2010). FIML analysis is a statistical approach that utilizes all available data to estimate parameters and standard errors (Buhi et al., 2008). Prior to the analyses, continuous variables were standardized. The experimental and control groups were represented as binary variables, with a value of 1 assigned to the experimental group and 0 to the control group. The magnitude of the intervention impact was assessed by comparing the standardized mean differences (Hedges, 2007). As no similar studies were found, the widely accepted classification of effect sizes was employed: small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$) (Cohen, 1992). Since treatment effects were assessed across four dependent variables, the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) was employed to control for multiple testing, and adjusted p-values were reported.

4. Results

Table 1 displays the means and standard deviations for the experimental and control groups in the pre- and post-tests. The experimental group received Automated Writing Evaluation (AWE) intervention while the control group received traditional writing instruction. The table shows the scores for task achievement, coherence and cohesion, lexicon, grammatical accuracy, global English proficiency (measured by OPT), and writing self-efficacy. The pretest means for both groups were similar for all measures, and no significant differences were found. However, at the posttest, the experimental group showed higher mean scores in all measures than the control group. The missing data ranged from 2 to 6% across both groups, with the higher missing rate resulting from students' absence at posttest. The missing data in this study were determined to be missing at random. The interclass correlation coefficients (ICC) for all measures were above 0.70, indicating acceptable levels of reliability.

Table 2 presents the correlations at the pretest (below diagonal) and posttest (above diagonal) in the study. The table shows the Pearson's correlation coefficients between the four variables measured in the study.

The table indicates that there is a statistically significant positive correlation between task achievement and coherence and cohesion at both the pretest ($r = 0.24$, $p < 0.05$) and the posttest ($r = 0.43$, $p < 0.01$).

TABLE 1 Means and standard deviations for each group in pre- and post-tests.

| | Pre-test | | | | | Post-test | | | | | |
| | Experimental G | | Control G | | | Experimental G | | Control G | | | |
	M	SD	M	SD	MIS	M	SD	M	SD	MIS	ICC
Task	4.26	0.83	4.12	0.61	6	5.42	0.69	4.78	0.83	11	0.82
Coherence	4.12	0.69	4.20	0.72	6	5.28	0.74	5.12	0.59	11	0.79
Lexicon	5.14	0.92	4.98	0.49	6	6.02	0.85	5.36	0.97	11	0.71
Accuracy	3.86	0.68	3.97	0.82	6	5.76	0.92	4.68	0.68	11	0.90
Global Eng	5.46	1.24	5.33	0.89	3						
WSE	3.57	0.72	3.71	0.92	3						

Task, task achievement; Coherence, coherence and cohesion; Accuracy, grammatical accuracy; Global Eng, Global English proficiency (OPT); WSE, writing self-efficacy.

TABLE 2 Correlations among the constructs at the pretest (below diagonal) and the posttest (above diagonal).

	1	2	3	4
(1) Task achievement	–	0.22*	0.37**	0.31**
(2) Coherence and Cohesion	0.24*	–	0.43**	0.38**
(3) Lexicon	0.27*	0.22*	–	0.41**
(4) Grammatical accuracy	0.32**	0.26*	0.33**	–

*$p < 0.05$; **$p < 0.01$.

There is also a significant positive correlation between task achievement and lexicon at the pretest ($r = 0.27$, $p < 0.05$) and the posttest ($r = 0.41$, $p < 0.01$). Similarly, there is a significant positive correlation between coherence and cohesion and lexicon at both the pretest ($r = 0.22$, $p < 0.05$) and the posttest ($r = 0.38$, $p < 0.01$).

Furthermore, there is a significant positive correlation between grammatical accuracy and task achievement at the pretest ($r = 0.32$, $p < 0.01$) and posttest ($r = 0.33$, $p < 0.01$). There is also a significant positive correlation between grammatical accuracy and coherence and cohesion at the pretest ($r = 0.26$, $p < 0.05$) but not at the posttest.

Table 3 reports the results of an analysis of the effects of AWE-based instruction on writing skills (posttest) in terms of task achievement, coherence and cohesion, lexicon, and grammatical accuracy.

In terms of the intervention effect, the results show that AWE-based instruction has a significant positive effect on task achievement ($B = 0.38$, $SE = 0.27$, $p = 0.044$), coherence and cohesion ($B = 0.46$, $SE = 0.32$, $p = 0.036$), lexicon ($B = 0.55$, $SE = 0.31$, $p = 0.009$), and grammatical accuracy ($B = 0.74$, $SE = 0.29$, $p = 0.003$).

The results also show that the pretest score is a significant predictor of task achievement ($B = 0.26$, $SE = 0.19$, $p = 0.197$), coherence and cohesion ($B = 0.34$, $SE = 0.26$, $p = 0.046$), and lexicon ($B = 0.33$, $SE = 0.24$, $p = 0.245$), but not of grammatical accuracy ($B = 0.19$, $SE = 0.13$, $p = 0.573$).

The interaction effect between the intervention and pretest score is not significant for task achievement ($B = 0.29$, $SE = 0.24$, $p = 0.621$) and coherence and cohesion ($B = 0.21$, $SE = 0.17$, $p = 0.263$), but is significant for lexicon ($B = -0.39$, $SE = 0.24$, $p = 0.048$) indicating that the effect of the intervention on Lexicon is weaker for participants who had higher pretest scores.

Global English proficiency is not a significant predictor of any of the outcome variables. However, writing self-efficacy (WSE) is a

significant predictor of task achievement ($B = 0.39$, $SE = 0.19$, $p = 0.024$) and Coherence and cohesion ($B = 0.52$, $SE = 0.30$, $p = 0.013$), but not of Lexicon ($B = 0.26$, $SE = 0.19$, $p = 0.427$) or grammatical accuracy ($B = 0.24$, $SE = 0.17$, $p = 0.092$).

The explained variance (R^2) shows that the AWE-based instruction accounts for 31% of the variance in task achievement, 28% in coherence and cohesion, 35% in lexicon, and 46% in grammatical accuracy.

Finally, the omnibus test for the overall model was statistically significant ($F = 17.12$, $p < 0.001$), indicating that the combination of predictors significantly improved the fit of the model compared to a null model. This suggests that the included predictors collectively contribute to the prediction of writing skills among the participants.

5. Discussion

The present study aimed to investigate the effectiveness of AWE-based instruction on the second language writing skills of Chinese EFL learners. More specifically, the researchers examined the effects of AWE-based instruction on task achievement, coherence and cohesion, lexicon, and grammatical accuracy, while also considering global English proficiency, writing self-efficacy, and pre-test scores as control variables. The results indicate that AWE-based instruction had a significant positive effect on L2 writing skills.

The positive effect of AWE-based instruction on task achievement is consistent with previous studies that have highlighted the role of automated feedback in enhancing learners' ability to meet specific writing task requirements effectively (Liao, 2016a; Jiang et al., 2020; Barrot, 2022; Jiang and Yu, 2022). Via providing immediate and targeted feedback, AWE systems enable learners to identify and address gaps in task achievement, leading to improved performance. Similarly, the positive impact of AWE-based instruction on coherence and cohesion supports previous research highlighting the role of technology in promoting cohesive and well-structured writing (Tuzi, 2004; Li et al., 2019; Kessler, 2020; Rahimi and Fathi, 2022). AWE systems can assist learners in identifying and rectifying issues related to paragraph organization, sentence connections, and the overall flow of ideas (Cotos, 2011). The improvement observed in lexicon and grammatical accuracy can be attributed to the automated features of AWE systems, which enable learners to receive detailed feedback on vocabulary use and grammatical errors (Zhang and Hyland, 2018; Fu et al., 2022). The immediate feedback provided by AWE systems allows learners to

TABLE 3 AWE-based instruction effects on the writing skills (posttest).

	Task achievement			Coherence and cohesion			Lexicon			Grammatical accuracy		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p
Intervention	0.38	0.27	0.044	0.46	0.32	0.036	0.55	0.31	0.009	0.74	0.29	0.003
Pretest score	0.26	0.19	0.197	0.34	0.26	0.046	0.33	0.24	0.245	0.19	0.13	0.573
Intervention × Pretest score	0.29	0.24	0.621	0.21	0.17	0.263	−0.39	0.24	0.048	−0.30	0.26	0.140
Global English	0.75	0.32	1	−0.32	0.24	0.186	−0.19	0.15	0.792	0.16	0.13	0.954
WSE	0.39	0.19	0.024	0.52	0.30	0.013	0.26	0.19	0.427	0.24	0.17	0.092
Explained variance (R^2)	0.31			0.28			0.35			0.46		

identify and correct lexical and grammatical issues, leading to enhanced language accuracy (Ranalli, 2018; Zhang, 2020).

Taken together, the findings of this study provide empirical evidence supporting the positive impact of AWE-based instruction on multiple components of writing skills among Chinese EFL learners. These results align with previous studies that have also reported the effectiveness of AWE tools on enhancing second language writing skills (e.g., Liao, 2016a,b; Ranalli et al., 2017; Li et al., 2019; Hassanzadeh and Fotoohnejad, 2021; Barrot, 2022; Fu et al., 2022; Ebadi et al., 2023).

One likely reason for the effectiveness of AWE-based instruction is that it provides immediate and personalized feedback to learners, allowing them to identify and correct their errors in real-time. This feature is especially beneficial for low-proficiency learners who may struggle with self-correction and need more guidance in their writing process.

This finding aligns with previous research on the benefits of technology in language learning, particularly in improving writing skills (Tuzi, 2004; Stapleton and Radia, 2010; Kessler, 2020; Fathi and Rahimi, 2022). This study is anchored in the Social Cognitive Theory (SCT) which suggests that people learn by observing and understanding the consequences of their actions (Bandura, 2003). In the context of language learning, AWE offers immediate feedback to students on their writing, allowing them to recognize the outcomes of their writing strategies and make the necessary adjustments (Bandura, 2001). Therefore, the findings of this study provide empirical evidence that supports SCT's belief that feedback is a fundamental aspect of the learning process. Nevertheless, it is important to note that AWE should not replace human feedback and evaluation entirely. Warschauer and Healey (1998) suggested that technology should supplement and support human instruction rather than replacing it completely. Thus, AWE should be employed in tandem with teacher feedback and instruction to provide a well-rounded approach to writing instruction in EFL environments.

Furthermore, the results indicate that pre-test scores significantly predicted task achievement, coherence and cohesion, and lexicon. Although this finding suggests that learners' initial proficiency levels exert influence on their subsequent writing performance, it is crucial to clarify the steps taken to mitigate the potential impact of participants' initial differences on the observed writing skill differences. To address this concern, the study employed a randomized controlled trial design, which ensured that participants were assigned to the experimental and control groups randomly. This random assignment aimed to distribute any potential initial skill disparities evenly across both groups. Additionally, the study employed repeated measures, allowing for within-subject comparisons over time, effectively controlling for individual differences. This design choice aimed to provide a comprehensive understanding of the intervention's impact by examining how each participant's skills evolved relative to their own baseline. The absence of a significant interaction effect between the intervention and pre-test scores for task achievement and coherence and cohesion indicates that the impact of AWE-based instruction remained consistent across varying proficiency levels in these aspects. However, the significant interaction effect for lexicon indicates that the intervention's effect on lexical improvement was weaker for participants with higher pre-test scores. This finding suggests that learners with higher initial lexical proficiency may have had less room for improvement in this specific aspect. This is consistent with previous research on the effectiveness of technology-enhanced language learning for low-proficiency learners (Huang et al., 2017; Zhang et al., 2022). These studies suggest that technology-enhanced language learning can provide more individualized and personalized instruction (Golonka et al., 2014) that caters to the specific needs of low-proficiency learners, thus leading to more effective learning outcomes.

On the other hand, writing self-efficacy was found to be a significant predictor of task achievement and coherence and cohesion. This finding highlights the importance of learners' beliefs and confidence in their writing abilities. Higher levels of writing self-efficacy may contribute to increased motivation and effort invested in writing tasks, leading to improved performance in specific writing skill components. These findings are consistent with theoretical frameworks emphasizing the role of self-efficacy beliefs in influencing learners' engagement and success in writing activities (Lee and Evans, 2019; Golparvar and Khafi, 2021; Tsao, 2021).

Moreover, this finding is consistent with the Sociocultural Theory, which suggests that learning is a social and cultural process (Vygotsky, 1978). This theory maintains that learners' language learning experiences are influenced by their social and cultural background, including their beliefs, prior knowledge, and experiences. Individuals with lower proficiency levels may have limited exposure to the target language and culture, which can restrict their language learning

opportunities. Technology-aided language learning can provide more tailored and personalized instruction, which can assist low-proficiency learners in overcoming these obstacles and promoting their language learning.

Taken together, this study indicates that AWE can be a useful tool in enhancing the writing skills of Chinese EFL learners, particularly those with lower proficiency levels. The results of this study align with the existing literature on technology-enhanced language learning and the significance of feedback in the learning process. The study's implications for EFL educators and researchers are to consider integrating AWE into their teaching and learning practices. Nonetheless, this study's limitations include using a single measure for writing skills and the need for further research on the long-term effectiveness of AWE on language learning outcomes. This study contributes to the current body of research on technology-enhanced language learning's potential to improve language skills and provides valuable insights into the use of AWE in EFL settings.

6. Conclusion and implications

In this study, the researchers probed the utility of an AWE-based instructional program on the writing skills of Chinese EFL students. The outcomes showed that the program was successful in improving L2 writing skills, with greater benefits for low-proficiency students. These findings have significant implications for second language writing instruction, suggesting that educators should incorporate AWE-based tools like Grammarly into their teaching practice. By providing instant feedback, AWE can support self-directed learning and personalized instruction, helping learners develop a comprehensive set of writing skills that includes more than just grammatical accuracy. Therefore, AWE can complement traditional writing instruction and improve learners' overall writing abilities.

In addition, this study has important implications for curriculum design and assessment. The inclusion of AWE in the curriculum can help students become more familiar with AI-based tools, which can prepare them for academic and professional contexts where such tools are commonly used. Additionally, AWE can serve as an objective assessment tool that can reliably measure student progress in writing skills, providing teachers with valuable feedback. As technology use in language education continues to expand, it is crucial to explore the potential advantages and disadvantages of various tools and approaches. The use of AWE-based instructional programs can offer a more effective and objective way of assessing and improving L2 writing skills, especially in situations where face-to-face instruction is not possible. Overall, this study demonstrates the potential of AWE-based instructional programs to improve second language writing skills and suggests that integrating AI-based tools can be a promising approach to enhance the effectiveness of second language writing instruction.

While the results of this study suggest that AWE-based instructional programs can be effective in improving second language writing skills, there are several limitations that should be considered. One limitation is that the study was conducted with a sample of Chinese EFL learners only, which limits the generalizability of the findings to other EFL contexts. To address this limitation, future research should replicate the study with different populations. Another limitation is that the study only examined short-term effects of AWE-based instruction, and it is unclear whether these effects would persist over time. Therefore, future studies should investigate the long-term effects of AWE-based instructional programs on second language writing skills. Furthermore, the study did not explore the attitudes of learners towards using AWE technology in writing instruction. Understanding students' acceptance and engagement with the technology is important to determine the effectiveness of AWE-based instruction. Therefore, future research should investigate learners' attitudes and perceptions towards AWE-based instruction. Lastly, the study did not examine the effects of AWE-based instruction on other aspects of writing, such as discourse organization and rhetorical strategies. Future studies should explore whether AWE-based instruction can improve these aspects of second language writing as well.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to PW, peiqi469@163.com.

Ethics statement

The studies involving humans were approved by the School of Foreign Languages, Tangshan Normal University, Tangshan, Hebei, China. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

All the authors equally contributed to completing this 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

Abdullah, F., Ward, R., and Ahmed, E. (2016). Investigating the influence of the most commonly used external variables of TAM on students' perceived ease of use (PEOU) and perceived usefulness (PU) of e-portfolios. *Comput. Hum. Behav.* 63, 75–90. doi: 10.1016/j.chb.2016.05.014

Al-Azawei, A., Parslow, P., and Lundqvist, K. (2017). Investigating the effect of learning styles in a blended e-learning system: an extension of the technology acceptance model (TAM). *Australas. J. Educ. Technol.* 33, 1–23. doi: 10.14742/ajet.2741

Al-Azawei, A., and Al-Azawi, R. (2021). Evaluating Facebook success in Iraq: An extension of the DeLone and McLean's model of information systems success (ISS). In *J. Phys. Conf. Ser.* IOP Publishing. 1804:012114. doi: 10.1088/1742-6596/1804/1/012114

Allan, D. (2004). *Oxford placement test*. Oxford: Oxford University Press.

Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annu. Rev. Psychol.* 52, 1–26. doi: 10.1146/annurev.psych.52.1.1

Bandura, A. (2003). "Social cognitive theory for personal and social change by enabling media" in *Entertainment—education and social change: History, research, and practice*. eds. A. Singhal, M. J. Cody, E. M. Rogers and M. Sabido (Mahwah, NJ: Lawrence Erlbaum), 75–96.

Barrot, J. S. (2022). Integrating technology into ESL/EFL writing through Grammarly. *RELC J.* 53, 764–768. doi: 10.1177/0033688220966632

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bitchener, J., and Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.

Buhi, E. R., Goodson, P., and Neilands, T. B. (2008). Out of sight, not out of mind: strategies for handling missing data. *Am. J. Health Behav.* 32, 83–92. doi: 10.5993/AJHB.32.1.8

Burstein, J., Chodorow, M., and Leacock, C. (2004). Automated essay evaluation: the criterion online writing service. *AI Mag.* 25:27. doi: 10.1609/aimag.v25i3.1774

Chen, H. J., Chiu, T. L., and Liao, P. (2009). Analyzing the grammar feedback of two automated writing evaluation systems: my access and criterion. *English Teaching & Learning*. 33, 1–43.

Cheung, Y. L. (2016). Teaching writing. In *English language teaching today*. eds. W. A. Renandya and H. P. Widodo. Springer. 179–194.

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd Edn.). Mahwah, NJ: L. Erlbaum Associates.

Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO J.* 28, 420–459. doi: 10.11139/cj.28.2.420-459

Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes? *Lang. Test.* 18, 207–224. doi: 10.1191/026553201678777086

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13:319. doi: 10.2307/249008

Dizon, G. (2016). Measuring Japanese EFL student perceptions of internet-based tests with the technology acceptance model. *TESL-Ej* 20:n2

Ebadi, S., Gholami, M., and Vakili, S. (2023). Investigating the effects of using Grammarly in EFL writing: the case of articles. *Comput. Sch.* 40, 85–105. doi: 10.1080/07380569.2022.2150067

Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York, NY: Guilford Press.

Fathi, J., Derakhshan, A., and Safdari, M. (2020). The impact of portfolio-based writing instruction on writing performance and anxiety of EFL students. *Pol. Psychol. Bull.* 51, 226–235. doi: 10.24425/ppb.2020.134729

Fathi, J., and Ebadi, S. (2020). Exploring EFL pre-service teachers' adoption of technology in a CALL program: obstacles, motivators, and maintenance. *Educ. Inf. Technol.* 25, 3897–3917. doi: 10.1007/s10639-020-10146-y

Fathi, J., and Rahimi, M. (2022). Electronic writing portfolio in a collaborative writing environment: its impact on EFL students' writing performance. *Comput. Assist. Lang. Learn.* 1-39, 1–39. doi: 10.1080/09588221.2022.2097697

Friedman, L. M., Furberg, C., and DeMets, D. L. (2010). Fundamentals of clinical trials (4th ed.).New York, NY: Springer. doi: 10.1007/978-3-319-18539-2

Fu, Q. K., Zou, D., Xie, H., and Cheng, G. (2022). A review of AWE feedback: types, learning outcomes, and implications. *Comput. Assist. Lang. Learn.* 1-43, 1–43. doi: 10.1080/09588221.2022.2033787

Geng, J., and Razali, A. B. (2022). Effectiveness of the automated writing evaluation program on improving undergraduates' writing performance. *Engl. Lang. Teach.* 15, 49–60. doi: 10.5539/elt.v15n7p49

Ghufron, M. A., and Rosyida, F. (2018). The role of Grammarly in assessing English as a foreign language (EFL) writing. *Ling. Cult.* 12, 395–403. doi: 10.21512/lc.v12i4.4582

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., and Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Comput. Assist. Lang. Learn.* 27, 70–105. doi: 10.1080/09588221.2012.700315

Golparvar, S. E., and Khafi, A. (2021). The role of L2 writing self-efficacy in integrated writing strategy use and performance. *Assess. Writ.* 47:100504. doi: 10.1016/j.asw.2020.100504

Graham, S., and Harris, K. R. (2014). Conducting high quality writing intervention research: twelve recommendations. *J. Writ. Res.* 6, 89–123. doi: 10.17239/jowr-2014.06.02.1

Grimes, D., and Warschauer, M. (2010). Utility in a fallible tool: a multi-site case study of automated writing evaluation. *J. Technol. Learn. Assess.* 8, 4–42

Han, J., and Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: the dynamics of change. *J. Second. Lang. Writ.* 40, 44–59. doi: 10.1016/j.jslw.2018.03.001

Hashemian, M., and Heidari, A. (2013). The relationship between L2 learners' motivation/attitude and success in L2 writing. *Procedia Soc. Behav. Sci.* 70, 476–489. doi: 10.1016/j.sbspro.2013.01.085

Hassanzadeh, M., and Fotoohnejad, S. (2021). Implementing an automated feedback program for a foreign language writing course: a learner-centric study: implementing an AWE tool in a L2 class. *J. Comput. Assist. Learn.* 37, 1494–1507. doi: 10.1111/jcal.12587

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *J. Educ. Behav. Stat.* 32, 341–370. doi: 10.3102/1076998606298043

Hsu, H. T., and Lin, C. C. (2022). Extending the technology acceptance model of college learners' mobile-assisted language learning by incorporating psychological constructs. *Br. J. Educ. Technol.* 53, 286–306. doi: 10.1111/bjet.13165

Huang, Y. Y., Liu, C. C., Wang, Y., Tsai, C. C., and Lin, H. M. (2017). Student engagement in long-term collaborative EFL storytelling activities: an analysis of learners with English proficiency differences. *J. Educ. Technol. Soc.* 20, 95–109.

Hyland, K. (2007). Genre pedagogy: language, literacy and L2 writing instruction. *J. Second. Lang. Writ.* 16, 148–164. doi: 10.1016/j.jslw.2007.07.005

Hyland, K. (2019). *Second language writing*. Cambridge: Cambridge university press.

Jiang, L., and Yu, S. (2022). Appropriating automated feedback in L2 writing: experiences of Chinese EFL student writers. *Comput. Assist. Lang. Learn.* 35, 1329–1353. doi: 10.1080/09588221.2020.1799824

Jiang, L., Yu, S., and Wang, C. (2020). Second language writing instructors' feedback practice in response to automated writing evaluation: a sociocultural perspective. *System* 93:102302. doi: 10.1016/j.system.2020.102302

Jingxin, G., and Razali, A. B. (2020). Tapping the potential of pigai automated writing evaluation (AWE) program to give feedback on EFL writing. *Univ. J. Educ. Res.* 8, 8334–8343. doi: 10.13189/ujer.2020.082638

Kessler, M. (2020). Technology-mediated writing: exploring incoming graduate students' L2 writing strategies with activity theory. *Comput. Compos.* 55:102542. doi: 10.1016/j.compcom.2020.102542

Khoshnevisan, B. (2019). The affordances and constraints of automatic writing evaluation (AWE) tools: a case for grammarly. *Artesol EFL J.* 2, 12–25.

Lee, Y. J. (2020). The long-term effect of automated writing evaluation feedback on writing development. *Eng. Teach.* 75, 67–92. doi: 10.15858/engtea.75.1.202003.67

Lee, M. K., and Evans, M. (2019). Investigating the operating mechanisms of the sources of L2 writing self-efficacy at the stages of giving and receiving peer feedback. *Mod. Lang. J.* 103, 831–847. doi: 10.1111/modl.12598

Lee, Y., Kozar, K. A., and Larsen, K. R. T. (2003). The technology acceptance model: past, present, and future. *Commun. Assoc. Inf. Syst.* 12, 752–780. doi: 10.17705/1CAIS.01250

Li, J., Link, S., and Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *J. Second. Lang. Writ.* 27, 1–18. doi: 10.1016/j.jslw.2014.10.004

Li, Z., Link, S., Ma, H., Yang, H., and Hegelheimer, V. (2014). The role of automated writing evaluation holistic scores in the ESL classroom. *System* 44, 66–78. doi: 10.1016/j.system.2014.02.007

Li, R., Meng, Z., Tian, M., Zhang, Z., Ni, C., and Xiao, W. (2019). Examining EFL learners' individual antecedents on the adoption of automated writing evaluation in China. *Comput. Assist. Lang. Learn.* 32, 784–804. doi: 10.1080/09588221.2018.1540433

Liao, H. C. (2016a). Using automated writing evaluation to reduce grammar errors in writing. *ELT J.* 70, 308–319. doi: 10.1093/elt/ccv058

Liao, H. C. (2016b). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System* 62, 77–92. doi: 10.1016/j.system.2016.02.007

Link, S., Mehrzad, M., and Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Comput. Assist. Lang. Learn.* 35, 605–634. doi: 10.1080/09588221.2020.1743323

Liu, G. Z., Rahimi, M., and Fathi, J. (2022). Flipping writing metacognitive strategies and writing skills in an English as a foreign language collaborative writing context: a mixed-methods study. *J. Comput. Assist. Learn.* 38, 1730–1751. doi: 10.1111/jcal.12707

Liu, O. L., Rios, J. A., Heilman, M., Gerard, L., and Linn, M. C. (2016). Validation of automated scoring of science assessments. *J. Res. Sci. Teach.* 53, 215–233. doi: 10.1002/tea.21299

Marangunić, N., and Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Univ. Access Inf. Soc.* 14, 81–95. doi: 10.1007/s10209-014-0348-1

Marzban, A., and Jalali, F. E. (2016). The interrelationship among L1 writing skills, L2 writing skills, and L2 proficiency of Iranian EFL learners at different proficiency levels. *Theory Pract. Lang. Stud.* 6:1364. doi: 10.17507/tpls.0607.05

Moon, J. W., and Kim, Y. G. (2001). Extending the TAM for a world-wide-web context. *Inf. Manag.* 38, 217–230. doi: 10.1016/S0378-7206(00)00061-6

Muthén, L. K., and Muthén, B. O. (1998–2012). *Mplus User's guide* (*7th Edn.*). Los Angeles, CA: Muthén & Muthén.

Nation, I. S., and Nation, I. S. P. (2001). *Learning vocabulary in another language* (*10*). Cambridge: Cambridge University Press.

Ngo, T. T. N., Chen, H. H. J., and Lai, K. K. W. (2022). The effectiveness of automated writing evaluation in EFL/ESL writing: a three-level meta-analysis. *Interact. Learn. Environ.* 1-18, 1–18. doi: 10.1080/10494820.2022.2096642

Nunes, A., Cordeiro, C., Limpo, T., and Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: a systematic review of studies from 2000 to 2020. *J. Comput. Assist. Learn.* 38, 599–620. doi: 10.1111/jcal.12635

Palermo, C., and Wilson, J. (2020). Implementing automated writing evaluation in different instructional contexts: a mixed-methods study. *J. Writ. Res.* 12, 63–108. doi: 10.17239/jowr-2020.12.01.04

Parmaxi, A. (2023). Virtual reality in language learning: a systematic review and implications for research and practice. *Interact. Learn. Environ.* 31, 172–184. doi: 10.1080/10494820.2020.1765392

Parra, G. L., and Calero, S. X. (2019). Automated writing evaluation tools in the improvement of the writing skill. *Int. J. Instr.* 12, 209–226. doi: 10.29333/iji.2019.12214a

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Lang. Learn.* 47, 101–143. doi: 10.1111/0023-8333.31997003

Qassemzadeh, A., and Soleimani, H. (2016). The impact of feedback provision by Grammarly software and teachers on learning passive structures by Iranian EFL learners. *Theory Pract. Lang. Stud.* 6, 1884–1894. doi: 10.17507/tpls.0609.23

Rahimi, M., and Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: a mixed methods study. *Comput. Assist. Lang. Learn.* 35, 2627–2674. doi: 10.1080/09588221.2021.1888753

Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Comput. Assist. Lang. Learn.* 31, 653–674. doi: 10.1080/09588221.2018.1428994

Ranalli, J., Link, S., and Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educ. Psychol.* 37, 8–25. doi: 10.1080/01443410.2015.1136407

Roscoe, R. D., Wilson, J., Johnson, A. C., and Mayra, C. R. (2017). Presentation, expectations, and experience: sources of student perceptions of automated writing evaluation. *Comput. Hum. Behav.* 70, 207–221. doi: 10.1016/j.chb.2016.12.076

Saricaoglu, A., and Bilki, Z. (2021). Voluntary use of automated writing evaluation by content course students. *ReCALL* 33, 265–277. doi: 10.1017/S0958344021000021

Stapleton, P., and Radia, P. (2010). Tech-era L2 writing: towards a new kind of process. *ELT J.* 64, 175–183. doi: 10.1093/elt/ccp038

Stevenson, M., and Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assess. Writ.* 19, 51–65. doi: 10.1016/j.asw.2013.11.007

Stevenson, M., and Phakiti, A. (2019). Automated feedback and second language writing. *Feed. Sec. Lang. Writ.* 125-142. doi: 10.1017/9781108635547.009

Storch, N. (2011). Collaborative writing in L2 contexts: processes, outcomes, and future directions. *Annu. Rev. Appl. Linguist.* 31, 275–288. doi: 10.1017/S0267190511000079

Tambunan, A. R. S., Andayani, W., Sari, W. S., and Lubis, F. K. (2022). Investigating EFL students' linguistic problems using Grammarly as automated writing evaluation feedback. *Indones. J. Appl. Linguist.* 12, 16–27. doi: 10.17509/ijal.v12i1.46428

Tang, J., and Rich, C. S. (2017). Automated writing evaluation in an EFL setting: lessons from China. *Jalt Call Journal* 13, 117–146. doi: 10.29140/jaltcall.v13n2.215

Tsao, J. J. (2021). Effects of EFL learners' L2 writing self-efficacy on engagement with written corrective feedback. *Asia Pac. Educ. Res.* 30, 575–584. doi: 10.1007/s40299-021-00591-9

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Comput. Compos.* 21, 217–235. doi: 10.1016/j.compcom.2004.02.003

Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT J.* 64, 314–320. doi: 10.1093/elt/ccp026

Venkatesh, V., and Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences.* 27, 451–481. doi: 10.1111/j.1540-5915.1996.tb00860.x

Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Waer, H. (2023). The effect of integrating automated writing evaluation on EFL writing apprehension and grammatical knowledge. *Innov. Lang. Learn. Teach.* 17, 47–71. doi: 10.1080/17501229.2021.1914062

Wallace, L. G., and Sheetz, S. D. (2014). The adoption of software measures: a technology acceptance model (TAM) perspective. *Inf. Manag.* 51, 249–259. doi: 10.1016/j.im.2013.12.003

Wang, Y. J., Shang, H. F., and Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Comput. Assist. Lang. Learn.* 26, 234–257. doi: 10.1080/09588221.2012.655300

Warschauer, M., and Healey, D. (1998). Computers and language learning: an overview. *Lang. Teach.* 31, 57–71. doi: 10.1017/S0261444800012970

Warschauer, M., and Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Lang. Teach. Res.* 10, 157–180. doi: 10.1191/1362168806lr190oa

Wu, B., and Chen, X. (2017). Continuance intention to use MOOCs: integrating the technology acceptance model (TAM) and task technology fit (TTF) model. *Comput. Hum. Behav.* 67, 221–232. doi: 10.1016/j.chb.2016.10.028

Xu, J., and Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: a sociocultural perspective. *Asia Pac. Educ. Res.* 31, 357–367. doi: 10.1007/s40299-021-00577-7

Yousefifard, S., and Fathi, J. (2021). Exploring the impact of blogging in English classrooms: focus on the ideal writing self of EFL learners. *Int. J. Instr.* 14, 913–932. doi: 10.29333/iji.2021.14452a

Yu, S. (2021). Feedback-giving practice for L2 writing teachers: friend or foe? *J. Second. Lang. Writ.* 52:100798. doi: 10.1016/j.jslw.2021.100798

Zhai, N., and Ma, X. (2022). Automated writing evaluation (AWE) feedback: a systematic investigation of college students' acceptance. *Comput. Assist. Lang. Learn.* 35, 2817–2842. doi: 10.1080/09588221.2021.1897019

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: student perceptions and revisions. *Assess. Writ.* 43:100439. doi: 10.1016/j.asw.2019.100439

Zhang, Z. V., and Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assess. Writ.* 36, 90–102. doi: 10.1016/j.asw.2018.02.004

Zhang, R., Zou, D., Cheng, G., and Xie, H. (2022). Implementing technology-enhanced collaborative writing in second and foreign language learning: a review of practices, technology and challenges. *Educ. Inf. Technol.* 27, 8041–8069. doi: 10.1007/s10639-022-10941-9

Beyond borders: exploring the impact of augmented reality on intercultural competence and L2 learning motivation in EFL learners

Song Liu[1]*[†], Shengbing Gao[1†] and Xiaoyan Ji[2†]

[1]School of Foreign Languages, Southeast University, Nanjing, China, [2]School of Humanities and Tourism, Jiangsu Vocational College of Business, Nantong, China

Introduction: This mixed-methods study investigates the impact of augmented reality (AR) on the development of intercultural competence and L2 (second language) learning motivation among Chinese English as a Foreign Language (EFL) learners. The research comprised forty-eight intermediate-level learners who were randomly assigned to either an experimental group, receiving AR-based language instruction, or a control group, receiving traditional instruction.

Methods: Both groups underwent pre- and post-tests to assess their intercultural competence and L2 learning motivation. The experimental group engaged with an AR application, which exposed learners to a variety of cultural scenarios, customs, and norms.

Results: The results indicate that the experimental group, exposed to AR-based instruction, demonstrated significantly higher levels of intercultural competence and L2 learning motivation in comparison to the control group.

Discussion: Qualitative data analysis further elucidated that AR-based instruction enhanced learners' engagement, motivation, and deepened their cultural understanding. This study highlights the potential of augmented reality as a powerful tool for fostering the development of intercultural competence and L2 learning motivation within the EFL context, suggesting promising opportunities for innovative pedagogical approaches in language education.

KEYWORDS

augmented reality, intercultural competence, L2 learning motivation, EFL learners, mixed methods study

1. Introduction

Technology-enhanced learning has much more gained popularity due to the growth of information technology and mobile application (Ozcelik and Acarturk, 2011; Hwang and Wu, 2014; Hwang et al., 2015; Gambo et al., 2017; Teng et al., 2018; Lei et al., 2022). Traditional second language (L2) instruction methods might create challenges for both students and learners in today's globalized world and they often fall short in providing authentic experiences that can enhance engagement and motivation (Wang and Vásquez, 2012; Fathi and Rahimi, 2022). As such, with the advancement of technology, new possibilities for language learning emerge, offering innovative and engaging approaches to address these challenges (Grgurović et al., 2013; Rahimi and Fathi, 2022). Augmented reality (AR), as a newly developing cutting-edge technology, has proven to be a beneficial tool for increasing electronic language learning and motivation (Mahadzir and Phung, 2013; Godwin-Jones, 2016; Li and Wong, 2021; Lin et al., 2022). Following Azuma (1997), AR is regarded as a three-dimensional technology that seamlessly blends both the virtual and real worlds in order to enhance

learners' understanding of the actual world with simulated items. As contextualized learning aids students in understanding and applying what they have learned to attain knowledge internalization, especially in the English as a foreign language (EFL) context (Chen and Li, 2010), the usage of AR in contextualized education is favorable for successful learning since AR combines virtual and real surroundings (Hsu, 2017; Chen et al., 2019; Reipschläger and Dachselt, 2019; Lee, 2022).

AR provides EFL students with cutting-edge audiovisual opportunities for learning and allows them to communicate with simulated information in real-world contexts which can lead to more immediate recall and improved comprehension of the learning material and motivation (Casella and Coelho, 2013; Zhao, 2018; Parmaxi and Demetriou, 2020). On the other hand, by adding technological data to the actual environment, AR presents fascinating novel possibilities that can improve the way learners see the outside world (Krüger et al., 2019; Zafar and Zachar, 2020). A number of studies have explored the role of AR in enhancing language learning performance in the English language contexts. For example, Richardson (2016) investigating the impact of AR game-based activities on EFL learners' language learning, indicated that AR gamification improved the learners' language learning performance. Sydorenko et al. (2019) also examining learners' language learning via AR, found that AR could significantly contribute to the learners' language learning, especially lexical resources.

More specifically, previous research has addressed the role of AR in enhancing English language learners' intercultural competence, their language learning motivation, and their positive perceptions toward AR language learning environments. For instance, Matveev et al. (2021) investigating the impact of using AR in developing learners' intercultural competence, indicated that the AR environment played a crucial role in improving the learners' intercultural competence. Hadjistassou et al. (2021a,b) further found that the AR environment improved the learners' intercultural competence. Kamarudin et al. (2021) investigating EFL students' perceptions of virtual language learning applying an AR application, found that the AR application promoted the learners' engagement in language learning activities. Yang and Mei (2018) also found learners' positive perceptions toward the AR language learning environment.

Although the idea of ubiquitous learning (i.e., learning anywhere and anytime) has recently become widely recognized (Hwang and Tsai, 2011), the use of AR in actual educational settings has been constrained by a dearth of research and development of educational AR environments. The review of the literature revealed that AR played a substantial role in developing English language learners' language learning (Yang and Mei, 2018; Sydorenko et al., 2019; Dalim et al., 2020; Li and Wong, 2021; Karacan and Polat, 2022), language learning engagement and motivation (Mahadzir and Phung, 2013; Richardson, 2016; Kamarudin et al., 2021; Wen, 2021), and intercultural competence (Liu et al., 2016; Hadjistassou et al., 2019; Matveev et al., 2021; Sabie et al., 2023). However, the impact of AR on learners' learning motivation and intercultural competence needs further investigation, especially in the EFL context. To address this and get more insight into the literature, the present study examined the impact of AR on EFL learners' learning motivation and intercultural competence. Qualitative data collection and analysis were further carried out exploring the learners' perceptions of the impact of AR on their intercultural competence and language learning motivation. The following research questions, therefore, were postulated to address the purposes of the study:

1. Are there any significant differences between AR and non-AR classes in developing EFL learners' intercultural competence and L2 learning motivation?
2. What are the perceptions of EFL learners toward the impact of AR on their intercultural competence and L2 learning motivation?

This study bridges this gap by examining the interplay between AR, intercultural competence, and L2 motivation—a triad that introduces unexplored dynamics and enriches the discourse on technology-driven language acquisition. In an increasingly interconnected world, the significance of intercultural competence for L2 learners is undeniable (Piątkowska, 2015; Avgousti, 2018; Liu et al., 2023). Our study recognizes the potential of technology-based approaches to amplify intercultural competence, especially for learners who face constraints in engaging with real-life cross-cultural interactions. Amidst this backdrop, our research acquires further distinction as it accentuates the importance of novel technologies like AR in enhancing intercultural competence. The unique two-fold purpose of AR—enhancing both communicative and intercultural competencies—might introduce a paradigm shift in pedagogical approaches.

In addition, central to our study's distinctiveness is the innovative use of a custom-developed AR application. This application emerges as a conduit for cultural awareness and intercultural competence, crafting a holistic learning experience. Its distinctiveness lies in its dedication to infusing cultural richness into every aspect of learning, transcending the confines of language to cultivate adeptness in navigating cross-cultural interactions.

Also, our methodological fusion emerges as a distinguishing element. By employing a mixed methods approach, we unearth a deeper understanding of AR's impact on intercultural competence and L2 learning motivation. Moreover, the demographics of our study participants—a group of undergraduate college students in China—add yet another layer of uniqueness to our research. This demographic, often characterized by intrinsic challenges in motivation towards English language communication and intercultural competence, defies initial expectations. Our findings, which highlight the engaging and motivating impact of AR on these participants, might reframe assumptions and underscore the dynamic nature of technology-enhanced language learning.

2. Literature review

2.1. Theoretical framework

The present study's interactive cultural activities in both groups are grounded in Vygotsky's (1984) social constructivism. Vygotsky's perspective highlights that interactions between more capable individuals and learners gradually internalize and trigger higher levels of awareness. He posits that functions in cultural development manifest first on a social level and subsequently on an individual level—initially between individuals (inter-psychological) and later within individuals (intra-psychological) (Vygotsky, 1986, p. 57). Central to Vygotsky's social constructivism is the concept of the Zone of Proximal Development (ZPD), wherein a gap exists between learners' independent problem-solving ability and their potential growth when collaborating with more skilled peers. Through peer engagement and assistance, learners can bridge this gap and achieve their ZPD. Researchers contend that learners can

alternate between the roles of more and less capable peers based on language learning tasks, leading to collaborative ZPD attainment across diverse skills (Kim, 2008). Sharing varying language abilities enables learners to collaborate, co-construct language skills, and achieve their ZPD.

In recent years, the rapid evolution of technology has introduced fresh possibilities for embedding Vygotsky's principles within educational contexts (Fathi and Rahimi, 2022; Shadiev and Yu, 2022). While the core tenets of social constructivism remain timeless, contemporary educational research has explored how digital tools, portable devices, games, and adaptive learning experiences can enhance the effectiveness of social interactions and collaborative learning, aligning seamlessly with Vygotsky's ideas (Lin and Lin, 2019; Liu et al., 2022). Digital platforms and technology-mediated environments offer learners unprecedented opportunities for collaborative activities, enabling them to fluidly transition between roles of knowledgeable peers and learners based on tasks (Tommerdahl et al., 2022). This versatility facilitates mutual support, collaborative problem-solving, and co-construction of knowledge, ultimately contributing to ZPD achievement.

Furthermore, modern educational technology, including Augmented Reality (AR), expands the horizons of collaborative learning (Yu et al., 2022). AR offers dynamic, immersive experiences that enable interactive engagement with cultural contexts and intercultural scenarios (Cai et al., 2022), aligning with Vygotsky's emphasis on shared experiences and interactions. The integration of AR technology in the experimental group of this study provided learners with culturally enriched content and collaborative intercultural learning experiences, underscoring the symbiosis between technology and social constructivism. Similarly, learners in the control group collaborated with peers to explore their ZPD within traditional classroom settings, showcasing the continued relevance of Vygotsky's principles across technology-enhanced and conventional learning environments.

In the evolving landscape of technology-enhanced learning, the fusion of Vygotsky's social constructivism with contemporary digital tools accentuates the importance of collaborative interactions, co-constructed knowledge, and ZPD attainment through technology-mediated means (Fathi et al., 2023; Shortt et al., 2023). By embracing this synthesis, the present study contributes to the ongoing discourse on effective pedagogical approaches that leverage technology while remaining rooted in established educational theories.

2.2. Augmented reality

AR is conceptualized as "a situation in which a real-world context is dynamically overlaid with coherent location or context-sensitive virtual information" (Klopfer and Squire, 2008, p. 205). Following Carmigniani and Furht (2011), direct or indirect time perceptions of the outside world are enhanced by computer-generated digital data in AR. It is asserted that the idea of AR may now be seen broadly and is not restricted to any one sort of technology. For instance, mobile technologies, such as smartphones and tablets, may include AR games, which are thought to be beneficial for language learning purposes. Sannikov et al. (2015), similarly, argue that learners' acquisition of skills and professional abilities is facilitated and expedited by the use of AR in mobile learning.

A substantial body of studies has examined the role of AR in second and foreign language contexts (Sydorenko et al., 2019; Parmaxi and Demetriou, 2020; Kamarudin et al., 2021; Karacan and Akoglu, 2021; Cai et al., 2022). Karacan and Polat (2022), for example, investigated the elements that influence pre-service English language teachers' expectations to include AR in their subsequent language instruction. The teachers received instruction on how to incorporate AR activities and discussions into their language sessions. The results of a questionnaire showed that the perceived utility of AR was the most important predictor of the teachers' willingness to utilize AR. The least important elements, however, were the favorable circumstances and ease of usage. Similarly, Li and Wong (2021) reviewing a number of studies about AR, indicated that a large number of studies have recently focused on the benefits of AR in educational settings. Li and Wong also revealed that most of the AR studies concentrated on the use of AR in developing learners' speaking performance and vocabulary knowledge. They further recommend that along with AR, educators need to design appropriate curricular and learning materials in order to improve learners' language proficiency.

In a similar vein, Kamarudin et al. (2021) examined how Saudi Arabian graduate and postgraduate EFL students behaved toward online learning via an AR application. They examined the social, personal, emotional, and cognitive variables that influenced the students' electronic learning at higher education institutes. The findings revealed that interpersonal, social, and emotional aspects influenced the students' behavioral goals and cognitive engagement with e-learning. Additionally, the findings showed that the moderating effect of cognitive involvement was used to explain how the students' behavioral intentions toward electronic learning were impacted by personal, social, and emotional aspects. The findings suggested that EFL students need to be encouraged to learn by the AR application. Yang and Mei (2018) also explored learners' perceptions of using AR for their language learning activities. Collecting the required data via semi-structured interviews, the findings indicated that the learners held positive perceptions of the application of AR for their language learning purposes.

Richardson (2016) examined the effects of AR-supported gamification in developing EFL learners' language learning using the Aurasma mobile application. The learners accomplished a number of challenging language learning tasks via AR. The required data were gathered through observation and feedback on the learners' language learning tasks accomplishments. The findings revealed that AR game-based activities enhanced the learners' language learning and their engagement in language learning activities. Dalim et al. (2020) also investigated the influence of AR and speech recognition technologies on English language learners' vocabulary knowledge gains and language learning engagement. An AR interface named TeachAR was developed for accomplishing the purposes of the study. Two experiments were applied to examine the effectiveness of combining AR and speech recognition in improving the learners' vocabulary knowledge. The findings indicated that the learners in the experimental group outperformed their counterparts in the control group in terms of vocabulary knowledge and engagement. The findings also demonstrated that the combination of AR and speech recognition technologies helped the learners to accomplish certain language learning tasks more rapidly.

In a similar vein, Baabdullah et al. (2022) examined 500 Saudi Arabian undergraduate EFL students' experience of using AR applications. The benefits of an AR environment on the personal, social, emotive, and cognitive dimensions of the EFL learners were examined. The findings demonstrated the strong influence of personal interactive, emotive, and cognitive benefits on the learners' AR experiences. The outcomes also showed a substantial connection between the learners' AR experience and their learning achievement. Ibáñez et al. (2020) also comparing the effectiveness of an AR learning environment and a web-based learning environment, indicated that the AR learning environment was more effective in developing the learners' learning performance than the web-based learning environment. Furthermore, the learners who used AR demonstrated a higher level of motivation in the learning activities in comparison with the learners who did not benefit from AR.

Sydorenko et al. (2019) explored English language learners' language learning via AR game-based activities. The data were gathered through video recordings of the learners' interaction with AR using their mobile devices. Analyzing the data through the technique of language-related episodes, the findings indicated that in the AR environment, the learners mostly concentrated on lexical items. Following Vygotsky's social constructivist theory of learning, the advice of more capable individuals and the immediate environment usually helped the learners understand the new lexical elements. Wen (2021) investigated English language learners' motivation and engagement in an AR-supported environment. The results showed the learners' high levels of motivation and engagement in AR-supported activities. The learners were more consistently involved with classroom activities that were made for their own settings as opposed to learning content knowledge created by experts.

Saleem et al. (2021) further explored 133 Pakistani university EFL students' perspectives on online learning using AR during the COVID-19 pandemic. For data collection and analysis purposes, the study adopted structural equation modeling. Overall, the findings demonstrated that the EFL university students held positive perceptions toward using AR during the COVID-19 pandemic. Similarly, Li and Liu (2022) investigated college students' perceptions of using AR for covering the content of their textbooks. The learners used mobile-based AR to browse the content of their textbooks more efficiently. The findings displayed that the students were strongly motivated and highly engaged in AR-supported activities, which could help them increase their learning performance. Moreover, Hsieh (2016) exploring both teachers' and students' perspectives on using AR materials via interviews, indicated that both teachers and students were willing and motivated to apply AR materials in their courses. The findings also revealed that using AR can positively influence both teachers' and learners' learning behavior.

The AR pop-up book is another option for helping students link the gap between the digital and physical worlds. By moving the book via the webcam and using detection—a two-dimensional pattern that carries information tied to the book page—students can use the AR pop-up book. Mahadzir and Phung (2013) created an AR pop-up book utilizing the ZooBurst technology and added a problem-solving strategy to aid with the motivation and language learning of English language students. The data were gathered through semi-structured interviews and observations of learners employing the AR pop-up book. The findings of their study showed that the learners' motivation and language learning had improved.

Some studies have also focused on the role of AR in enhancing learners' intercultural competence (Liu et al., 2016; Miranda Bojórquez et al., 2016; Hadjistassou et al., 2021a,b). For instance, Liu et al. (2016) examined the role of AR in improving English as a second language learners' cultural awareness, communication skills, and language improvements. Gathering the data through video recordings, the findings revealed that the learners were highly engaged in both virtual and physical worlds and could successfully accomplish the required language learning tasks, which subsequently increased the learners' cultural awareness, communication skills, and language learning. Matveev et al. (2021) examined the utilization of AR in forming learners' multicultural competence. The Zome application was applied to engage the learners in an AR environment in order to enhance their multicultural competence which included the development of intercultural interaction and a new worldview. The findings revealed the positive role of AR in enhancing the formation of multicultural competence of the learners.

Sabie et al. (2023) investigated the role of AR in fostering intercultural exchanges among learners with different cultural backgrounds. An augmented reality application was utilized to link individuals with different cultures and help them interact with each other, so as to enhance their intercultural competence. The results indicated that the users were highly engaged in the AR environment because of the application's narrative, visuals, and interactive features. Additionally, the users were more enthusiastic about exploring their own and other users' cultural elements which boosted their confidence in connecting with individuals from other cultures. Hadjistassou et al. (2021a,b) explored how AR enhanced the intercultural competence of learners from two different academic institutions in Sweden and Cyprus. The findings demonstrated that the AR-based application provided game-based activities for the learners through which they were able to improve their intercultural competence. Similarly, Hadjistassou et al. (2019) examined the impact of AR on the intercultural competence of individuals with various cultural backgrounds. The learners were from two academic institutions in the United Kingdom and Cyprus. The findings indicated that AR could significantly contribute to both groups of learners' intercultural competence.

2.3. AR-based and game-based learning technologies

Building upon the existing body of knowledge, a thorough exploration of the similarities and differences between AR-based learning and game-based learning technologies unveils interactions, shared traits, and distinct features that significantly shape the learning experience. AR dynamically merges real-world contexts with virtual information, resulting in a dynamic blend that enhances perception through digital data (Chen et al., 2019). On the flip side, game-based learning capitalizes on the intrinsic appeal of gaming to drive educational efforts (Kessler et al., 2022).

In the ever-evolving realm of mobile technologies, the convergence of AR and gamification takes center stage, exemplified by interactive platforms such as smartphones and tablets (Pellas et al., 2019). Within this dynamic landscape, AR-enhanced games thrive, offering promising avenues for nurturing language acquisition and cultural engagement (Chen, 2019). This cooperative partnership

harnesses the inherent engagement of gaming to seamlessly intertwine interactive entertainment with language learning (Yu et al., 2022).

Comparatively, game-based learning, rooted in gamification principles, taps into the natural motivational allure of games to facilitate educational advancement. This approach adeptly incorporates educational content into the fabric of game narratives, fostering engagement, participation, and skill development (Hung et al., 2018). This synergy finds resonance in AR-enhanced language learning games, where real-world contexts intermingle with linguistic challenges, cultural experiences, and interactive scenarios (Cai et al., 2022). These fusion nurtures a comprehensive language proficiency and intercultural awareness.

Both AR-based learning and game-based learning harness interactivity and immersive experiences to cultivate learners' motivation and dedication to learning (Chen et al., 2018). Rooted in the inherent joy and reward systems of gaming, both modalities fortify commitment and enthusiasm. Furthermore, both AR and game-based approaches facilitate experiential learning by immersing learners in genuine contexts, promoting practical language usage and cultural insights (Pellas et al., 2019). This immersion effectively bridges theoretical knowledge with real-world application. Additionally, interaction takes center stage in both approaches; AR-based learning involves dynamic engagement with digital overlays within real-world settings, while game-based learning emphasizes interactive challenges, narratives, and problem-solving tasks (Yu et al., 2022).

Nevertheless, AR surpasses its counterpart by enriching real-world contexts with virtual information, allowing for a deeper understanding of cultural intricacies and language application within genuine scenarios (Lin and Wang, 2023). Conversely, game-based learning frequently employs fictional scenarios to facilitate learning, providing learners with imaginative landscapes for language exploration. Narratives often drive game-based learning, embedding language and cultural learning within stories (Kessler et al., 2022). In contrast, AR-based learning immerses learners in authentic environments, bridging language skills with tangible real-world situations (Wen, 2021). Finally, AR seamlessly melds digital components with the physical world, facilitating instant connections between virtual tasks and the physical environment (Yang and Mei, 2018). Conversely, game-based learning flourishes within imaginative domains, imparting language learning within abstract contexts (Xu et al., 2020). This distinction underscores the varied ways in which these two approaches elevate the learning journey.

2.4. Intercultural competence in language learning

In the L2 domain, the learning process extends far beyond the mere acquisition of grammatical rules and vocabulary. In the increasingly interconnected world nowadays, achieving fluency in a foreign language is intimately intertwined with the ability to comprehend and navigate diverse cultural landscapes (O'Dowd, 2006; Wang, 2023). Beyond linguistic proficiency, the intricate interplay between language and culture transcends national boundaries, manifesting in the daily behaviors, interactions, and communication patterns of individuals (Kramsch, 2013; Byram and Golubeva, 2020). This interconnection finds its essence in the concept of intercultural competence—an essential capacity that enables effective and culturally

apt engagement with individuals from a variety of cultural backgrounds (Byram, 1997, 2020).

Intercultural competence goes beyond the domain of linguistic prowess, delving into the realms of cultural sensitivity, awareness, and adaptability (Byram et al., 2013). Within the L2 context, it occupies a pivotal position in facilitating successful communication and cultivating profound interactions (O'Dowd, 2006; Wang, 2023). As the global landscape continues to evolve, individuals find themselves frequently engaged in interactions with people from diverse cultures—for academic, professional, or personal reasons. By nurturing intercultural competence, L2 learners acquire the skills to interpret and appreciate cultural subtleties, thereby enhancing the quality and effectiveness of their communicative endeavors (Tecedor and Vasseur, 2020).

The profound interrelationship between culture and language acquisition has been eloquently emphasized by Risager (2013), who asserts that language and culture are inherently intertwined. Language serves not only as a conduit for communication but also as a channel for conveying deeper meanings and inherited ways of life (Avgousti, 2018). An individual's cultural background profoundly shapes their perceptions, interactions, and responses to diverse situations, with language serving as the medium through which these cultural inclinations find expression and mutual understanding (Avgousti, 2018; Byram, 2020).

Critical to effective intercultural communication, intercultural competence transcends mere linguistic mastery. Koester and Lustig (2010) stress that the L2 acquisition goes beyond the intricacies of language structure; it requires a nuanced understanding of the cultural dimensions that underpin communication. Intercultural competence empowers individuals to navigate cultural differences, discern subtle nuances, and communicate effectively across cultural boundaries. Byram (1997) introduces the concept of the "intercultural speaker," an individual skilled in cross-cultural engagement. This competence encompasses dispositions, knowledge, and skills cultivated through exposure, reflection, and guided exploration (Byram et al., 2013; Byram and Golubeva, 2020). It involves an ongoing interplay between one's own culture and the foreign culture, fostering a deeper appreciation for both. The intercultural speaker is an active participant, not a passive observer, who critically evaluates and adeptly adapts to varying cultural contexts (Byram, 2020).

To amplify intercultural competence, L2 education must transcend its traditional focus solely on linguistic proficiency. Byon (2007) underscores the significance of integrating cultural awareness into language teaching. Learners engage in interactive activities that simulate real-life cultural interactions, fostering enhanced cultural behaviors and attitudes (Shadiev and Yu, 2022). Educators play a pivotal role in guiding students through these immersive experiences (Liu et al., 2023).

In the assessment of intercultural competence, Portalla and Chen (2010) utilized the Intercultural Communication Effectiveness Measure, a comprehensive multidimensional construct. This framework delineates six distinct subscales—behavioral flexibility, interaction relaxation, interactant respect, message skills, identity maintenance, and interaction management—each encapsulating diverse facets of intercultural competence. Behavioral flexibility, a pivotal aspect, delves into participants' adaptability within a panorama of intercultural interactions, showcasing their prowess in navigating diverse cultural contexts (Chen and Starosta, 2000). Interaction

relaxation, another crucial element, is concerned with the emotional dimension of intercultural competence, exploring participants' comfort levels during cross-cultural interactions. This offers insights into the emotional resonance experienced when engaging with individuals from diverse cultural backgrounds—an essential aspect of effective intercultural communication.

Interactant respect, a fundamental metric, assesses participants' embrace of respect and open-mindedness when interacting with individuals from diverse cultural backgrounds. It deals with the demeanor and approach individuals adopt when engaging with others, underscoring the significance of mutual respect in fostering intercultural harmony (Chen and Starosta, 1997). Message skills, a dimension under scrutiny, examines participants' ability to convey ideas and comprehend messages effectively within intercultural contexts (Chen, 2007).

Identity maintenance, another facet, investigates participants' ability to uphold their cultural identity while engaging with cultures different from their own (Portalla and Chen, 2010). This dimension highlights the delicate balance between adaptation and preservation, revealing the extent to which individuals remain anchored to their cultural roots within the realm of cultural diversity (Wood, 2008). Lastly, interaction management, the final dimension, explores the art of effectively navigating intercultural interactions. This encompasses the spectrum from conflict resolution to collaborative prowess, offering insights into how individuals orchestrate effective communication and collaboration across cultural boundaries (Koester and Olebe, 1988).

Collectively, these dimensions, as elucidated by Portalla and Chen (2010), weave an intricate tapestry of intercultural competence. In their diversity, they mirror the multifaceted nature of effective intercultural communication and underscore the pivotal role such competence plays in fostering harmony, understanding, and successful collaboration within a diverse global landscape. In essence, intercultural communicative competence (ICC) encompasses a range of dimensions—intercultural communicative awareness, sensitivity, and effectiveness (ICE). ICE, closely aligned with the behavioral facet of ICC, revolves around the verbal and non-verbal communication behaviors crucial for intercultural collaboration (Chen and Starosta, 2007). This construct encapsulates the holistic essence of intercultural competence, anchoring it as a vital cornerstone in the realm of language learning.

3. Method

3.1. Participants

The current mixed methods study included 48 undergraduate students who were enrolled in a mandatory College English Band III course at a comprehensive interdisciplinary university situated in a suburban region of mainland China. The primary aim of the course was to enhance their proficiency in the English language. These individuals were categorized as intermediate level based on their English language scores derived from the annual national college entrance examinations in China. With an average age of 19.68 years (SD = 2.31), the participants were selected from two separate classes taught by an instructor who maintained independence from the researcher. In terms of prior experience

with AR technologies, approximately 65% of the participants reported having some familiarity with AR applications, primarily from casual usage and exposure to mobile applications or games. Employing a quasi-experimental design, the participants were divided into either the experimental group ($n = 25$) or the control group ($n = 23$).

To ensure comparability between the groups, an independent-sample t-test was conducted employing scores from a recent college English language proficiency quiz completed by the participants. The findings revealed that students from both classes demonstrated a comparable level of proficiency in the English language, $t(46) = -2.35$, $p = 0.152$. Among the participants in the experimental group, eight students voluntarily agreed to partake in semi-structured interviews subsequent to the intervention. The selection of these participants was based on their willingness to engage in the interview process and their diverse backgrounds, including both prior AR experience and newcomers to the technology. It is noteworthy that student participation was entirely voluntary, and throughout the study, stringent measures were implemented to preserve the anonymity and privacy of the participants, adhering to ethical considerations and ensuring confidentiality.

3.2. Instruments

3.2.1. Intercultural communicative competence

To assess participants' intercultural competence in both groups, the Intercultural Communication Effectiveness Measure developed by Portalla and Chen (2010) was employed as a robust and established instrument. This measure comprised 20 items, each rated on a 5-point Likert scale, spanning from "strongly disagree" to "strongly agree." The questionnaire encompassed six distinct subscales, namely behavioral flexibility, interaction relaxation, interactant respect, message skills, identity maintenance, and interaction management, which collectively capture various dimensions of intercultural competence.

The internal consistency of the entire scale was rigorously evaluated using the widely accepted Cronbach's Alpha formula. The pre-test demonstrated a reliability coefficient of 0.86, indicating a high level of internal consistency, while the post-test exhibited a reliability coefficient of 0.82, signifying a satisfactory level of internal consistency for the scale. These findings underscore the reliability and stability of the instrument in measuring intercultural competence among the participants.

3.2.2. Second language motivated behavior

To gauge participants' motivation in learning English as a second language, this study employed the Motivated Behavior Scale developed by Taguchi et al. (2009), a widely recognized assessment tool in the field. The scale consisted of 10 self-report items, each specifically crafted to measure distinct facets of EFL learners' motivated behavior. Participants were requested to indicate their level of agreement with each item on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

One representative item from the scale was: "I am willing to invest substantial effort into mastering English." To ensure the reliability of the scale, the internal consistency was examined using Cronbach's alpha coefficient, yielding a reported reliability index of 0.88, indicating high internal consistency and stability of the instrument.

3.2.3. Semi-structured interview

In order to gain a nuanced understanding of participants' experiences, perceptions, and attitudes towards the AR Language Learning App and its influence on intercultural competence and L2 learning motivation, semi-structured interviews were employed. This qualitative research instrument facilitated an in-depth exploration of individual viewpoints, allowing participants to elaborate on their interactions with the app.

The semi-structured interviews adhered to a flexible interview guide, ensuring consistency while also permitting spontaneous and open-ended responses (see Appendix for details). These interviews were conducted individually with a selected subset of participants from the experimental group. A purposive sampling approach was employed to deliberately select participants who actively engaged with the app, representing a diverse range of experiences and perspectives. Eight participants willingly volunteered to take part in the interviews, thereby providing valuable insights from varied vantage points.

To create a comfortable and private environment conducive to open and honest dialog, the interviews were conducted in a relaxed setting. With the participants' consent, the sessions were audio-recorded to ensure accurate data collection. The interviews lasted between 30 and 45 min, allowing ample time for participants to express their thoughts and reflect upon their experiences with the AR Language Learning App.

During the analysis phase, the audio recordings were meticulously transcribed verbatim, capturing both verbal and non-verbal cues. Thematic analysis was employed to identify prevalent patterns, recurring themes, and noteworthy excerpts related to intercultural competence and L2 learning motivation. The data derived from the semi-structured interviews served to complement and enrich the quantitative findings, fostering a more comprehensive understanding of participants' perspectives and the overall impact of the AR Language Learning App.

3.2.4. AR language learning app

In this study, we utilized a custom-developed AR Language Learning App specifically designed for the purposes of the research. The app was created by a team of experienced developers and instructional designers in collaboration with language teaching experts. It incorporated augmented reality technology to provide learners with an immersive and interactive language learning experience. This app was designed to deliver culturally rich content, interactive language exercises, quizzes, cultural challenges, and scavenger hunts to enhance language acquisition and intercultural competence. Upon launching the app, learners were greeted by an aesthetically pleasing dashboard, providing good access to a diverse range of features and functionalities. The application was thoughtfully designed to deliver a comprehensive and engaging learning experience, incorporating the following elements:

1. **Immersive Cultural Content:** The app curated an array of culturally enriched content, offering virtual tours of iconic cultural landmarks, interactive simulations of real-life scenarios, and multimedia presentations showcasing cultural traditions and practices.
2. **Interactive Language Exercises:** Learners could actively participate in a variety of interactive language exercises tailored to enhance their linguistic proficiency. These exercises

encompassed vocabulary drills, grammar quizzes, and language comprehension activities, all carefully crafted to be both engaging and pedagogically effective.

3. **Quizzes and Assessments:** Beyond language skills, the application featured quizzes and assessments that encompassed intercultural knowledge evaluation. This comprehensive approach ensured a holistic assessment of learners' progress, measuring both language competence and intercultural awareness.
4. **Cultural Challenges:** Promoting intercultural competence, the app presented learners with thought-provoking cultural challenges. These challenges encouraged learners to delve into the intricate cultural nuances of the language they were mastering, fostering a deeper understanding of intercultural communication dynamics.
5. **Scavenger Hunts:** Within the augmented reality environment, the app introduced captivating scavenger hunts, seamlessly merging language learning with real-world exploration. Learners were tasked with locating cultural artifacts and landmarks, thereby reinforcing language acquisition through practical, context-driven experiences.

3.3. Procedure

In this study, we divided the participants into two groups: the experimental group and the control group. The experimental group received the AR Language Learning App intervention, while the control group underwent traditional instruction without the use of AR technology.

Guided by the expertise of the same seasoned instructor, both groups embarked on their language learning paths with tailored approaches and activities attuned to their instructional methodologies. For the experimental group, the teacher introduced the AR Language Learning App to the participants and provided detailed instructions on how to use it. The app was designed with a user-friendly interface, allowing learners to easily navigate through its sections and activities. During the language classes, the teacher actively facilitated the learners' engagement with the app, which offered a variety of culturally immersive situations and environments to explore. These included virtual shopping experiences, simulated dining scenarios, social gatherings, and simulated business interactions. Within these contexts, learners had the opportunity to interact with virtual characters representing individuals from different cultures. The teacher encouraged learners to have conversations with these virtual characters, promoting language practice and providing insights into cultural norms, customs, and communication styles.

To enhance language learning and intercultural competence, the app included interactive language exercises, quizzes, cultural challenges, and scavenger hunts. Learners were presented with specific language tasks within the app's augmented reality environment. These tasks required them to label objects, construct sentences, respond to cultural questions, and identify cultural artifacts or landmarks in their physical surroundings. The activities were designed to actively engage learners in language acquisition and deepen their understanding of cultural diversity.

In the control group, the teacher employed established teaching techniques such as lectures, discussions, pair and group work, and exercises, similar to the experimental group. The instructional content focused on developing language skills in listening, speaking, reading, and writing, as well as grammar and vocabulary. To ensure parity with the experimental group, the teacher integrated cultural topics and contexts into the instruction. This involved discussing cultural practices, traditions, and customs during relevant language activities and exercises.

Throughout the control group sessions, the teacher facilitated interactive activities that aimed to reinforce language skills and knowledge while incorporating cultural aspects. These activities included communicative tasks, practice exercises, and drills, which were designed to improve language proficiency and provide opportunities for language practice within cultural themes. The teacher provided feedback on learners' language performance through traditional means, such as verbal feedback, written corrections, and assessments.

It is important to note that the teacher played a crucial role in implementing the distinct instructional methods for each group while ensuring that cultural topics were integrated into the control group's instruction. In the experimental group, the teacher focused on guiding learners' interaction with the AR Language Learning App and promoting engagement with its culturally rich content. In the control group, the teacher utilized established teaching techniques, integrated cultural topics into language activities, and utilized traditional instructional materials to support language acquisition.

Evidently, the AR application's content was purposefully curated to provide learners with dynamic opportunities for language acquisition and intercultural understanding. Through its immersive scenarios, interactive language exercises, and cultural challenges, the app aimed to promote active engagement, motivation, and a deeper connection with both the language and the rich tapestry of diverse cultural contexts. This comprehensive approach ensured that the content presented not only facilitated language learning but also enriched the learners' understanding of cultural intricacies, contributing to their overall motivation and intercultural competence. The instructor's role in implementing distinct methodologies for each group further ensured the continuity and comparability of the instructional strategies across both groups.

Overall, the unique content presented to both groups held the potential to significantly influence their motivation to participate in the learning activities. In the experimental group, the app's captivating scenarios and interactive tasks offered a fresh, innovative learning avenue. Engaging with virtual characters and navigating augmented reality environments could have fostered a sense of excitement and curiosity, driving students' intrinsic motivation to actively participate in language learning. The immersive experiences likely kindled a sense of adventure and discovery, potentially amplifying students' enthusiasm to engage with the content and engage in language practice.

In the control group, the integration of cultural dimensions within traditional instruction also played a pivotal role in shaping motivation. The discussions on cultural practices and traditions could have enhanced students' sense of connection to the language they were learning. By contextualizing language within cultural contexts, the students might have perceived the learning process as more meaningful and relevant, further fueling their motivation to participate.

In essence, the unique content presented to both groups had the potential to influence motivation through different pathways. The experimental group's engagement with immersive augmented reality scenarios could have triggered curiosity and excitement, while the control group's exposure to culturally woven traditional instruction could have evoked a deeper sense of purpose and connection. The interplay between content, instructional methods, and motivation forms a complex yet intriguing facet of this study's exploration into enhanced language learning experiences.

3.4. Data analysis

A combination of quantitative and qualitative analysis methods was employed to evaluate the impact of AR technology on intercultural competence and L2 learning motivation. The quantitative data were summarized using descriptive statistics, offering insights into participants' intercultural competence levels and L2 learning motivation. Paired samples t-tests were conducted to examine changes within each group before and after the intervention, providing an assessment of the effectiveness of the AR Language Learning App and traditional instruction. To compare intercultural competence and L2 learning motivation between the experimental and control groups, a one-way ANCOVA was performed, taking into account pre-test scores as covariates. The qualitative data obtained from semi-structured interviews were analyzed using content analysis, revealing valuable information about participants' experiences and attitudes. By employing this comprehensive approach, a robust evaluation of the impact of AR technology on intercultural competence and L2 learning motivation was achieved.

4. Results

4.1. Quantitative results

Table 1 presents the descriptive statistics of the variables included in the study, namely the intercultural communicative competence (ICC) and L2 motivation. As seen in Table 1, the pre-test ICC mean score for the experimental group ($M = 2.88$, SD $= 0.49$) was slightly higher than the control group ($M = 2.70$, SD $= 0.57$). In the post-test, the experimental group showed an increase in the mean score

TABLE 1 Descriptive statistics.

	Group	N	Mean	Std. deviation	Std. error mean
Pre. ICC	Experimental	25	2.8864	0.49866	0.09973
	Control	23	2.7087	0.57754	0.12043
Post. ICC	Experimental	25	3.7620	0.85138	0.17028
	Control	23	3.1500	0.69364	0.14463
Pre. Motivation	Experimental	25	2.9807	0.35214	0.07043
	Control	23	3.0477	0.40009	0.08343
Post. Motivation	Experimental	25	3.6901	0.43032	0.08606
	Control	23	3.3010	0.62620	0.13057

($M = 3.76$, SD $= 0.85$) compared to the pre-test. Similarly, the control group also exhibited an increase in the mean score ($M = 3.15$, SD $= 0.69$). Regarding L2 motivation, the experimental group had a mean score of 2.98 (SD $= 0.35$) on the pre-test, while the control group had a slightly higher mean score of 3.04 (SD $= 0.40$). For motivation post-test scores, the experimental group had a higher mean score of 3.69 (SD $= 0.43$) than the control group with a mean score of 3.30 (SD $= 0.62$).

Table 2 presents the results of the paired samples t-test conducted to assess the differences in intercultural competence and motivation between the experimental and control groups before and after the intervention.

For the experimental group, the paired samples t-test revealed a significant increase in intercultural competence from pre-intervention ($M = 2.88$, SD $= 0.49$) to post-intervention ($M = 3.76$, SD $= 0.85$), $t(24) = -5.29$, $p < 0.001$. This change was associated with a moderate effect size, Cohen's d $= 0.46$. Similarly, the experimental group demonstrated a significant increase in motivation scores from pre-intervention ($M = 2.98$, SD $= 0.35$) to post-intervention ($M = 3.69$, SD $= 0.43$), $t(24) = -39.44$, $p < 0.001$, with a large effect size, Cohen's d $= 0.64$.

In the control group, the paired samples t-test indicated a significant increase in intercultural competence from pre-intervention ($M = 2.70$, SD $= 0.57$) to post-intervention ($M = 3.15$, SD $= 0.69$), $t(22) = -3.13$, $p = 0.005$. This change was associated with a moderate effect size, Cohen's d $= 0.25$. The control group also exhibited a significant increase in motivation scores from pre-intervention ($M = 3.04$, SD $= 0.40$) to post-intervention ($M = 3.30$, SD $= 0.62$), $t(22) = -2.27$, $p = 0.033$, with a small to moderate effect size, Cohen's d $= 0.19$.

These findings suggest that both the experimental and control groups experienced increases in intercultural competence and motivation following the intervention. However, the experimental group demonstrated larger mean differences in both intercultural competence and motivation compared to the control group. To explore the between-group differences, ANCOVA was used.

Table 3 displays the results of the ANCOVA run to examine the effects of the AR-based instruction on ICC. The results revealed a statistically significant impact of Group variable on the ICC scores [$F(1, 45) = 5.783$, $p = 0.020$, $\eta p^2 = 0.114$]. This ηp^2 value suggests that approximately 11.4% of the variance in ICC scores can be attributed to the Group variable (i.e., independent variable), reflecting a moderate effect size. This finding underscores a notable difference in ICC scores between the experimental and control groups after accounting for pre-test ICC scores. Specifically, the experimental group demonstrated a more substantial increase in ICC compared to the control group.

Similarly, Table 4 presents the results of the ANCOVA conducted to assess the effects of the AR on the L2 motivation scores. The analysis showed a significant effect on the L2 motivation scores [$F(1, 45) = 17.284$, $p < 0.001$, $\eta p^2 = 0.278$]. The ηp^2 value, approximately 27.8%, signifies a substantial effect size. This outcome highlights a significant difference in L2 motivation scores between the experimental and control groups after controlling for pre-test L2 motivation scores. The experimental group exhibited significantly higher L2 motivation scores compared to the control group, indicative of the strong positive impact of integrating AR technology in language learning environments.

Overall, the results from the ANCOVA analyses support the hypothesis that the AR-based instruction had a more positive impact on both intercultural competence and L2 motivation compared to traditional instruction. The experimental group showed higher ICC and L2 motivation scores compared to the control group, indicating the potential benefits of incorporating AR technology in language learning settings.

4.2. Qualitative results

Additionally, the qualitative analysis focused on extracting themes and insights from the semi-structured interviews conducted with a subset of participants from the experimental group. Thematic analysis was used to identify patterns and commonalities in participants' experiences, engagement, motivation, and cultural understanding facilitated by the AR Language Learning App. As revealed by the thematic analysis, participants consistently expressed a heightened level of engagement and motivation while using the AR Language Learning App. They conveyed enthusiasm about the interactive and immersive nature of the app, which made language learning more enjoyable and captivating. One participant, for instance, articulated, "I felt genuinely motivated to learn when I could see and interact with the virtual characters and environments. It made the whole process more exciting and fueled my curiosity."

Also, the AR-based instruction effectively facilitated the development of cultural understanding among participants. Engaging in virtual conversations with the culturally diverse virtual characters provided learners with valuable insights into different cultural norms, customs, and communication styles. Participants acknowledged that this exposure allowed them to develop a deeper appreciation and understanding of various cultures. One participant shared, "Interacting with the virtual characters gave me a new perspective on cultural diversity. It taught me how to adapt my language and behavior in different cultural contexts, which significantly improved my intercultural competence."

TABLE 2 Paired samples test for intercultural competence and motivation.

Group	Pair	M	SD	t	df	Sig. (2-tailed)	Cohen's d
Experimental	Pre.ICC - Post.ICC	−0.87560	0.82740	−5.291	24	0.000	0.46
Experimental	Pre. Motivation - Post. Motivation	−0.70943	0.08992	−39.449	24	0.000	0.64
Control	Pre.ICC - Post.ICC	−0.44130	0.67498	−3.136	22	0.005	0.25
Control	Pre. Motivation - Post. Motivation	−0.25333	0.53337	−2.278	22	0.033	0.19

TABLE 3 ANCOVA results for the ICC scores.

Source	Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
Pre.ICC	4.134	1	4.134	7.802	0.008	0.148
Group	3.065	1	3.065	5.783	0.020	0.114
Error	23.847	45	0.530			

TABLE 4 ANCOVA results for the L2 motivation scores.

Source	Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
Motivation1	6.619	1	6.619	46.163	0.000	0.506
Group	2.478	1	2.478	17.284	0.000	0.278
Error	6.452	45	0.143			

Some participants reported that the AR app enhanced their ability to apply language skills and cultural knowledge in real-world contexts. The incorporation of cultural challenges and scavenger hunts encouraged participants to explore their physical environment and identify cultural artifacts or landmarks. This aspect of the app helped participants bridge the gap between virtual experiences and the real world. As one participant mentioned, "The scavenger hunts made me more attentive to cultural elements in my surroundings. I started noticing nuances in everyday life that I had never paid attention to before. It made me feel more connected to the language and culture." Furthermore, it was found the AR-based instruction provided participants with a personalized and adaptive learning experience. The app's progress tracking system allowed learners to monitor their performance and track their learning progress. Students appreciated the instant feedback provided by the app, enabling them to identify areas for improvement and tailor their learning strategies accordingly. A participant, for example, emphasized, "The app's feedback system helped me recognize my strengths and weaknesses in language learning. It motivated me to work on my weaker areas and celebrate my progress along the way."

These emerged themes and sample excerpts from the semi-structured interviews highlight the positive impact of AR technology on EFL participants' engagement, motivation, and cultural understanding. The qualitative findings align with the quantitative results, further reinforcing the idea that integrating augmented reality technology in language instruction significantly contributes to the development of intercultural competence and L2 learning motivation in EFL learners.

5. Discussion

This study examined the effects of using AR in enhancing EFL learners' intercultural competence and language learning motivation by adopting a mixed-methods approach for data collection and analysis. Initially, the quantitative findings indicated that the AR-mediated class significantly improved the learners' intercultural competence and outperformed its non-AR counterpart in that regard.

The findings in this regard are in agreement with the findings of Hadjistassou et al. (2019), Matveev et al. (2021), and Sabie et al. (2023) who corroborated the positive effects of AR on learners' intercultural competence. This augmentation in competence can be attributed, at least in part, to the novel nature of the AR environment. The immersive experience in the AR class motivated learners to engage in more vibrant communication with their virtual peers, fostering an environment conducive to improving intercultural competence. In essence, the virtual exposure within the AR class encouraged students to interact within a new realm, potentially yielding beneficial outcomes for their intercultural competence.

In line with Vygotsky's (1984) social constructivism, the virtual learners in the AR class served as mediators for the other learners' intercultural competence. In the beginning, the learners engaged in communicative intercultural activities with their virtual partners, which, as suggested by Vygotsky (1984), might have helped them in controlling both their own and their peers' intercultural competence. This control entails navigating the intricacies of cross-cultural communication, effectively adapting and responding within diverse contexts. Moreover, the developmental process of learners unfolded in stages, transitioning from external regulation to self-regulation over their intercultural competence. This transition encompassed a shift from relying on external prompts, such as communicative activities facilitated by virtual peers, to independently engaging in intercultural endeavors. Those who mastered self-regulation displayed an advanced ability to function autonomously and successfully execute intercultural tasks without external aid. Ultimately, those who exhibited adept self-control demonstrated enhanced autonomy and efficacy in navigating and completing intercultural activities, fostering a sense of self-reliance.

The findings also underscored a significant disparity in the performance of AR-enabled learners compared to their non-AR counterparts. AR learners showcased a higher level of self-regulation in managing their intercultural competence, a result that suggests the role of AR in cultivating an environment conducive to developing learners' self-regulatory skills. These skills empower learners to actively engage with intercultural content, adapt their responses, and independently undertake intercultural activities, collectively enhancing their overall intercultural competence.

According to Liu et al.'s (2016) findings, the results mentioned above may possibly be a result of the cultural understanding that the students received from their communicative intercultural encounters with their virtual peers. Due to the fact that the students and their virtual classmates had various cultural backgrounds and capabilities, they were able to exchange cultural information, which the students found to be interesting. The AR learners' interest in learning novel cultural concepts in a virtual setting might have potentially enhanced their intercultural competence. Another factor contributing to the AR learners' superior intercultural competency may be their successful internalization of cultural aspects, which, as mentioned above, was the direct result of the learners' intercultural activities in the AR environment.

From a broader perspective, these findings align with contemporary educational research, which has explored the augmentation of social interactions and collaborative learning through digital tools, portable devices, games, and adaptive learning experiences (Lin and Lin, 2019). This perspective underscores the unprecedented opportunities for

learners to participate in collaborative activities, fluidly transitioning between roles as knowledgeable peers and learners depending on the task (Tommerdahl et al., 2022). This fluidity fosters mutual support, collaborative problem-solving, and knowledge co-construction, ultimately contributing to the attainment of their ZPD. Augmented Reality, as a modern technology, further amplifies the horizons of collaborative learning (Yu et al., 2022) by providing dynamic and immersive experiences that facilitate interactive engagement with cultural contexts and intercultural scenarios (Cai et al., 2022).

The quantitative findings of this study also shed light on the enhancement of learners' language learning motivation through the AR-supported class, surpassing its non-AR counterpart in this aspect. These results align with the research conducted by Dalim et al. (2020), Kamarudin et al. (2021), Mahadzir and Phung (2013), Richardson (2016), and Wen (2021), collectively demonstrating that AR-based classes significantly contribute to learners' engagement and motivation in language learning. The AR learners, immersed in communicative intercultural activities within an augmented environment, encountered novel experiences that contributed to the enhancement of their intercultural competence, a factor they found motivating. Conversely, the AR learners were able to leverage various features of the augmented environment to elevate their motivation across different intercultural activities, thereby enhancing both their intercultural competence and motivation.

AR amalgamates advanced technology and captivating educational content, prompting a pivotal query: What serves as the true motivator for learners—is it AR's innovative technology or its engaging content? Unraveling these influences is pivotal to comprehending the multifaceted impact of AR on learning. The allure of AR technology stirs curiosity and interest among learners (Yu et al., 2022). This innovative approach broadens the horizons of learning, nurturing active involvement that propels curiosity-driven exploration (Karacan and Akoglu, 2021). The captivating environment crafted by AR spurs learners to participate actively, a phenomenon substantiated by research demonstrating the motivating potential of emerging technologies (Chen, 2019; Ibáñez et al., 2020; Cai et al., 2022).

Delving into the educational content of AR, a crucial insight emerges: thoughtfully curated content lies at the heart of its impact. The AR application functions as a versatile platform encompassing a spectrum of intercultural encounters, language exercises, interactive scenarios, and tailored challenges. Prior research underscores this concept, underscoring the pivotal role of engaging content in fostering motivation (Li and Liu, 2022; Lin and Wang, 2023). However, at the heart of this issue lies the dynamic interplay between the innovative technology of AR and its educational content. As learners' curiosity develops, they come to realize that their motivation is not solely ignited by novelty; rather, it is nurtured and enriched by the substantial and valuable educational material that AR offers. This symbiotic interaction creates a connection where engaging content enhances the allure of AR, resulting in an uninterrupted source of motivation. Research strongly emphasizes the critical significance of integrating technology with compelling content to maintain long-lasting enthusiasm (Pellas et al., 2019).

This discussion extends beyond the conventional juxtaposition of technology against content, advocating for a holistic viewpoint. While new technology sparks initial interest, absorbing content is the propelling force that nurtures and intensifies motivation over time.

This harmonious synthesis positions AR as a realm where technology and content converge to shape learners' objectives. This perspective encourages further investigation, prompting a more profound exploration of the instructional design implications.

Overall, within the domain of AR's impact, the interplay between technology and content emerges as the pivotal catalyst propelling learners' motivation. This fusion results in an enriched learning journey that resonates with learners, enriches intercultural competence, and fuels motivation for language acquisition. This thorough exploration contributes to the ongoing discourse on technology-infused learning, fostering a deeper understanding of the diverse dimensions that shape education in the digital age. Ultimately, the integration of AR seamlessly melds technology and content, propelling learners toward gratifying learning experiences.

The qualitative findings reveal the positive impact of the AR-based instruction on engagement, motivation, and cultural understanding. Thematic analysis of interviews uncovered key insights into participants' experiences. They consistently reported higher engagement and motivation, finding the interactive and immersive nature of the app enjoyable. This suggests that augmented reality technology effectively enhances motivation and active participation in language learning (Erbas and Demirer, 2019; Ibáñez et al., 2020).

Moreover, the AR-based instruction facilitated cultural understanding by allowing learners to engage in virtual conversations with diverse characters. This exposure provided valuable insights into cultural norms, customs, and communication styles, fostering appreciation and improving intercultural competence. The app also encouraged the application of language skills and cultural knowledge in real-world contexts through cultural challenges and scavenger hunts. By bridging the gap between virtual and real experiences, language learning became more meaningful and relevant (Liu et al., 2023; Parmaxi, 2023). Participants highly valued the personalized and adaptive learning experience offered by the app. The progress tracking system and instant feedback empowered learners to monitor their performance, identify areas for improvement, and tailor their strategies. This individualized approach enhanced self-awareness and motivated learners (Xie et al., 2019).

The qualitative findings complement the quantitative results, offering a comprehensive understanding of AR technology's effects on intercultural competence and L2 learning motivation. Integrating augmented reality technology in language instruction shows promise in creating engaging, immersive experiences, fostering cultural understanding, and promoting learner autonomy. It is recommended that establishing an AR environment can involve students in a pleasant digital setting where they can communicate with their peers with greater efficiency, which can effectively boost their intercultural competence. The current study found that the EFL students were actively participating in user-friendly intercultural activities with their peers, which could effectively contribute to their intercultural competence and motivation. These communicative intercultural activities are believed to help students improve their capacity for self-regulating their intercultural competence. Vygotsky (1984) asserted that students with different skills and capabilities can help other students achieve their highest degree of performance. By fostering positive perceptions of the learners' cultures and abilities, the AR class helped the learners in the current study engage in communicative intercultural tasks more effectively. Due to the diversity of cultural backgrounds among the learners who took part in

the present research, the learners could greatly increase their peers' intercultural competence and motivation.

6. Conclusion and implication

The goal of the current study was to examine an AR language learning environment that was created to assist EFL students in developing their intercultural competence and learning motivation. In line with Vygotsky's social constructivism, the results showed that the AR class was more effective than the non-AR class in enhancing the intercultural competence and learning motivation of EFL learners. The findings were due to the innovative and engaging environment which was augmented by some virtual features. The learners also confirmed that they held positive perceptions toward language learning and intercultural activities experienced in the AR environment.

The findings pointed to a number of fruitful educational implications, especially for the EFL context. The AR class is suggested to be applied in interactive EFL intercultural courses since it is in line with modern ideas of student-centeredness and substantially enhances EFL students' intercultural competence and language learning motivation. In order to improve the intercultural competence and motivation of EFL students, EFL educators, teachers, and students are encouraged to utilize an AR environment for their communicative intercultural activities. To provide EFL students with engaging communicative intercultural activities with other students, EFL educators and teachers might set up an AR class. This makes it possible for EFL students to participate in more communicative intercultural tasks that can more successfully increase their intercultural competence and language learning motivation. By taking an AR class made particularly for communicative intercultural activities, EFL students could benefit more from peer intercultural mediation. They can receive peer and AR mediation on intercultural issues and enhance their intercultural competence accordingly. Additionally, the more engaging and interesting communicative intercultural activities in the AR environment can boost EFL students' intercultural competence and enthusiasm for language learning.

Although this study contributes valuable insights into the impact of AR on intercultural competence and L2 learning motivation among Chinese EFL learners, several limitations should be acknowledged. First, the participant pool consisted of undergraduate college students enrolled in a mandatory English course in China. This specific demographic may not fully represent the diversity of English language learners across various age groups and proficiency levels. The findings may be more applicable to similar contexts and may not be generalizable to other EFL learner populations, such as adult learners or those in different educational settings.

Second, the study employed a mixed methods approach, which allowed for a comprehensive exploration of the research questions. However, the scope of the qualitative analysis was limited to participants' perceptions and experiences. A more in-depth qualitative exploration, potentially incorporating interviews and focus group discussions, could provide richer insights into the findings. Third, the duration of the study was relatively short, encompassing a single semester. Longer-term investigations could offer a deeper understanding of the sustainability of the observed effects over extended periods of time. Additionally, the study did not explore the potential influence of prior experience with AR technology on participants' responses to AR-based language instruction. Future research could investigate whether prior familiarity with AR impacts learners' engagement and motivation.

Fourth, while the custom-developed AR Language Learning App enriched the learning experience through its cultural content, the effectiveness of individual app features and modules was not isolated in this study. Further research could delve into the specific impact of different components of the app on learners' intercultural competence and motivation. Finally, this study focused on the immediate impact of AR on intercultural competence and L2 learning motivation. Long-term effects and transfer of skills to real-life intercultural interactions were not within the scope of this investigation. A longitudinal study that traces participants' experiences and behaviors beyond the instructional period could offer deeper insights into the sustained impact of AR-enhanced learning.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to SL: dndxjyjy@163.com.

Ethics statement

The studies involving humans were approved by School of Foreign Languages, Southeast University, Jiangning District 211189, Nanjing, China. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

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

Avgousti, M. I. (2018). Intercultural communicative competence and online exchanges: a systematic review. *Comput. Assist. Lang. Learn.* 31, 819–853. doi: 10.1080/09588221.2018.1455713

Azuma, R. T. (1997). A survey of augmented reality. *Presence Teleop. Virt.* 6, 355–385. doi: 10.1162/pres.1997.6.4.355

Baabdullah, A. M., Alsulaimani, A. A., Allamnakhrah, A., Alalwan, A. A., Dwivedi, Y. K., and Rana, N. P. (2022). Usage of augmented reality (AR) and development of e-learning outcomes: an empirical evaluation of students' e-learning experience. *Comput. Educ.* 177:104383. doi: 10.1016/j.compedu.2021.104383

Byon, A. S. (2007). The use of culture portfolio project in a Korean culture classroom: Evaluating stereotypes and enhancing cross-cultural awareness. *Lang. Cult. Curriculum* 20, 1–19. doi: 10.2167/lcc323.0

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.

Byram, M. (2020). *Teaching and assessing intercultural communicative competence: revisited*. Bristol, Blue Ridge Summit: Multilingual Matters.

Byram, M., and Golubeva, I. (2020). "Conceptualising intercultural (communicative) competence and intercultural citizenship" in *The Routledge handbook of language and intercultural communication* (New York, NY: Routledge), 70–85.

Byram, M., Holmes, P., and Savvides, N. (2013). Intercultural communicative competence in foreign language education: questions of theory, practice and research. *Lang. Learn. J.* 41, 251–253. doi: 10.1080/09571736.2013.836343

Cai, Y., Pan, Z., and Liu, M. (2022). Augmented reality technology in language learning: a meta-analysis. *J. Comput. Assist. Learn.* 38, 929–945. doi: 10.1111/jcal.12661

Carmigniani, J., and Furht, B. (2011). "Augmented reality: an overview" in *Handbook of augmented reality*. ed. B. Furht (New York, NY: Springer), 3–46.

Casella, G., and Coelho, M. (2013). "Augmented heritage: situating augmented reality mobile apps in cultural heritage communication" in *Proceedings of the 2013 international conference on information systems and Design of Communication*, 138–140.

Chen, G. M. (2007). "A review of the concept of intercultural effectiveness" in *The influence of culture in the world of business*. ed. M. Hinner (Berlin: Peter Lang), 95–116.

Chen, Y. (2019). Augmented reality technique assists target language learning. In: Innovative Technologies and Learning: Second International Conference, ICITL 2019, Proceedings 2. 2019, (pp. 558–567). Springer International Publishing: Tromsø.

Chen, S. Y., Hung, C. Y., Chang, Y. C., Lin, Y. S., and Lai, Y. H. (2018). "A study on integrating augmented reality technology and game-based learning model to improve motivation and effectiveness of learning English vocabulary" in *2018 1st International Cognitive Cities Conference (IC3)* (Piscataway: IEEE), 24–27.

Chen, C. M., and Li, Y. L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interact. Learn. Environ.* 18, 341–364. doi: 10.1080/10494820802602329

Chen, G. M., and Starosta, W. J. (1997). A review of the concept of intercultural sensitivity. *Hum. Commun.* 1, 1–16.

Chen, G. M., and Starosta, W. J. (2000). The development and validation of the intercultural sensitivity scale. *Hum. Commun.* 3, 1–15.

Chen, G. M., and Starosta, W. J. (2007). "Intercultural communication competence: a synthesis" in *The global intercultural communication reader* (New York, NY: Routledge), 235–258.

Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H., and Tian, M. (2019). An overview of augmented reality technology. *J. Phys. Conf. Ser.* 1237:022082. doi: 10.1088/1742-6596/1237/2/022082

Dalim, C. S. C., Sunar, M. S., Dey, A., and Billinghurst, M. (2020). Using augmented reality with speech input for non-native children's language learning. *Int. J. Hum. Comput. Stud.* 134, 44–64. doi: 10.1016/j.ijhcs.2019.10.002

Erbas, C., and Demirer, V. (2019). The effects of augmented reality on students' academic achievement and motivation in a biology course. *J. Comput. Assist. Learn.* 35, 450–458. doi: 10.1111/jcal.12350

Fathi, J., and Rahimi, M. (2022). Electronic writing portfolio in a collaborative writing environment: its impact on EFL students' writing performance. *Comput. Assist. Lang. Learn.* 1-39, 1–39. doi: 10.1080/09588221.2022.2097697

Fathi, J., Rahimi, M., and Liu, G. Z. (2023). A preliminary study on flipping an English as a foreign language collaborative writing course with video clips: its impact on writing skills and writing motivation. *J. Comput. Assist. Learn.* 39, 659–675. doi: 10.1111/jcal.12772

Gambo, J. M., Bahreman, N. T., Watties-Daniels, D., Neal, M., and Swoboda, S. M. (2017). Can mobile technology enhance learning and change educational practice? *Comput. Inform. Nurs.* 35, 375–380. doi: 10.1097/CIN.0000000000000380

Godwin-Jones, R. (2016). Augmented reality and language learning: From annotated vocabulary to place-based mobile games. *Lang. Learn. Technol.* 20, 9–19.

Grgurović, M., Chapelle, C. A., and Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL* 25, 165–198. doi: 10.1017/S0958344013000013

Hadjistassou, S., Avgousti, M. I., and Louca, P. (2019). "ReDesigning intercultural exchanges through the use of augmented reality" in *CALL and Complexity*, 157.

Hadjistassou, S., Avgousti, M. I., and Louca, P. (2021b). Designing augmented reality applications to facilitate intercultural telecollaborations. *J. Learn. Dev.* 8, 58–73. doi: 10.56059/jl4d.v8i1.413

Hadjistassou, S., Louca, P., Joannidou, S., and Muñoz, P. J. M. (2021a). "Design-based research as a framework for developing and deploying augmented reality applications and scenarios for intercultural exchange" in *CALL and Professionalisation: Short Papers from EUROCALL*, vol. 2021, 106.

Hsieh, M. C. (2016). "Teachers' and students' perceptions toward augmented reality materials" in *2016 5th IIAI international congress on advanced applied informatics (IIAI-AAI)* (San Jose del Cabo, Mexico: IEEE), 1180–1181.

Hsu, T. C. (2017). Learning English with augmented reality: do learning styles matter? *Comput. Educ.* 106, 137–149. doi: 10.1016/j.compedu.2016.12.007

Hung, H. T., Yang, J. C., Hwang, G. J., Chu, H. C., and Wang, C. C. (2018). A scoping review of research on digital game-based language learning. *Comput. Educ.* 126, 89–104. doi: 10.1016/j.compedu.2018.07.001

Hwang, G. J., Lai, C. L., and Wang, S. Y. (2015). Seamless flipped learning: a mobile technology-enhanced flipped classroom with effective learning strategies. *J. Comput. Educ.* 2, 449–473. doi: 10.1007/s40692-015-0043-0

Hwang, G. J., and Tsai, C. C. (2011). Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010. *Br. J. Educ. Technol.* 42, E65–E70. doi: 10.1111/j.1467-8535.2011.01183.x

Hwang, G. J., and Wu, P. H. (2014). Applications, impacts and trends of mobile technology-enhanced learning: a review of 2008–2012 publications in selected SSCI journals. *Int. J. Mobile Learn. Organ.* 8, 83–95. doi: 10.1504/IJMLO.2014.062346

Ibáñez, M. B., Portillo, A. U., Cabada, R. Z., and Barrón, M. L. (2020). Impact of augmented reality technology on academic achievement and motivation of students from public and private Mexican schools. A case study in a middle-school geometry course. *Comput. Educ.* 145:103734. doi: 10.1016/j.compedu.2019.103734

Kamarudin, S., Shoaib, H. M., Jamjoom, Y., Saleem, M., and Mohammadi, P. (2021). Students' behavioural intention towards e-learning practices through augmented reality app during COVID-19 pandemic in Saudi Arabia. *Interact. Learn. Environ.* 1-17, 1–17. doi: 10.1080/10494820.2021.2016863

Karacan, C. G., and Akoglu, K. (2021). Educational augmented reality technology for language learning and teaching: a comprehensive review. *Shanlax Int. J. Educ.* 9, 68–79. doi: 10.34293/education.v9i2.3715

Karacan, C. G., and Polat, M. (2022). Predicting pre-service English language teachers' intentions to use augmented reality. *J. Digit. Learn. Teach. Educ.* 38, 139–153. doi: 10.1080/21532974.2022.2083731

Kessler, M., Loewen, S., and Gönülal, T. (2022). Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience. *Comput. Assist. Lang. Learn.* 1-25, 1–25. doi: 10.1080/09588221.2023.2215294

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *Mod. Lang. J.* 92, 114–130. doi: 10.1111/j.1540-4781.2008.00690.x

Klopfer, E., and Squire, K. (2008). Environmental detectives—the development of an augmented reality platform for environmental simulations. *Educ. Technol. Res. Dev.* 56, 203–228. doi: 10.1007/s11423-007-9037-6

Koester, J., and Lustig, M. (2010). *Intercultural competence: Interpersonal communication across cultures*. Pearson: Boston.

Koester, J., and Olebe, M. (1988). The behavioral assessment scale for intercultural communication effectiveness. *J. Intercult. Relat.* 12, 233–246. doi: 10.1016/0147-1767(88)90017-X

Kramsch, C. (2013). "Teaching culture and intercultural competence" in *The encyclopedia of applied linguistics*. ed. C. A. Chapelle (Oxford: Blackwell Publishing Ltd.), 5555–5560.

Krüger, J. M., Buchholz, A., and Bodemer, D. (2019). "Augmented reality in education: three unique characteristics from a user's perspective" in *Proceedings of the 27th International Conference on Computational in Education*, 412–422.

Lee, J. (2022). Problem-based gaming via an augmented reality mobile game and a printed game in foreign language education. *Educ. Inf. Technol.* 27, 743–771. doi: 10.1007/s10639-020-10391-1

Lei, X., Fathi, J., Noorbakhsh, S., and Rahimi, M. (2022). The impact of mobile-assisted language learning on English as a foreign language learners' vocabulary learning attitudes and self-regulatory capacity. *Front. Psychol.* 13:872922. doi: 10.3389/fpsyg.2022.872922

Li, M., and Liu, L. (2022). Students' perceptions of augmented reality integrated into a mobile learning environment. *Library Hi Tech*. doi: 10.1108/LHT-10-2021-0345

Li, K. C., and Wong, B. T. M. (2021). A literature review of augmented reality, virtual reality, and mixed reality in language learning. *Int. J. Mobile Learn. Organ.* 15, 164–178. doi: 10.1504/IJMLO.2021.114516

Lin, J. J., and Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: a systematic review and meta-analysis. *Comput. Assist. Lang. Learn.* 32, 878–919. doi: 10.1080/09588221.2018.1541359

Lin, V., Liu, G. Z., and Chen, N. S. (2022). The effects of an augmented-reality ubiquitous writing application: a comparative pilot project for enhancing EFL writing instruction. *Comput. Assist. Lang. Learn.* 35, 989–1030. doi: 10.1080/09588221.2020.1770291

Lin, Y. J., and Wang, H. C. (2023). Applying augmented reality in a university English class: learners' perceptions of creativity and learning motivation. *Innov. Lang. Learn. Teach.* 17, 291–305. doi: 10.1080/17501229.2022.2040513

Liu, G. Z., Fathi, J., and Rahimi, M. (2023). Enhancing EFL learners' intercultural communicative effectiveness through telecollaboration with native and non-native speakers of English. *Comput. Assist. Lang. Learn.* 1-31, 1–31. doi: 10.1080/09588221.2022.2164778

Liu, Y., Holden, D., and Zheng, D. (2016). Analyzing students' language learning experience in an augmented reality mobile game: an exploration of an emergent learning environment. *Procedia Soc. Behav. Sci.* 228, 369–374. doi: 10.1016/j.sbspro.2016.07.055

Liu, G. Z., Rahimi, M., and Fathi, J. (2022). Flipping writing metacognitive strategies and writing skills in an English as a foreign language collaborative writing context: a mixed-methods study. *J. Comput. Assist. Learn.* 38, 1730–1751. doi: 10.1111/jcal.12707

Mahadzir, N. N., and Phung, L. F. (2013). The use of augmented reality pop-up book to increase motivation in English language learning for national primary school. *J. Res. Method Educ.* 1, 26–38. doi: 10.9790/7388-0112638

Matveev, V. V., Gribkov, D. N., Lomakin, D. E., and Kharunzheva, E. V. (2021). The use of augmented reality for the formation of multicultural competence of undergraduates. *Perspect. Scie. Educ.* 54, 487–504. doi: 10.32744/pse.2021.6.33

Miranda Bojórquez, E., Vergara Villegas, O. O., Cruz Sanchez, V. G., García-Alcaraz, J. L., and Favela Vara, J. (2016). Study on mobile augmented reality adoption for Mayo language learning. *Mob. Inf. Syst.* 2016, 1–15. doi: 10.1155/2016/1069581

O'Dowd, R. (2006). *Telecollaboration and the development of intercultural communicative competence*. Berlin: Langenscheidt.

Ozcelik, E., and Acarturk, C. (2011). Reducing the spatial distance between printed and online information sources by means of mobile technology enhances learning: using 2D barcodes. *Comput. Educ.* 57, 2077–2085. doi: 10.1016/j.compedu.2011.05.019

Parmaxi, A. (2023). Virtual reality in language learning: a systematic review and implications for research and practice. *Interact. Learn. Environ.* 31, 172–184. doi: 10.1080/10494820.2020.1765392

Parmaxi, A., and Demetriou, A. A. (2020). Augmented reality in language learning: a state-of-the-art review of 2014–2019. *J. Comput. Assist. Learn.* 36, 861–875. doi: 10.1111/jcal.12486

Pellas, N., Fotaris, P., Kazanidis, I., and Wells, D. (2019). Augmenting the learning experience in primary and secondary school education: a systematic review of recent trends in augmented reality game-based learning. *Virtual Reality* 23, 329–346. doi: 10.1007/s10055-018-0347-2

Piątkowska, K. (2015). From cultural knowledge to intercultural communicative competence: changing perspectives on the role of culture in foreign language teaching. *Intercult. Educ.* 26, 397–408. doi: 10.1080/14675986.2015.1092674

Portalla, T., and Chen, G. (2010). The development and validation of the intercultural effectiveness scale. *Intercult. Commun. Stud.* 19, 21–37.

Rahimi, M., and Fathi, J. (2022). Employing e-tandem language learning method to enhance speaking skills and willingness to communicate: the case of EFL learners. *Comput. Assist. Lang. Learn.* 1-37, 1–37. doi: 10.1080/09588221.2022.2064512

Reipschläger, P., and Dachselt, R. (2019). "Designar: immersive 3d-modeling combining augmented reality with interactive displays" in *Proceedings of the 2019 ACM international conference on interactive surfaces and spaces*, 29–41.

Richardson, D. (2016). Exploring the potential of a location based augmented reality game for language learning. *Int. J. Game-Based Learn.* 6, 34–49. doi: 10.4018/IJGBL.2016070103

Risager, K. (2013). "Culture and context" in *The encyclopedia of applied linguistics*. ed. C. A. Chapelle (Oxford: Blackwell Publishing Ltd), 1613–1616.

Sabie, D., Sheta, H., Ferdous, H. S., Kopalakrishnan, V., and Ahmed, S. I. (2023). "Be our guest: intercultural heritage exchange through augmented reality (AR)" in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1–15.

Saleem, M., Kamarudin, S., Shoaib, H. M., and Nasar, A. (2021). Influence of augmented reality app on intention towards e-learning amidst COVID-19 pandemic. *Interact. Learn. Environ.* 31, 3083–3097. doi: 10.1080/10494820.2021.1919147

Sannikov, S., Zhdanov, F., Chebotarev, P., and Rabinovich, P. (2015). Interactive educational content based on augmented reality and 3D visualization. *Proc. Comput. Sci.* 66, 720–729. doi: 10.1016/j.procs.2015.11.082

Shadiev, R., and Yu, J. (2022). Review of research on computer-assisted language learning with a focus on intercultural education. *Comput. Assist. Lang. Learn.*, 1–31. doi: 10.1080/09588221.2022.2056616

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., and Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. *Comput. Assist. Lang. Learn.* 36, 517–554. doi: 10.1080/09588221.2021.1933540

Sydorenko, T., Hellermann, J., Thorne, S. L., and Howe, V. (2019). Mobile augmented reality and language-related episodes. *TESOL Q.* 53, 712–740. doi: 10.1002/tesq.507

Taguchi, T., Magid, M., and Papi, M. (2009). "The L2 motivational self system among Japanese, Chinese and Iranian learners of English: a comparative study" in *Motivation, language identity and the L2 self*. eds. Z. Dörnyei and E. Ushioda (Bristol: Multilingual Matters), 66–97.

Tecedor, M., and Vasseur, R. (2020). Videoconferencing and the development of intercultural competence: insights from students' self-reflections. *Foreign Lang. Ann.* 53, 761–784. doi: 10.1111/flan.12495

Teng, C. H., Chen, J. Y., and Chen, Z. H. (2018). Impact of augmented reality on programming language learning: efficiency and perception. *J. Educ. Comput. Res.* 56, 254–271. doi: 10.1177/0735633117706109

Tommerdahl, J. M., Dragonflame, C. S., and Olsen, A. A. (2022). A systematic review examining the efficacy of commercially available foreign language learning mobile apps. *Comput. Assist. Lang. Learn.* 1-30, 1–30. doi: 10.1080/09588221.2022.2035401

Vygotsky, L. S. (1984). *The problem of age*. New York: Plenum.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge: MIT Press.

Wang, H. C. (2023). Facilitating English L2 learners' intercultural competence and learning of English in a Taiwanese university. *Lang. Teach. Res.* 27, 1032–1048. doi: 10.1177/1362168820969359

Wang, S., and Vásquez, C. (2012). Web 2.0 and second language learning: what does the research tell us? *CALICO J.* 29, 412–430. doi: 10.11139/cj.29.3.412-430

Wen, Y. (2021). Augmented reality enhanced cognitive engagement: designing classroom-based collaborative learning activities for young language learners. *Educ. Technol. Res. Dev.* 69, 843–860. doi: 10.1007/s11423-020-09893-z

Wood, J. T. (2008). *Communication in our lives*. Belmont, CA: Wadsworth.

Xie, H., Chu, H. C., Hwang, G. J., and Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: a systematic review of journal publications from 2007 to 2017. *Comput. Educ.* 140:103599. doi: 10.1016/j.compedu.2019.103599

Xu, Z., Chen, Z., Eutsler, L., Geng, Z., and Kogut, A. (2020). A scoping review of digital game-based technology on English language learning. *Educ. Technol. Res. Dev.* 68, 877–904. doi: 10.1007/s11423-019-09702-2

Yang, S., and Mei, B. (2018). Understanding learners' use of augmented reality in language learning: insights from a case study. *J. Educ. Teach.* 44, 511–513. doi: 10.1080/02607476.2018.1450937

Yu, J., Denham, A. R., and Searight, E. (2022). A systematic review of augmented reality game-based learning in STEM education. *Educ. Technol. Res. Dev.* 70, 1169–1194. doi: 10.1007/s11423-022-10122-y

Zafar, S., and Zachar, J. J. (2020). Evaluation of HoloHuman augmented reality application as a novel educational tool in dentistry. *Eur. J. Dent. Educ.* 24, 259–265. doi: 10.1111/eje.12492

Zhao, Q. (2018). The application of augmented reality visual communication in network teaching. *Int. J. Emerg. Technol. Learn.* 13:57. doi: 10.3991/ijet.v13i07.8780

Appendix

Interview Questions

1. How did the AR Language Learning App impact your overall language learning experience? Can you describe any specific instances where the app helped you overcome challenges or enhance your understanding?
2. What aspects of the AR Language Learning App did you find most valuable in terms of increasing your engagement and motivation to learn?
3. In what ways did the app contribute to your development of intercultural competence?
4. How did engaging in virtual conversations with culturally diverse virtual characters in the app enhance your cultural understanding and language skills? Can you discuss any specific insights or perspectives you gained through these interactions?
5. Looking back on your experience with the app, how do you believe it compared to other language learning methods or tools you have used? What advantages did the app offer that may have been lacking in other approaches?

frontiers | Frontiers in Education

Dynamics of automatized measures of creativity: mapping the landscape to quantify creative ideation

Ijaz Ul Haq and Manoli Pifarré*

Faculty of Education, Psychology and Social Work, University of Lleida, Lleida, Spain

The growing body of creativity research involves Artificial Intelligence (AI) and Machine learning (ML) approaches to automatically evaluating creative solutions. However, numerous challenges persist in evaluating the creativity dimensions and the methodologies employed for automatic evaluation. This paper contributes to this research gap with a scoping review that maps the Natural Language Processing (NLP) approaches to computations of different creativity dimensions. The review has two research objectives to cover the scope of automatic creativity evaluation: to identify different computational approaches and techniques in creativity evaluation and, to analyze the automatic evaluation of different creativity dimensions. As a first result, the scoping review provides a categorization of the automatic creativity research in the reviewed papers into three NLP approaches, namely: text similarity, text classification, and text mining. This categorization and further compilation of computational techniques used in these NLP approaches help ameliorate their application scenarios, research gaps, research limitations, and alternative solutions. As a second result, the thorough analysis of the automatic evaluation of different creativity dimensions differentiated the evaluation of 25 different creativity dimensions. Attending similarities in definitions and computations, we characterized seven core creativity dimensions, namely: novelty, value, flexibility, elaboration, fluency, feasibility, and others related to playful aspects of creativity. We hope this scoping review could provide valuable insights for researchers from psychology, education, AI, and others to make evidence-based decisions when developing automated creativity evaluation.

KEYWORDS

review, creativity process, ideation, evaluation, artificial intelligence

1. Introduction

Creativity as a 21st century skill is increasingly becoming an explicit part of educational policy initiatives and curricula (Plucker et al., 2023). Creativity is a multifaceted concept, and research in this area has made remarkable progress in understanding the different components embedded in creativity phenomena, such as idea generation through collaborative creative (co-creative) processes (Sawyer, 2011, 2022). Furthermore, research also revealed the significance of another important component of creativity: creativity evaluation (Guo et al., 2023), which is the ability to accurately identify creative ideas, solutions, or characteristics among individuals to understand their creative strengths and potential (Kim et al., 2019). In the educational context, creativity evaluation is an essential step for teachers and students because it is helpful to monitor, refine, and implement creative ideas, which could improve students' creative performance in the creative process (Rominger et al., 2022).

Creativity evaluation poses a challenging problem in creativity research. Creativity evaluation mainly involves four dimensions: fluency (number of meaningful ideas), flexibility (number of different categories), elaboration (detailed ideas), and novelty (uniqueness of ideas) (Bozkurt Altan and Tan, 2021). To evaluate these creativity dimensions, various manual creativity evaluations (paper-based) and psychological tests have been commonly used (Rafner et al., 2022). Examples are the Torrance Tests of Creative Thinking (Torrance, 2008), Creativity Assessment Packet (CAP) (Williams, 1980), and Divergent Production abilities (DP), (Guilford, 1967). Other ways to evaluate creativity include a rating scale (Gong and Zhang, 2017; Birkey and Hausserman, 2019), a survey and questionnaire (De Stobbeleir et al., 2011; Gong et al., 2019), using a grading rubric (Vo and Asojo, 2018), and subjective scoring of creativity dimensions (George and Wiley, 2020). However, these manual creativity evaluations face some challenges, e.g., being error-prone (experts' ratings do not always agree on what is creative) and time-consuming (Said-Metwaly et al., 2017; Doboli et al., 2020). These challenges can be tackled using automated creativity evaluation supported by AI techniques which can also enrich co-creation by providing real-time feedback to guide students to develop novel solutions (George and Wiley, 2020; Kenworthy et al., 2023).

Artificial intelligence (AI) focuses on enabling machines to perform tasks that typically demand human intelligence. Within AI, machine learning (ML) algorithms learn from data to make predictions. Notably, computer vision is used for analyzing figural data, and NLP is used for analyzing textual data. Given our focus on textual ideas, NLP enables machines to comprehend, interpret, analyze, and generate human language (Braun et al., 2017). NLP contains a variety of approaches and techniques such as text similarity, text classification, topic modeling, information extraction, and text generation, each with its computational techniques spanning from statistical methods to predictive and deep learning models. NLP provides different opportunities to compute variables related to creativity dimensions. Among these, the following five variables could be computed in the vector space provided by NLP: (1) Contextual and semantic similarity are applied to measure the uniqueness of ideas and originality (Hass, 2017; Doboli et al., 2020); (2) text clustering could identify different categories in the text; (3) text classification is used to compute novelty (Simpson et al., 2019); (4) keyword searching is mainly used to compute elaboration (Dumas et al., 2021); and (5) information retrieval could be applied to score the level of idea elaboration (Vartanian et al., 2020). These implications of NLP in co-creative processes can be used to automatically evaluate creativity and support co-creation by providing feedback (Bae et al., 2020; Kang et al., 2021; Kovalkov et al., 2021).

Considering the above implications of NLP, current research focuses on studying how different computational techniques can measure creativity dimensions (Doboli et al., 2020). Research on this topic has been very productive and has designed other computational techniques to measure creativity dimensions, e.g., (1) novelty is measured by keyword similarity (Prasch et al., 2020), part of speech tagging (Karampiperis et al., 2014; Camburn et al., 2019), and different ML classifiers, such as Bayesian

classifiers, random tree, and Support Vector Machine (SVM) (Manske and Hoppe, 2014; Simpson et al., 2019; Doboli et al., 2020); (2) originality dimension is measured by Latent Semantic Analysis (LSA) (Dunbar and Forster, 2009), Global Vectors for word representation (GloVe) (Dumas et al., 2021), and part of speech tagging (Georgiev and Casakin, 2019); (3) fluency dimension is measured by LSA (Dumas and Dunbar, 2014; LaVoie et al., 2020); (4) elaboration dimension is measured by parts of speech tagging (Dumas et al., 2021); and (5) level of details dimension is measured by text-mining methods (Camburn et al., 2019).

This study aims to tackle the following four main challenges that current research faces when designing computational techniques to measure creativity: (1) a range of computational techniques evaluating various creativity dimensions; (2) there is no consensus about the use of a specific technique for computing a specific creativity dimension; (3) some of the studies do not expose and argue the rationale that supports the use of a specific technique to compute a specific creativity dimension, e.g., evaluation of the category switch dimension of creativity using LSA (Dunbar and Forster, 2009); and (4) the need to consider the limitations of computational techniques that could affect the evaluation of creativity dimensions (Olivares-Rodríguez et al., 2017; Doboli et al., 2020). Considering these challenges, as per our knowledge, no existing literature review addresses the above four challenges. Therefore, this exploration led us to two research questions: (1) What NLP approaches and techniques are used to automatically measure creativity? and (2) What creativity dimensions are computed automatically, and how? These research questions enable us to address the previous four challenges in automatic creativity evaluation. Furthermore, these research questions help to understand the concept of NLP approaches and creativity dimensions, their applications in evaluating creativity dimensions, identify research gaps and limitations, and propose alternative solutions for advancing the evaluation and promotion of creativity. Therefore, we chose a scoping review because it helps to understand key concepts and identify knowledge gaps (Munn et al., 2018) to inspire innovation and improve the education of future generations through advanced technologies.

2. Research objectives

This scoping review aims to meet the following two objectives.

1. To identify and categorize different ML approaches used in automatic creativity evaluation, highlighting their application scenarios and limitations of computational approaches and techniques. This categorization could contribute to a deeper understanding of the contribution that different ML approaches can make to automatic creativity.
2. To analyze the definition and computation of different creativity dimensions used in automatic creativity evaluation research. This analysis can help establish a joint agreement on creativity dimensions and their computation, which will pave the way for advancements in automatic creativity evaluation.

3. Method

This section describes the sampling method we used to collect and compile the state-of-the-art approaches to automatic creativity evaluation. Our methodological framework follows the PRISMA technique (Dickson and Yeung, 2022) by conducting a scoping review to find relevant and significant research papers by identifying the following four core concepts.

1. Creativity: The articles must be related to creativity, especially the creative process (Sawyer, 2011).
2. Measurement/evaluation/assessment of creativity dimensions.
3. Technology: We selected those studies that are assisted or evaluated with technology support. This core concept aims to review the technological support for creativity evaluation and explore future research in the creative process.
4. Domain: We focused on the creativity process applicable in the educational sector that helps to enhance students' creativity. Other fields such as medicine, finance, and business were excluded from the search query.

Exploring the current literature considering the above four core concepts, peer-reviewed journals and conference papers are included in this mapping study. Regarding the time span, we searched from 2005 to 2021, although interestingly, according to our inclusion–exclusion criteria, the oldest study included is from 2009, and most are from recent past years. It indicates that automatic creativity evaluation has recently grabbed researchers' attention and is still an open and active research problem.

We excluded articles focused on the person's or organization's creativity evaluation. We excluded domains other than education, e.g., medicine and finance. Articles in other languages apart from English published before 2005 and articles with no technological role and creativity were also excluded.

For this mapping study, we extracted articles published in Scopus with the search query: [(creativ* OR "Creative Process" OR "Novelty" OR "Flexibility" OR "Fluency" OR "Elaboration" OR "Originality") AND (Measur*OR Evaluat* OR Asses* OR Calcul* OR Analys* OR Scor* OR Qunat*) AND (Automat* OR Comput* OR Machin* OR Natural* OR Artificial* OR Deep learning OR Mathemat* OR Mining) AND (E-learning OR educa* OR Learn* OR School OR students*)].

The search query resulted in 364 research articles. By applying the inclusion and exclusion criteria while reading the title, abstract, keywords, and conclusion, the search is filtered to 65 articles. Furthermore, the authors read, checked, and discussed the selected articles and conducted all the screening stages to answer the two research questions. The consensus among the authors developed by solving discrepancies since member checking is a well-established procedure to build up "trustworthiness" in qualitative research (Toma, 2011). After this process, a total of 26 articles were finally included in this scoping review. The overall article selection procedure through the PRISMA technique is depicted in Figure 1.

4. Results

4.1. Approaches and techniques used in automatic creativity evaluation (RQ1)

The compilation of computational approaches and techniques in automatic creativity evaluation research to answer the first research question gives the following three results;

The first result reveals that creativity evaluation research spreads over three different NLP approaches, namely, (1) text similarity, which measures the relatedness and closeness among words, sentences, or paragraphs presented in a numerical space; (2) text classification, which is a supervised learning approach (needs data training) that requires ML algorithms [such as the K-nearest neighbor (KNN) algorithm and random forest] to analyze text automatically and then to assign a set of predefined tags or categories; and (3) text mining that uses NLP to examine and transform extensive unstructured text data to discover new information and patterns. These three NLP approaches and their computational techniques identified in the studies included in this review are displayed in Figure 2.

As a second result, the scoping review shows that text similarity is the most common approach (69% of the reviewed studies), followed by text classification (27%), and text mining is less commonly used (only 4% of the studies), as shown in Figure 2.

As a third result, our scoping review has identified and categorized the computation techniques used in the three NLP approaches (text similarity, text classification, and text mining) and the creativity dimensions that were evaluated automatically. In the following sections, we present the mapping that we have built after a thorough analysis of all the studies included in the scoping review.

Regarding the text similarity approach, NLP converts textual ideas into a numerical vector space. To do this conversion, the studies revised the use of a wide range of techniques that could be classified into the next three categories: string-based similarity, corpus-based similarity, and knowledge-based similarity. These three categories and their computational techniques identified in the reviewed studies are shown in Figure 3, and Table 1 maps automatic creativity evaluation studies into the three categories and techniques used.

In the first category, string-based similarity (6% of the text similarity approach of reviewed studies) matches exact keywords or alphabet strings, e.g., Longest Common Substring (LCS) or N-gram (a subsequence of n items from a given sequence of text). The string similarity of ideas with the existing ideas in the database is computed by using keyword matching (Prasch et al., 2020).

In the second category, corpus-based similarity is mostly used (72% of textual similarity), and the results are presented in Table 1. The corpus-based similarity is classified into two sub-categories: On the one hand, the statistical-based models, e.g., LSA, present corpus in the word-document matrix as words in row vectors and each document as a column vector, and weighting schemes and dimension reduction schemes are applied before calculating the cosine similarity among word vectors (Martin and Berry, 2007; Wagire et al., 2020). On the other hand, the deep learning-based models (both word and sentence embeddings) use supervised (which need to be trained on data), semi-supervised,

FIGURE 1
Screening procedure of the articles using the PRISMA technique.
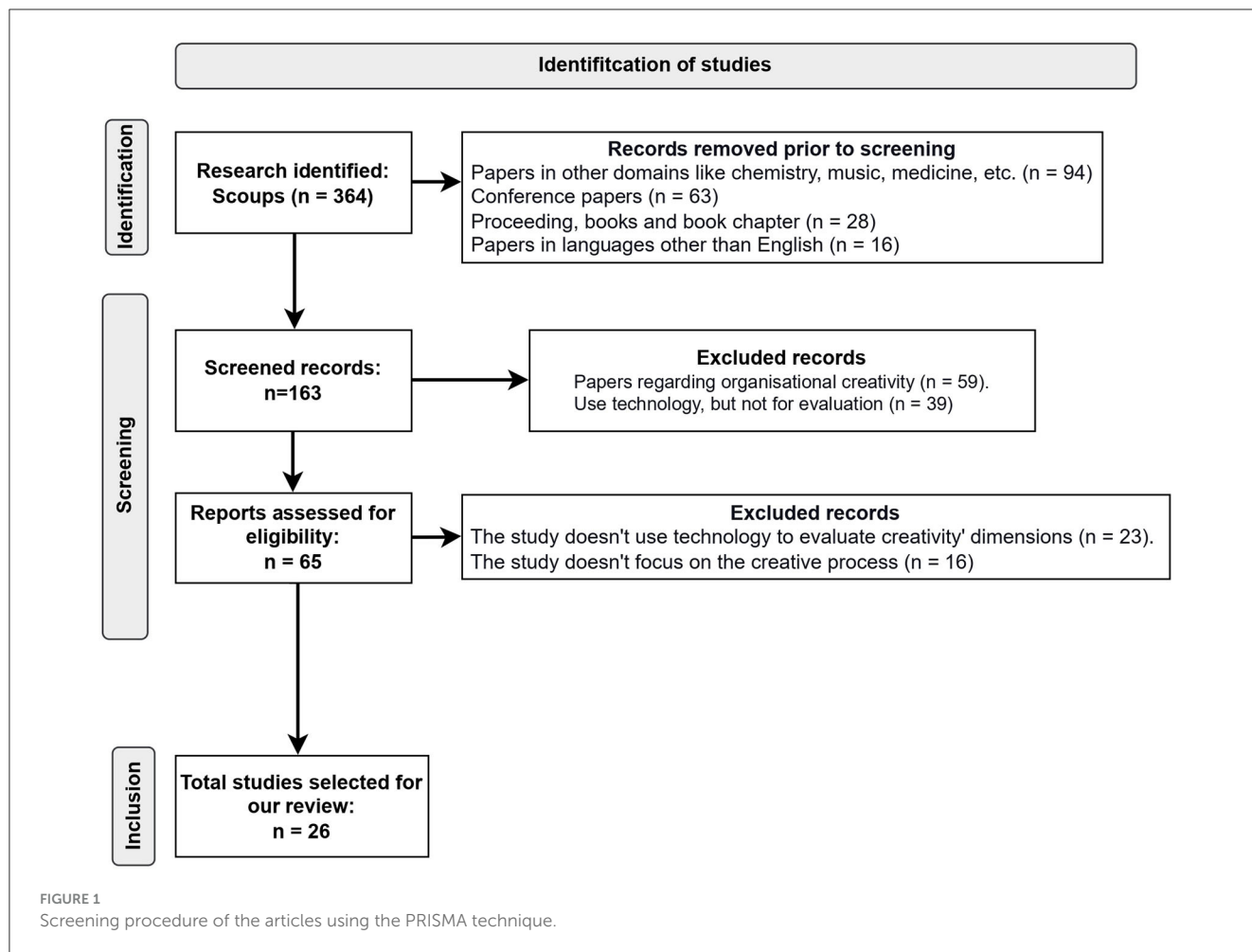
or unsupervised methods (no prior training) that are trained on a large corpus, e.g., Wikipedia and common crawl dataset. Deep learning models such as Word2Vec (Mikolov et al., 2013) or GloVe (Pennington et al., 2014) use knowledge from large datasets, encode the data, and find similarities in words or sentences. The GloVe model showed reliable results as compared with the experts' scores, especially for single-word creativity tasks (Beaty and Johnson, 2021; Johnson and Hass, 2022).

In the third category, knowledge-based similarity (used in 22% of text similarity approaches in reviewed studies, as presented in Table 1) using the knowledge of ontologies represents the textual data on a semantic network graph consisting of nodes representing semantic memory and lines. Ontologies are the dictionaries of millions of words and are lexically associated, e.g., WordNet, Wikipedia, and DBpedia.

Text classification is the second NLP approach used by 27% of the reviewed studies in automatic creativity evaluation depicted in Figure 1. Classification is an ML technique that categorizes text into predefined categories. The classification consists of four main steps: (1) data collection, pre-processing (data acquisition, cleaning, and labeling), and data presentation (feature selection, dividing into training and testing datasets); (2) applying classifier models; (3) evaluation of classifiers; and (4) prediction (output of the testing data). These four steps are influential factors when

applying text classification in automatic creativity evaluation. Table 2 gives an overview of the classification approach, the datasets, classifiers, evaluations, and creativity dimensions in creativity evaluation research.

Text mining is the third approach in automatic creativity evaluation, which is the practice of analyzing a vast collection of textual data to capture key concepts, trends, patterns, and hidden relationships. In the scoping review, text mining is used (Dumas et al., 2021). The studies used four mining techniques, e.g., all words count, stop list inclusion (defined terms that are not meaningful), counting part of speech, and applying inverse document frequency (a technique to extract rare and important documents).

4.2. Creativity dimensions are computed automatically (RQ2)

In the studies included in this scoping review of automatic creativity evaluation, we differentiated 25 different creativity dimensions. These 25 dimensions of creativity are displayed in the second column (Manifestation) of Table 3. We analyze the similarities in the conceptual definition and computational approach employed in various studies that consider different
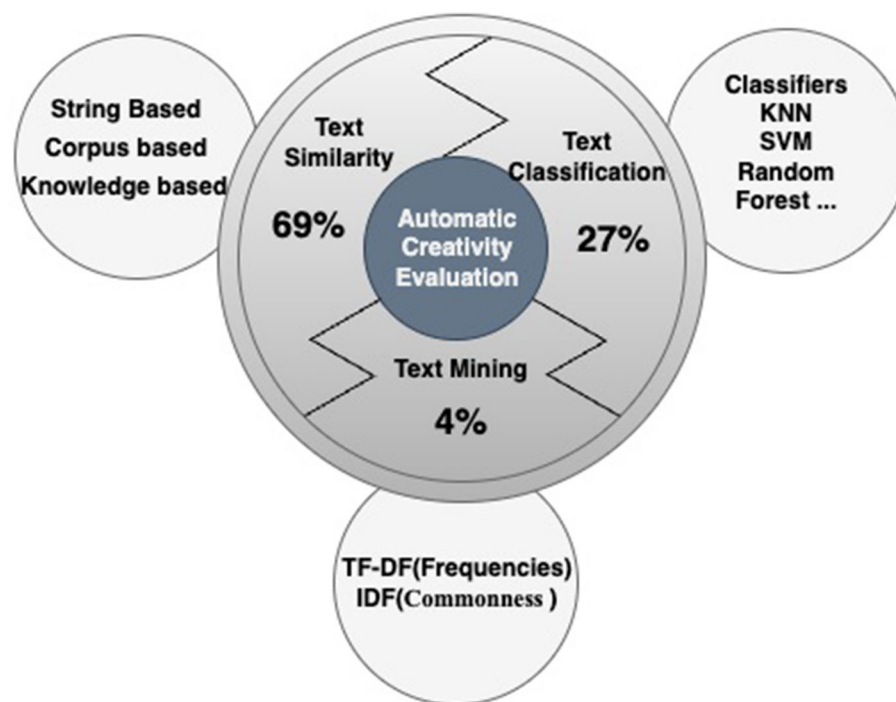
FIGURE 2
Different NLP approaches in creativity evaluation.
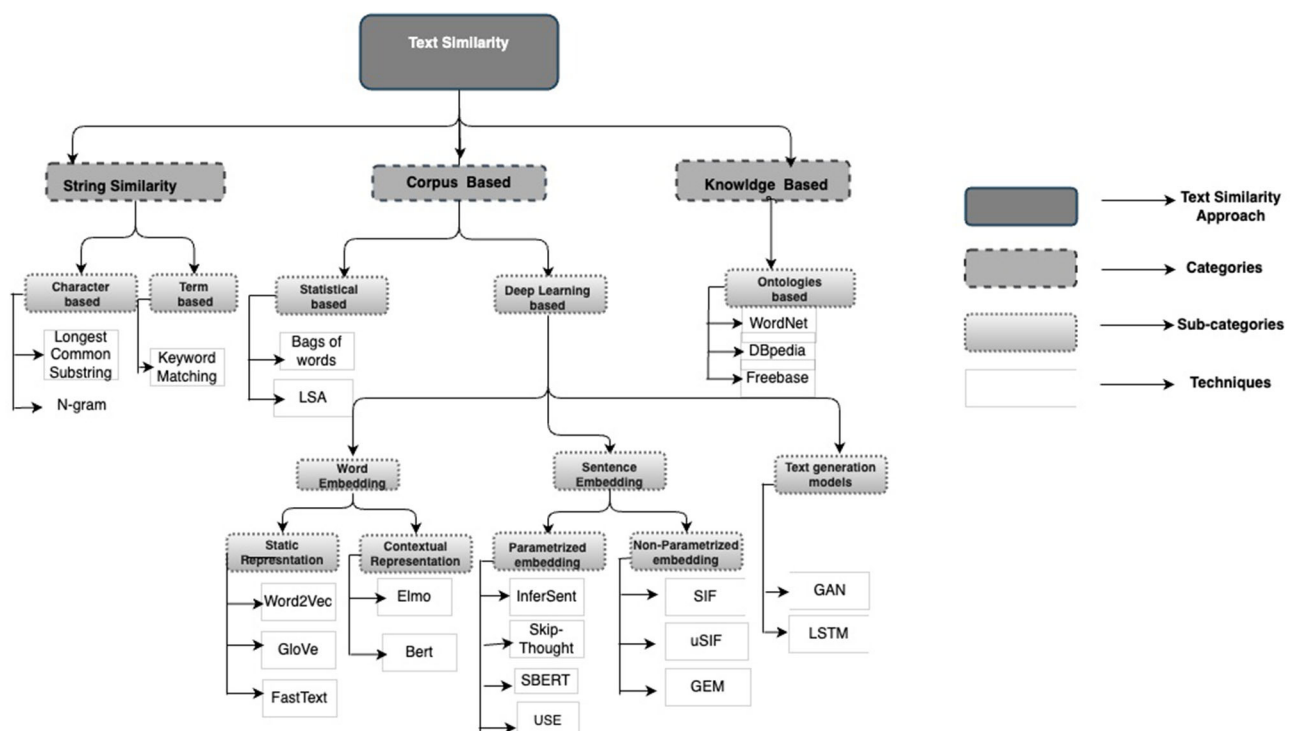
FIGURE 3
Text similarity approaches, categories, sub-categories, and their computational techniques.

dimensions for assessing creativity. This analysis allows us to categorize these 25 manifestations of creativity into seven core creativity dimensions, namely, novelty, value, flexibility, elaboration, fluency, feasibility, and others related to playful aspects of creativity such as humor or recreational efforts, which are displayed in the first column of Table 3 (Core Dimension).

TABLE 1 Categorizing of review studies in text similarity approaches and percentages of studies included in the review that use each approach.

Text similarity categories	Sub-categories	Vectorization techniques	Dimensions	Studies
String-based 6%		Keyword matching	Novelty, usefulness	Prasch et al., 2020
Knowledge-based 22%		Part of speech tagging	Novelty, level of details	Camburn et al., 2019
		Part of speech tagging	Originality, value, overall value, feasibility	Karampiperis et al., 2014
		Clustering in the knowledge graph	Novelty, surprise, rarity, recreational effort	Georgiev and Casakin, 2019
		Semantic network	Flexibility	Cosgrove et al., 2021
Corpus-based 72%	Statical based	LSA	Category switch, variety, original, prune originality, common use	Dunbar and Forster, 2009
		LSA	Fluency, originality	Dumas and Dunbar, 2014
		LSA	Similarity, fluency	LaVoie et al., 2020
		Vectorization of linguistic features	Similarity	Zuñiga et al., 2017
	Deep learning	Word2Vec	Originality, flexibility, fluency	Sung et al., 2022
		GloVe	Originality	Acar et al., 2021; Beaty and Johnson, 2021; Dumas et al., 2021
		GloVe	Similarity of text	Olson et al., 2021
		GloVe	Diversity (Novelty)	Johnson and Hass, 2022
		Universal sentence encoder	Novelty	Kenworthy et al., 2023
		GAN	Novelty, value, surprise	Franceschelli and Musolesi, 2022
		LSTM	originality	Marrone et al., 2022

TABLE 2 Text classification-based creativity evaluation studies.

Datasets	Classifiers	Evaluation	Dimensions	Studies
In 4,099,877 solutions from Project Euler Website	Linear regression and SVM	Comparison with expert rating	Novelty, usefulness, quality	Manske and Hoppe, 2014
Two datasets were used: 1. ideas: 1,480; 2, domain dataset: a collection of 1,144 sports datasets from Wikipedia	SVM, neural networks (NN), logistic regression, decision trees, KNN, and Naive Bayes	F-measure is a measure of a test's accuracy. Precision and recall are calculated	Novelty	Doboli et al., 2020
Semeval-2017 jokes, 4,030 short texts, and VU Amsterdam Metaphor Corpus	Bayesian approach	Bayesian approach is compared to the best-worst scaling method	Novelty, humor	Simpson et al., 2019
Internet movies database and Rotten Tomatoes dataset contained textual, image, and numerical attributes	SVM, random forest, ridge regression, Bayesian regression, and K-nearest regression	Correlation analysis	Novelty, value, influence, unexpectedness	Shrivastava et al., 2017
User queries Wikipedia as knowledge source	Random trees	Sensitivity, Specificity	Diversity	Olivares-Rodríguez et al., 2017
203 responses present in the multiplex lexical network	Logistic regression, random forest, and SVM classifiers	Entropy	Fluency	Stella and Kenett, 2019
1,214 recipes, 2,130 ingredients, and 235 cooking techniques	K-neighbor classifier, SVM, multi-layer perceptron classifier, and the random forest	The scoring function of classifiers, random forest, has the best results. No other evaluation	Novelty, adaptiveness, style, transcendence, realization	Jimenez-Mavillard and Suarez, 2022

Furthermore, the results obtained to answer research question two are illustrated in Figure 4, which displays the percentage of the seven core creativity dimensions identified in this review. These results show that novelty is the most evaluated dimension in the studies compiled in this scoping review.

TABLE 3　Characterization of 25 creativity dimensions into seven core creativity dimensions (first column) and creativity dimensions manifested (second column) based on similarities in definitions (third column) and computation (fourth column).

Core dimension	Dimension manifestation	Dimension definition	Dimension computation	Study
Novelty	Novelty	Novelty is an idea with respect to prior ideas or deviation from existing solutions	Textual similarity of a given solution to all existing or previous solutions	Manske and Hoppe, 2014; Prasch et al., 2020; Kenworthy et al., 2023
		A measure of how unique a concept is relative to others	Span (Path length) is the sum of distances of each entity or unique words from the central entity or topic (e.g., predefined hierarchical topical categories of Wikipedia).	Camburn et al., 2019
		The deviation from existing knowledge/experience	Average semantic distance between the dominant terms included in the textual representation of the story, compared to the average semantic distance of the dominant terms in all stories	Karampiperis et al., 2014
		-	Pairwise text similarity using linguistic features	Simpson et al., 2019
		Novelty is defined as a unique solution	From surprise and relevance score surprise term is computed from document term frequency in idea data, and relevance term is calculated from domain dataset (sports was collected from Wikipedia)	Doboli et al., 2020
		How an artifact is different from others	Calculation of the distance between a given artifact and the other artifacts in a descriptive space	Shrivastava et al., 2017; Franceschelli and Musolesi, 2022
		Novelty to originality score and defines that the creative method led to more innovative products	The classifier models learn from ingredients and techniques and classify them as novel or not novel in the case study of culinary products	Jimenez-Mavillard and Suarez, 2022
	Originality	Similarity to existing ideas	Semantic distance between the responses	Dunbar and Forster, 2009; Song et al., 2020; Beaty and Johnson, 2021
		Originality is referred as a novelty	The semantic distance among ideas	Dumas and Dunbar, 2014
		Statistically infrequent responses	Semantic distance between the responses	Acar et al., 2021
		A response that is more unusual within a given context would be more Original.	Semantic distance between a given responses	Dumas et al., 2021
	Similarity	The similarity of meaning between multiple texts	The similarity of the new response was measured with topic clusters, rubrics, and example responses	LaVoie et al., 2020
		The similarity of the original poem to the translated poem	Similarity distance is calculated between original (English language) and translated poems (Spanish)	Zuñiga et al., 2017
		Similar contexts have smaller distances	Semantic distance between different words	Olson et al., 2021
	Diversity	Semantic distance among user queries	Semantic similarity is estimated of each user-issued query to the k most relevant concepts for the challenge using distance formulas	Olivares-Rodríguez et al., 2017
		The degree to which participants engaged in semantic context search	Semantic diversity refers to the degree to which the contexts surrounding words vary in their meanings	Johnson and Hass, 2022

(Continued)

TABLE 3 (Continued)

Core dimension	Dimension manifestation	Dimension definition	Dimension computation	Study
	Rarity	A rare combination of properties	The sum of weights on the min-weight closure of the cluster graph is compared to the maximum sum of weights in the story	Karampiperis et al., 2014
	Common use	Common uses of objects in Object use tasks	Each response was compared to the most common use of the corresponding object (collected previously from Common Use Judges)	Dunbar and Forster, 2009
	Surprise or unexpectedness	Unexpectedness or surprise defines how different the artifact or some of its attributes are from expected behavior	The similarity of a given artifact with other artifacts	Shrivastava et al., 2017; Franceschelli and Musolesi, 2022
	Influence	How impactful or inspiring it has been	The similarity of an artifact with other artifacts occurs later	Shrivastava et al., 2017
Value	Value	A measure of how artifact is valued by domain experts for artifact	Datapoint is highly valuable if its combination of correlated dimensions leads to a better rating prediction.	Shrivastava et al., 2017; Franceschelli and Musolesi, 2022
	Overall value	Overall value of the outcome of the designs from design ideation	Semantic analysis of verbalizations can be promising to measure the semantic value.	Georgiev and Casakin, 2019
	Quality	Quality is related to reliability, maintainability, extensibility, and adaptability.	Quality and Usefulness are computed from two metrics. Static Code Metrics: Line of codes Dynamic Code Metrics: number of visited lines	Manske and Hoppe, 2014
	Usefulness	The correct solutions to programming tasks		Manske and Hoppe, 2014
	Adaptiveness and Style	Adaptiveness is the solution to solve a problem. Style is elegance and other aesthetic qualities	Adaptiveness as useful solutions style as quality	Jimenez-Mavillard and Suarez, 2022
Elaboration	Elaboration	The degree to which they explain and embellish their responses	Counting based on: (1) Unweighted Word, (2) Stop listed Inclusion, (3) Part of Speech Inclusion, and (4) Inverse Frequency Weighting.	Dumas et al., 2021
	Level of details	Level of details of the ideas	Count of named entities. Examples of entities are person, place, things, money, etc	Camburn et al., 2019
Flexibility	Flexibility	Semantic memory structure	1. Cosine similarity to estimate the edges between nodes in semantic network. 2. Number of similar clusters	Sung et al., 2022
	Category switch	Number of changes in the category of use between responses	The similarity scores between successive response pairs were averaged for each object	Dunbar and Forster, 2009
	Variety	Measure the variety of responses produced by each person	The similarity scores between every single pair of responses for an object were also averaged as a measure of the variety of responses produced by each person	Dunbar and Forster, 2009
Fluency	Fluency	Number of ideas	Counting the number of ideas	Dumas and Dunbar, 2014; Stella and Kenett, 2019; Sung et al., 2022
Feasibility	Feasibility	Feasibility can be materialized or achieved in real practice	Polysemy, abstraction, and IC are highly correlated to the feasibility score	Georgiev and Casakin, 2019
	Transcendence and realism	Transforming into reality	Development of the product and its communication with the other products	Jimenez-Mavillard and Suarez, 2022

(Continued)

TABLE 3 (Continued)

Core dimension	Dimension manifestation	Dimension definition	Dimension computation	Study
Other	Humor	Humor is funniness	Pairwise comparison of text	Karampiperis et al., 2014
	Recreational effort	Difficult to achieve	The number of different clusters that each story contains as compared to the maximum number of clusters in a story of the whole group	Simpson et al., 2019

FIGURE 4
Percentage distribution of each core creativity dimension in the reviewed studies.
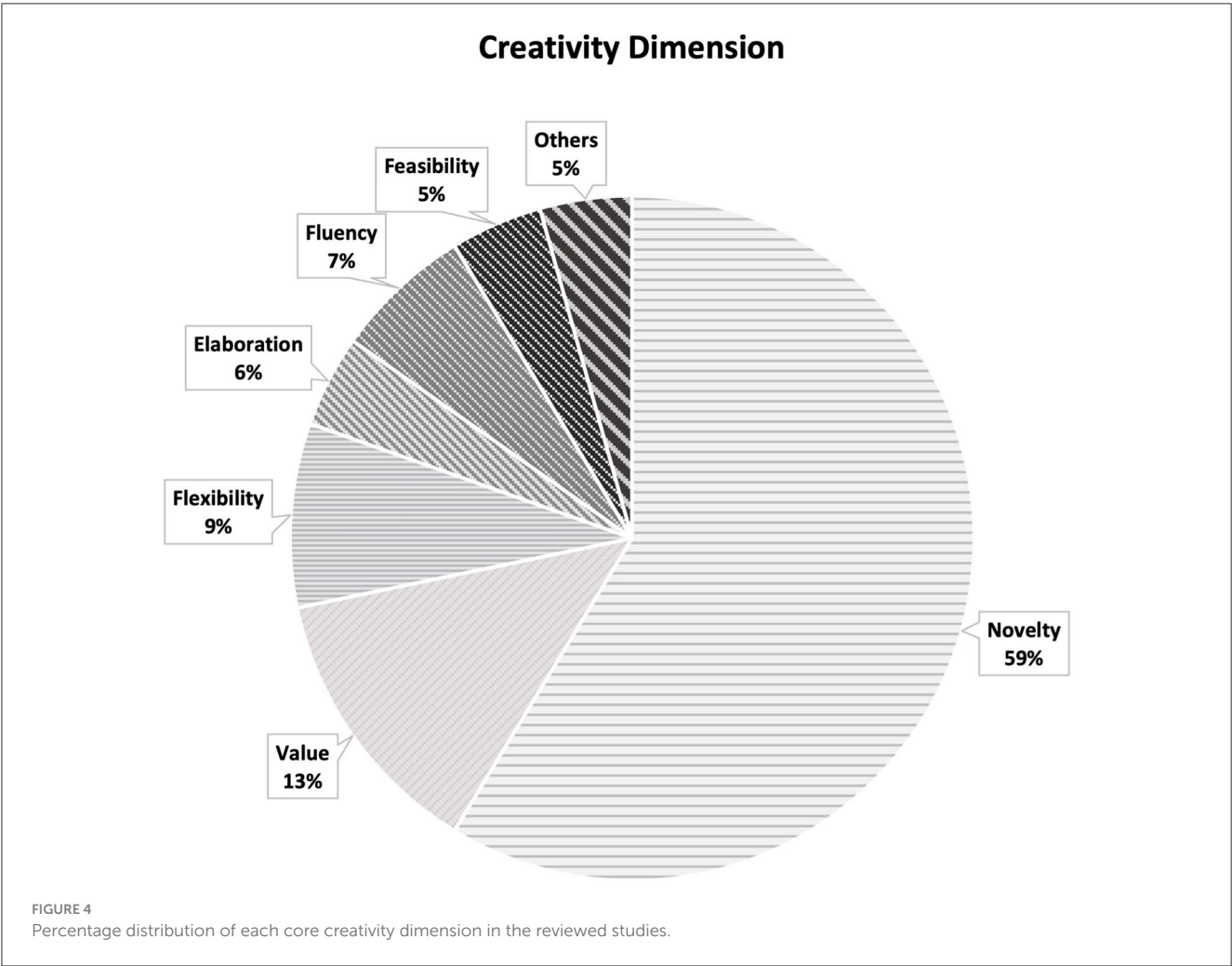
5. Discussion

5.1. Approaches and techniques used in automatic creativity evaluation

The scoping review identified three main NLP approaches used in automatic creativity evaluation, namely, (1) text similarity, (2) text classification, and (3) text mining. In the next sections, we discuss the contribution of each computational approach to automatic creativity evaluation, argue their applications, discuss their limitations, identify research gaps, and make further recommendations for automatic creativity evaluations.

Regarding the text similarity approach, the scoping review revealed that it is used in 69% of the studies, which helps understand creative thinking (Li et al., 2023). Our analysis concluded that the widespread use of textual similarity in automatic creativity evaluation is because automatic creativity evaluation is more focused on evaluating originality, novelty, similarity, or diversity dimensions of creativity. The computations of these dimensions involve assessing the similarity of an idea with the existing ideas. The text similarity approach provides a variety of computational techniques to measure the similarity of ideas, as shown in Figure 3.

Concerning the three categories of text similarity, namely, string similarity, corpus-based similarity, and knowledge-based

similarity as set out in Table 3, the scoping review shows differences in the process of similarity computation that have an impact on how they are applied. On the one hand, string-based and knowledge-based similarities have limited application in automatic creativity evaluation because string-based only considers syntactic similarity (not semantic) and knowledge-based only extracts from text-specific entities, such as a person's name, place, and money (Camburn et al., 2019). During ideation, the knowledge-based approach might focus on entities rather than technical terms or scientific jargon within the sentence used by sentences solving a scientific challenge. For example, when brainstorming about renewable energy solutions, the knowledge-based approach might not capture specific terms such as "photovoltaics" or "wind turbines." On the other hand, corpus-based techniques are widely used, so in the following, we elaborate on corpus-based techniques.

Regarding corpus-based similarity, it has been commonly used in automatic evaluation because it provides a wide range of computational techniques, from simple statistical to deep learning models, as shown in Figure 2. Considering that a statistical model such as LSA is applied to examine semantic similarity, memory, and creativity (Beaty and Johnson, 2021), it has shown a more reliable scoring technique of originality on divergent thinking tasks than human ratters (Dunbar and Forster, 2009; Dumas and Dunbar, 2014; LaVoie et al., 2020), as shown in Table 1. We argue that LSA uses statistical techniques, including Probabilistic Latent Semantic Analysis (Hofmann, 1999), Latent Dirichlet Allocation (Blei et al., 2003), and Non-Negative Matrix Factorization (Lee and Seung, 1999), which limit its implication because these consider words statistics (e.g., co-occurrence of words) instead of word contextual and semantic meaning. These limitations are addressed by deep learning models, which we discuss below.

Recently, drastic changes in NLP research with the development of deep learning models based on deep neural architectures have unlocked ways to model text with more nuance and complexity. This advancement started with the development of word embedding models such as GloVe or Word2Vec pre-trained, including Wikipedia, news articles, and web pages. These predictive models use a neural network with one or more hidden layers to learn the vector representations of words. The GloVe showed comparable results to human experts' scores in single-word creativity tasks (Beaty and Johnson, 2021; Olson et al., 2021). However, word embedding models do not differentiate between a list of keywords and a meaningful sentence; hence, they cannot capture the semantic and contextual meaning of the whole sentence (idea) in the vector space. The vectorization of the whole sentence is one major innovation in text modeling: The transformer architecture generally outperforms word embedding models on standard tasks, and often by large margins (Wang et al., 2018, 2019), which utilizes a concept called attention (Vaswani et al., 2017). Attention makes it computationally tractable for a transformer model to consider a long sequence of text by selecting the most important parts of the sequence. Attention allows the training of large models on words and the complex contexts in which those words occur. This development resulted mainly in two kinds of categories, pre-trained sentence embedding models and text generation models which are discussed below.

Sentence embedding models vectorize the whole sentence into a vector space that keeps the semantic and contextual meaning of the entire sentence. The sentence embedding models are unsupervised techniques that do not require external data, e.g., Unsupervised Smooth Inverse Frequency (uSIF) (Ethayarajh, 2018) and Geometric Sentence embedding (GEM) (Yang et al., 2018). Some transformers allowed the tuning of parameters or training on their datasets to improve performance (if a large dataset is available), e.g., Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), Sentence Transformer (Reimers and Gurevych, 2019), MPNet (Song et al., 2020), Skip-Thought (ST) (Kiros et al., 2015), InferSent (Conneau et al., 2017), and Universal Sentence Encoder (USE) (Cer et al., 2018). In creativity research, the USE model is used to evaluate the novelty of ideas (Kenworthy et al., 2023). We argue that more exploration is needed to apply different, or combinations of sentence embedding models to evaluate creative ideas in an open-ended co-creation.

Text generation models generate new text that is similar to a given text prompt, such as Generative Pre-trained Transformer (GPT-3) (Brown et al., 2020), Text-to-Text Transfer Transformer (T5) (Raffel et al., 2020), and Long Short-Term Memory (LSTM) (Huang et al., 2022). In creativity research, one of the text-generated models, the Generative Adversarial Network (GAN) (Aggarwal et al., 2021), is used by Franceschelli and Musolesi (2022) to evaluate novelty, surprise, and relevance. We present two criticisms regarding using text generation models for evaluating open-ended ideas. First, text generation is specialized to generate text from a given text that could be useful for dialog generation, machine translation, chatbots, and prompt-based learning (Liu et al., 2023). Second, as the model becomes better at generating text with an improved understanding of language, it is more likely to generate text that closely resembles the input data rather than producing more novel or creative outputs. However, we argue that text generation models are not tested on a larger scale in creativity research, so future investigations could help understand these limits.

Finally, two conclusions are drawn from the above discussion. First, for single-word tasks in creativity research, word embedding models can be used, especially the GloVe embedding model, which is widely used. Word embedding models represent words in a high-dimensional vector space, enabling the computation of their contextual and semantic similarity with other words. Second, for open-ended co-creation resulting in ideas of sentence structure, sentence embedding models can be useful in three ways: (a) In open-ended ideation, mostly the ideas are in sentence structure, so these sentence models present the whole sentence in a vector space, capturing the semantic and contextual meaning of the whole sentence; (b) sentence embedding models outperform the word embedding models for textual similarity tasks; and (c) sentence embedding models can also be applied to small datasets and open-ended problems because these models are pre-trained over large corpora. Finally, we recommend not only validating sentence embedding models but also applying text generation models within a broader context of co-creation.

We concluded that sentence embedding models offer a powerful measure that can be used alongside statistical (Acar et al., 2021), word embedding models (Organisciak et al., 2023), and standard subjective scoring methods of the creative process and its output (Kenett, 2019).

Text classification approach refers to the automated categorization or labeling of textual data into predetermined classes or categories using machine learning classifiers. A large dataset is used for text classification, which is divided into training and testing (the usual ratio is 70% training and 30% testing datasets). An ML classifier learns from the training dataset and then uses the knowledge learned during training to categorize the testing dataset. Therefore, integrating text classification into automatic creativity evaluation depends on four key factors: the dataset, the selection of appropriate machine learning classifiers, the accuracy of the ML classifier, and the creativity dimensions being evaluated. These factors in the reviewed studies using the text classification approach are highlighted in Table 2.

Using text classification, it is essential to consider the dataset factor for three reasons: First, the datasets used for classification need pre-processing and labeling. Pre-processing includes removing noisy or irrelevant information, and labeling includes giving a class label to each idea. Second, a large dataset is required to train the ML classifiers. The prediction capability of ML classifiers increases with an increase in the amount of data used for training. All studies reviewed in Table 2 except Stella and Kenett (2019) use more than a thousand ideas or solutions for the classification problem. A smaller dataset may need better or more balanced results. Third, ML classifiers trained on one type of data cannot be applied to another kind of data. For example, classifiers trained on datasets from the linguistic domain cannot be used to test data from the scientific domain.

Furthermore, classifier selection and accuracy are also critical. Regarding classifier selection, the working methods of ML classifiers are different and dependent on the nature of the dataset, e.g., SVM works well for multiclass classification, and random forest excels in scenarios involving numerical and categorical features. Similarly, logistic regression works on linear problems; the K-neighbor classifier is best for text, and SVM can also work for multiclass dataset classification. The Bayesian approach is a simple and fast algorithm. The reviewed studies lack arguments for using a specific classifier in their studies. Regarding the accuracy of ML, there is a risk of not getting high accuracy. Different automatic evaluators are used to evaluate model accuracy, such as confusion matrix, entropy, and sensitivity, as shown in Table 2. It is suggested to apply several classifiers, and one with high accuracy can be used for prediction in a similar domain.

Finally, the text classification approach can be applied to evaluate different dimensions of creativity; however, it requires a large, labeled dataset, which limits its application in creativity research. We also argue that the dataset's preparation and labeling might be expensive, which mitigates the advantages of automatic evaluation over manual creativity evaluation, e.g., accuracy, cost, and time. Furthermore, the text classification problems are domain-dependent. So, for creativity tasks, such as object use tasks and alternate use tasks, some public datasets are available that could apply to similar tasks. However, it is not useful for small and open-ended creative tasks because it is not enough to train an ML classifier and is domain-independent. In short, large dataset preparation, labeling, and domain dependence make the text classification approach less reliable and expensive than manual creativity evaluation.

Text mining employs NLP statistical computation to discover new information and patterns. It uses statistical indicators such as the frequency of words, word patterns, and correlation within words. Dumas et al. (2021) implemented four text-mining techniques and measured the elaboration score in Alternate Use Tasks (AUT). Elaboration was computed in four different ways: (1) unweighted word count method: count the number of words; (2) stop listed inclusion: a preliminary agreed list of stop words; (3) parts of speech include verbs, nouns, adjectives, and adverbs; and (4) inverse frequency weighting: commonness of a word in an initial corpus of text.

The above text-mining techniques are the basic statistical operations in NLP. Text mining holds the potential to handle a massive amount of data to discover new information, patterns, trends, relationships, etc., that could be useful in creativity research. Text-mining applications include search engines, product suggestion analysis, social media analytics, and trend analysis.

5.2. Automatically computed creativity dimensions

The scoping review noted 25 creativity dimensions computed automatically. However, our analysis reveals that these creativity dimensions are not sufficiently based on previous creativity research and theory. Therefore, we have found some theoretical and methodological inconsistencies that should be tackled in future research. In this line of argument, first, we highlight that some of the creativity dimensions studied in the scoping review are defined and computed, building links with the challenges or the creativity tasks designed for the experiment but not with a strong theoretical framework. For example, a category switch is defined as the similarity difference between two successive responses in object use tasks (Dunbar and Forster, 2009). Another example is the creativity dimensions of quality (reusability) and usefulness (Degree of completion) that are defined and computed in the context of programming problems (Manske and Hoppe, 2014). Second, another reason for the inconsistency among the dimensions of creativity is the variation in manifestations employed across the reviewed articles. Specifically, it has been observed that dimensions such as novelty (Prasch et al., 2020), similarity (LaVoie et al., 2020), and originality (Beaty and Johnson, 2021) are defined in a similar manner, a strong focus on the similarity between ideas or solutions. Moreover, these dimensions are often measured using semantic textual similarity, although different computational techniques are performed.

To mitigate these shortcomings, this scoping review has thoroughly analyzed the conceptual and computational framework used in each study and contributed to the emergence of seven core creativity dimensions that could be automatically evaluated and bring more consistency to this research area. These seven core creativity dimensions are novelty, elaboration, flexibility, value, feasibility, fluency, and others related to playful aspects of creativity, such as humor and recreational efforts. Following, we discuss each core creativity dimension identified and highlight the key aspects of its conceptual definition and computational approach.

Novelty is the first core dimension in automatic creativity research that is most evaluated in 59% of the reviewed studies. Despite this high interest, our revision indicates multifariousness in defining and measuring novelty. As a consequence of that, the reviewed studies refer to novelty using the following different words or manifestations, namely, (1) uniqueness: the uniqueness of a concept related to the other concepts (Camburn et al., 2019); (2) originality: how different the outcome is from standard/other solutions (Georgiev and Casakin, 2019) or semantic distance among ideas (Beaty and Johnson, 2021); (3) similarity: the similarity of meaning between multiple texts (LaVoie et al., 2020) or similarity distance between the texts (Olson et al., 2021); (4) diversity: the diversity of users' entered queries; (5) rarity: the rare combination or rare ideas (Karampiperis et al., 2014) or unique solution (Doboli et al., 2020); (6) common use: the difference between common and uncommon solutions; (7) surprise: that how much an artifact is different from existing attributes (Shrivastava et al., 2017); and (8) influence or the comparison of an artifact with other artifacts (Shrivastava et al., 2017).

Nonetheless, the diversity in labeling and defining the novelty dimension, our analysis identified the next six characteristics that could be included in defining novelty and assisting its automatic evaluation: (1) deviation from the standard, routine way of solving a given problem (Manske and Hoppe, 2014); (2) semantic distance between ideas (Beaty and Johnson, 2021); (3) similarity of meaning between multiple texts (LaVoie et al., 2020); (4) Semantic similarity of the user query to the concepts in the challenge; (5) combination of properties (Karampiperis et al., 2014); and (6) surprise and unexpected ideas (Shrivastava et al., 2017). These six characteristics involved in the definition of novelty in the studies reviewed give an account of the complexity of defining the novelty dimension and acknowledge the challenges in developing automatic measures for novelty.

Despite these challenges, the scoping review has highlighted some common computing approaches and techniques to measure novelty as a core dimension and they can be synthesized in the next five characteristics: (1) distance of the new solution to the existing solution (Manske and Hoppe, 2014); (2) semantic distance among ideas (Beaty and Johnson, 2021; Olson et al., 2021); (3) semantic similarity of user queries and relevant concepts in Wikipedia; (4) semantic distance between the clusters in a story; and (5) semantic distance between the consecutive fragments of the story (Karampiperis et al., 2014). It concludes that when developing an automatic evaluation of novelty, the semantic distance of a solution to existing solutions should be considered.

Value is the second core dimension identified in automatic creativity evaluation. The scoping review identified the next four concepts related to value (Shrivastava et al., 2017; Franceschelli and Musolesi, 2022): (1) overall value, which relates how an artifact is perceived by society (Georgiev and Casakin, 2019); (2) quality, this concept is mainly used for programming solutions when they embody specific attributes such as reliability, characterized by error-free operation; maintainability, denoting ease of maintenance; extensibility, encompassing scalability and simplified modification; and, adaptability, reflecting the flexibility to integrate new technologies seamlessly (Manske and Hoppe, 2014); (3) the concept of usefulness which is linked to the notion of correctness; and (4) the concept of adaptiveness, it pertains

to useful solutions that effectively address specific problems (Jimenez-Mavillard and Suarez, 2022). In sum, these four concepts share a common meaning of usefulness and quality that could be considered the value dimension of creativity. Furthermore, from a computer science perspective, value, quality, usefulness, adaptiveness, and style are the non-functional characteristics related to quality attributes. These quality attributes have different computations depending on the nature of the task, e.g., quality and useful programming solutions are the reusability and scalability of computer programs (Manske and Hoppe, 2014), and usefulness is the degree of completing the task (Prasch et al., 2020). Therefore, the value dimension needs clear definitions and computation metrics like other dimensions.

The third core dimension used in automatic creativity evaluation is flexibility. Flexibility refers to one of the key executive functions of creative thinking (Boot et al., 2017), which drives individuals to follow diverse directions, dimensions, and pathways (Acar et al., 2021), more likely to produce highly creative ideas (Zhang et al., 2020). Creativity research defines flexibility in two distinct ways. First, it involves category switching (Dunbar and Forster, 2009; Acar et al., 2019; Mastria et al., 2021), which refers to the ability to transition from one semantic concept to another. Second, flexibility is also measured by the number of semantic categories, varieties (Dunbar and Forster, 2009), or topics generated during the creative process. Owing to variations in the definition of flexibility across creativity research, different computational approaches are employed to compute this dimension. On one side, flexibility as a category switch is a measure of the similarity of one idea to all existing ideas. Therefore, semantic similarity approaches are used to evaluate flexibility, such as LSA (Dunbar and Forster, 2009), network graphs (Cosgrove et al., 2021), and sentence embedding models. On the other side, flexibility identifies semantic categories, varieties, or topics that can be evaluated using text clustering (Sung et al., 2022) or topic modeling techniques [e.g., Latent Dirichlet Allocation (LDA); Chauhan and Shah, 2021] to categorize or extract different topics from the textual ideas. We argue that flexibility as a category switch could be the easiest way to compute because it acquires simple text similarities rather than identifying categories in the text, which involve more variables and algorithms.

Regarding elaboration as a core creativity dimension in automatic creativity evaluation, it is defined as the degree of elaboration to which the participants embellish their responses (Camburn et al., 2019; Dumas et al., 2021) or which gives further details on adding reasoning or cause to an idea. Automatic creativity evaluation captures the level of detail of an idea by counting the number of words used in the idea (Camburn et al., 2019). The scoping review has identified four different methods for evaluating the level of idea elaboration: (1) counting all words in an idea (Counting unweighted measures); (2) counting stop words (words that do not have semantic meanings); (3) counting nouns, verbs, and adverbs; and (4) specifying and counting adjectives (parts of speech inclusion) and uncommon words with high weight (inverse frequency weighting). An idea with more words is considered an elaborated idea. We argue that the above-adopted computation of elaboration may not capture conjunctions (Tuzcu, 2021) or reasoning words (Sedova et al., 2019; Hennessy et al., 2020), adding more explanation to the ideas. Therefore, we suggest

the semantic search to specify the words that cause reasoning or words that give reason to the idea, such as because, therefore, and since.

Fluency is defined as the number of ideas generated during an ideation process. This scoping review showed that fluency is one of the core dimensions that finds consensus on its conceptual definition (number of ideas) and computational approach (counting ideas) (Dumas and Dunbar, 2014; Stella and Kenett, 2019). Creativity research claims that when there are more ideas, there is a greater chance of producing original ideas or products (Dumas and Dunbar, 2014). Fluency measurement is easy to implement and is independent of other ideas such as elaboration. Compared to novelty and flexibility, which require comparison with different ideas, fluency can be easily computed for each idea.

Feasibility is defined as the solution that is achievable in real practice (Georgiev and Casakin, 2019). The scoping review found transcendence and realization have been used as manifestations of feasibility as they refer to the achievement in real practice or transforming into reality (Jimenez-Mavillard and Suarez, 2022). These dimensions share the same characteristic of transforming an idea or solution into real practice, which is significant in creativity research. The creativity research highlights the significance of putting ideas into practice; however, the automatic computation of feasibility (Georgiev and Casakin, 2019), transcendence, and realization (Jimenez-Mavillard and Suarez, 2022) does not provide any rationale from the creativity research. Feasibility is mostly a product-oriented dimension and is mostly used in the ideation process, but finding transformable ideas into real practice is still a challenge to address. Therefore, it is a dimension that needs further research to automatically measure feasible, transcendent, and realistic ideas.

Finally, other dimensions associated with the playful aspects of creativity, such as humor (Simpson et al., 2019) and recreational effort (Karampiperis et al., 2014), were identified in the reviewed articles. Humor, representing the funniness of ideas, is typically measured through pairwise text comparison techniques. At the same time, recreational effort is defined as a solution that is difficult to achieve and is measured using clustering methods. These dimensions contribute to the playful nature of creativity, so it is essential to establish clear definitions and develop suitable computational approaches from both psychological and computer science perspectives.

6. Conclusion

This article has the objective of conducting a scoping review of automatic creativity evaluation from creativity and computer science perspectives. To meet this objective, we defined two research questions: The first identifies the NLP approaches and techniques used in automatic creativity, and the second analyzes which and how different creativity dimensions are computed.

The first research question's contributions are multi-fold: (1) identifying the existing ML approaches and techniques in automatic creativity evaluation; (2) categorizing the approaches into different groups for deep compilation, e.g., text similarity, text classification, and text mining. Among these, text similarity is commonly used; (3) classifying creativity evaluation studies

into different techniques accordingly, e.g., classifying studies in text similarity approaches using various techniques such as string similarity, corpus-based similarity, and knowledge-based similarity. Our results showed that corpus-based methods are widely used for automatic creativity evaluation. Corpus-based techniques, LSA (Dunbar and Forster, 2009; Dumas and Dunbar, 2014; LaVoie et al., 2020) and GloVe algorithm (Beaty and Johnson, 2021; Olson et al., 2021), have shown a positive correlation with human experts' similarity scores; (4) identifying the limitations of the critical challenge and identifying alternative techniques, for example, statistical and word embedding techniques are generally used, but they cannot capture the semantic and contextual meaning of a whole sentence; and (5) providing a broad overview of all existing automatic creativity to give a deeper understanding of all the approaches. We concluded that word embedding models, especially GloVe, work better for single-word tasks, and for open-ended ideas in sentence structure, sentence embedding models could provide promising results.

The second research question's contributions are also multi-fold: first, we have examined what creativity dimensions are automatically evaluated in the different articles analyzed in this scoping review. In contrast to creativity research, which has standardized tests that evaluate four specific dimensions, 25 different creativity dimensions are found in automatic creativity evaluation. Second, the scoping review has analyzed how these dimensions are defined and measured in automatic creativity evaluation. We found similarities in the definitions and computations of different creativity dimensions. Finally, based on a thorough analysis of the definitions and computations used in the studies, we characterized the 25 dimensions into seven core dimensions. This analysis helps elaborate a coherent and consistent framework about core creativity dimensions and their computation.

The overall contributions of this scoping review bridge the realms of computer science and education. For computer scientists, this review provides insights to refine existing NLP approaches and provides opportunities for developing more novel NLP methods for evaluating and promoting creativity. Meanwhile, educators can use these automatic evaluations as pedagogical tools in real-world classroom practices. The implications of automatic creativity evaluation could help assess and nurture creativity, which is becoming an explicit part of educational policy initiatives and curricula. Ultimately, this scoping review leverages AI as a valuable tool in evaluating and enhancing creativity capable of equipping future citizens with the necessary competencies to generate innovative solutions to the world's complex economic, environmental, and social challenges.

6.1. Limitations and future work

This scoping review has two limitations, which may have conditioned our results. The first limitation could be the search keyword strategy, which may be insufficient to include key articles in our field of study. Second, the exclusion and inclusion criteria may suffer from the omission of relevant studies that could have answered our research questions. We tried to mitigate this risk

by carefully constructing an inclusive search string and providing explicit inclusion and exclusion criteria with co-authors' consensus.

In future, concluding from this scoping review, we intend to design experimental research to evaluate the reliability of deep learning models such as sentence embedding models to measure the novelty of ideas in an open-ended co-creative process. Furthermore, we also suggest using text generation models to recommend diverse hints to improve divergent thinking in the creative process. Regarding the automatic evaluation of creativity dimensions, our review highlighted that there is still a research gap in studies that fully automate the main core dimensions of creativity. So, we plan to simultaneously measure different core creativity dimensions by evaluating idea datasets with ML techniques. Finally, the development of reliable and automatic evaluation of the different dimensions of creativity could be the seed for the design and the delivery of real-time recommendations during the creative process that could trigger students' 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.

Author contributions

IU has contributed in the conceptualization of the paper, methodology and investigation; he has participated in writing the original manuscript, revision and edition. MP is the principal investigator of the research project and she has designed the project, she has also contributed in the conceptualization of the paper, methodology and investigation; she has participated in writing the manuscript, revision and edition. Both authors contributed to the article and approved the submitted version.

Funding

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

Acar, S., Berthiaume, K., Grajzel, K., Dumas, D., Flemister, C., and Organisciak, P. (2021). Applying automated originality scoring to the verbal form of torrance tests of creative thinking. *Gifted Child Quart.* 67, 3–17. doi: 10.1177/00169862211061874

Acar, S., Runco, M. A., and Ogurlu, U. (2019). The moderating influence of idea sequence: A re-analysis of the relationship between category switch and latency. *Person. Indiv. Differ.* 142, 214–217. doi: 10.1016/j.paid.2018.06.013

Aggarwal, A., Mittal, M., and Battineni, G. (2021). Generative adversarial network: An overview ofvtheory and applications. *Int. J. Inform. Manage. Data Insights* 1, 100004. doi: 10.1016/j.jjimei.2020.100004

Bae, S. S., Kwon, O.-H., Chandrasegaran, S., and Ma, K.-L. (2020). "Spinneret: aiding creative ideationvthrough non-obvious concept associations," in *Proceedings of the 2020 CHI Conference on HumanvFactors in Computing Systems* 1–13. doi: 10.1145/3313831.3376746

Beaty, R. E., and Johnson, D. R. (2021). Automating creativity assessment with semdis: An open platformvfor computing semantic distance. *Behav. Res. Methods* 53, 757–780. doi: 10.3758/s13428-020-01453-w

Birkey, R., and Hausserman, C. (2019). "Inducing creativity in accountants' task performance: The effects of background, environment, and feedback," in *Advances in Accounting Education: Teaching and Curriculum Innovations* (Emerald Publishing Limited) 109–133. doi: 10.1108/S1085-462220190000022006

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022. doi: 10.5555/944919.944937

Boot, N., Baas, M., Mühlfeld, E., de Dreu, C. K., and van Gaal, S. (2017). Widespread neural oscillations in the delta band dissociate rule convergence from rule divergence during creative idea generation. *Neuropsychologia* 104, 8–17. doi: 10.1016/j.neuropsychologia.2017.07.033

Bozkurt Altan, E., and Tan, S. (2021). Concepts of creativity in design-based learning in STEM education. *Int. J. Technol. Design Educ.* 31, 503–529. doi: 10.1007/s10798-020-09569-y

Braun, D., Hernandez Mendez, A., Matthes, F., and Langen, M. (2017). "Evaluating natural language understanding services for conversational question answering systems," in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue* (Saarbrucken, Germany: Association for Computational Linguistics) 174–185. doi: 10.18653/v1/W17-5522

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Proc. Syst.* 33, 1877–1901. doi: 10.48550/arXiv.2005.14165

Camburn, B., He, Y., Raviselvam, S., Luo, J., and Wood, K. (2019). "Evaluating crowdsourced design concepts with machine learning," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* 7. doi: 10.1115/DETC2019-97285

Cer, D., Yang, Y., Kong, S.-,y., Hua, N., Limtiaco, N., John, R. S., et al. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175.*

Chauhan, U., and Shah, A. (2021). Topic modeling using latent dirichlet allocation: A survey. *ACM Comput. Surv.* 54, 1–35. doi: 10.1145/3462478

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., and Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364.*

Cosgrove, A. L., Kenett, Y. N., Beaty, R. E., and Diaz, M. T. (2021). Quantifying flexibility in thought: The resiliency of semantic networks differs across the lifespan. *Cognition* 211, 104631. doi: 10.1016/j.cognition.2021.104631

De Stobbeleir, K. E., Ashford, S. J., and Buyens, D. (2011). Self-regulation of creativity at work: The role of feedback-seeking behavior in creative performance. *Acad. Manage. J.* 54, 811–831. doi: 10.5465/amj.2011.64870144

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Dickson, K., and Yeung, C. A. (2022). PRISMA 2020 updated guideline. *Br. Dental J.* 232, 760–761. doi: 10.1038/s41415-022-4359-7

Doboli, S., Kenworthy, J., Paulus, P., Minai, A., and Doboli, A. (2020). "A cognitive inspired method for assessing novelty of short-text ideas," in *2020 International Joint Conference on Neural Networks (IJCNN)* (IEEE), 1–8. doi: 10.1109/IJCNN48605.2020.9206788

Dumas, D., and Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Think. Skills Creat.* 14, 56–67. doi: 10.1016/j.tsc.2014.09.003

Dumas, D., Organisciak, P., Maio, S., and Doherty, M. (2021). Four text-mining methods for measuring elaboration. *J. Creat. Behav.* 55, 517–531. doi: 10.1002/jocb.471

Dunbar, K., and Forster, E. (2009). "Creativity evaluation through latent semantic analysis," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 31.

Ethayarajh, K. (2018). "Unsupervised random walk sentence embeddings: A strong but simple baseline," in *Proceedings of The Third Workshop on Representation Learning for NLP* 91–100. doi: 10.18653/v1/W18-3012

Franceschelli, G., and Musolesi, M. (2022). Deepcreativity: measuring creativity with deep learning techniques. *Intell. Artif.* 16, 151–163. doi: 10.3233/IA-220136

George, T., and Wiley, J. (2020). Need something different? Here's what's been done: Effects of examples and task instructions on creative idea generation. *Memory Cogn.* 48, 226–243. doi: 10.3758/s13421-019-01005-4

Georgiev, G. V., and Casakin, H. (2019). "Semantic measures for enhancing creativity in design education," in *Proceedings of the Design Society: International Conference on Engineering Design* (Cambridge: Cambridge University Press), 369–378. doi: 10.1017/dsi.2019.40

Gong, Z., Shan, C., and Yu, H. (2019). The relationship between the feedback environment and creativity: a self-motives perspective. *Psychol. Res Behav. Manag.* 12, 825–837. doi: 10.2147/PRBM.S221670

Gong, Z., and Zhang, N. (2017). Using a feedback environment to improve creative performance: a dynamic affect perspective. *Front. Psychol.* 8, 1398. doi: 10.3389/fpsyg.2017.01398

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *J. Creat. Behav.* 1, 3–14. doi: 10.1002/j.2162-6057.1967.tb00002.x

Guo, Y., Lin, S., Williams, Z. J., Zeng, Y., and Clark, L. Q. C. (2023). Evaluative skill in the creativeprocess: A cross-cultural study. *Think. Skills Creativ.* 47, 101240. doi: 10.1016/j.tsc.2023.101240

Hass, R. W. (2017). Tracking the dynamics of divergent thinking via semantic distance: Analytic methods and theoretical implications. *Memory Cogn.* 45, 233–244. doi: 10.3758/s13421-016-0659-y

Hennessy, S., Howe, C., Mercer, N., and Vrikki, M. (2020). Coding classroom dialogue: Methodological considerations for researchers. *Learning, Cult. Soc. Interact.* 25, 100404. doi: 10.1016/j.lcsi.2020.100404

Hofmann, T. (1999). "Probabilistic latent semantic indexing," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR'99* (New York, NY, USA: Association for Computing Machinery), 50–57. doi: 10.1145/312624.312649

Huang, R., Wei, C., Wang, B., Yang, J., Xu, X., Wu, S., et al. (2022). Well performance prediction based on long short-term memory (lstm) neural network. *J. Petroleum Sci. Eng.* 208, 109686. doi: 10.1016/j.petrol.2021.109686

Jimenez-Mavillard, A., and Suarez, J. L. (2022). A computational approach for creativity assessment of culinary products: the case of elbulli. *AI Soc.* 37, 331–353. doi: 10.1007/s00146-021-01183-3

Johnson, D. R., and Hass, R. W. (2022). Semantic context search in creative idea generation. *J. Creat. Behav.* 56, 362–381. doi: 10.1002/jocb.534

Kang, Y., Sun, Z., Wang, S., Huang, Z., Wu, Z., and Ma, X. (2021). "Metamap: Supporting visual metaphor ideation through multi-dimensional example-based exploration," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* 1–15. doi: 10.1145/3411764.3445325

Karampiperis, P., Koukourikos, A., and Koliopoulou, E. (2014). "Towards machines for measuring creativity: The use of computational tools in storytelling activities," in *2014 IEEE 14th International Conference on Advanced Learning Technologies* 508–512. doi: 10.1109/ICALT.2014.150

Kenett, Y. N. (2019). What can quantitative measures of semantic distance tell us about creativity? *Curr. Opin. Behav. Sci.* 27, 11–16. doi: 10.1016/j.cobeha.2018.08.010

Kenworthy, J. B., Doboli, S., Alsayed, O., Choudhary, R., Jaed, A., Minai, A. A., et al. (2023). Toward the development of a computer-assisted, real-time assessment of ideational dynamics in collaborative creative groups. *Creativ. Res. J.* 35, 396–411. doi: 10.1080/10400419.2022.2157589

Kim, S., Choe, I., and Kaufman, J. C. (2019). The development and evaluation of the effect of creative problem-solving program on young children's creativity and character. *Think. Skills Creativ.* 33, 100590. doi: 10.1016/j.tsc.2019.100590

Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., et al. (2015). "Skip-thought vectors," in *Advances in Neural Information Processing Systems* 28.

Kovalkov, A., Paaßen, B., Segal, A., Pinkwart, N., and Gal, K. (2021). Automatic creativity measurement in scratch programs across modalities. *IEEE Trans. Learn. Technol.* 14, 740–753. doi: 10.1109/TLT.2022.3144442

LaVoie, N., Parker, J., Legree, P. J., Ardison, S., and Kilcullen, R. N. (2020). Using latent semantic analysis to score short answer constructed responses: Automated scoring of the consequences test. *Educ. Psychol. Measur.* 80, 399–414. doi: 10.1177/0013164419860575

Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565

Li, Y., Du Ying, X. I. E., Liu, C., Yang, Y., Li, Y., and Qiu, J. (2023). A meta-analysis of the relationship 649 between semantic distance and creative thinking. *Adv. Psychol. Sci.* 31, 519. doi: 10.3724/SP.J.1042.2023.00519

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* 55, 1–35. doi: 10.1145/3560815

Manske, S., and Hoppe, H. U. (2014). "Automated indicators to assess the creativity of solutions to programming exercises," in *2014 IEEE 14th International Conference on Advanced Learning Technologies* 497–501. doi: 10.1109/ICALT.2014.147

Marrone, R., Cropley, D. H., and Wang, Z. (2022). Automatic assessment of mathematical creativity using natural language processing. *Creat. Res. J.* 2022, 1–16. doi: 10.1080/10400419.2022.2131209

Martin, D. I., and Berry, M. W. (2007). "Mathematical foundations behind latent semantic analysis," in *Handbook of Latent Semantic Analysis* 35–56.

Mastria, S., Agnoli, S., Zanon, M., Acar, S., Runco, M. A., and Corazza, G. E. (2021). Clustering and switching in divergent thinking: Neurophysiological correlates underlying flexibility during idea generation. *Neuropsychologia* 158, 107890. doi: 10.1016/j.neuropsychologia.2021.107890

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., and Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* 18, 1–7. doi: 10.1186/s12874-018-0611-x

Olivares-Rodríguez, C., Guenaga, M., and Garaizar, P. (2017). Automatic assessment of creativity in heuristic problem-solving based on query diversity. *DYNA* 92, 449–455. doi: 10.6036/8243

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., and Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proc. Nat. Acad. Sci.* 118, e2022340118. doi: 10.1073/pnas.2022340118

Organisciak, P., Newman, M., Eby, D., Acar, S., and Dumas, D. (2023). How do the kids speak? Improving educational use of text mining with child-directed language models. *Inf. Learn. Sci.* 124, 25–47. doi: 10.1108/ILS-06-2022-0082

Pennington, J., Socher, R., and Manning, C. D. (2014). "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP), 1532–1543. doi: 10.3115/v1/D14-1162

Plucker, J. A., Meyer, M. S., Karami, S., and Ghahremani, M. (2023). "Room to run: Using technology to move creativity into the classroom," in *Creative Provocations: Speculations on the Future of Creativity, Technology and Learning* (Springer) 65–80. doi: 10.1007/978-3-031-14549-0_5

Prasch, L., Maruhn, P., Brünn, M., and Bengler, K. (2020). "Creativity assessment via novelty and usefulness (canu) – approach to an easy to use objective test tool," in *Proceedings of the Sixth International Conference on Design Creativity (ICDC)* 019–026. doi: 10.35199/ICDC.2020.03

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 687 5485–5551. doi: 10.48550/arXiv.1910.10683

Rafner, J., Biskjær, M. M., Zana, B., Langsford, S., Bergenholtz, C., Rahimi, S., et al. (2022). Digital games for creativity assessment: strengths, weaknesses and opportunities. *Creat. Res. J.* 34, 28–54. doi: 10.1080/10400419.2021.1971447

Reimers, N., and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084.*

Rominger, C., Benedek, M., Lebuda, I., Perchtold-Stefan, C. M., Schwerdtfeger, A. R., Papousek, I., et al. (2022). Functional brain activation patterns of creative metacognitive monitoring. *Neuropsychologia* 177, 108416. doi: 10.1016/j.neuropsychologia.2022.108416

Said-Metwaly, S., Van den Noortgate, W., and Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity.* 4, 238–275. doi: 10.1515/ctra-2017-0013

Sawyer, R. K. (2011). *Explaining creativity: The science of human innovation* (Oxford university press) Sawyer, R. K. (2021). The iterative and improvisational nature of the creative process. *J. Creat.* 31, 100002. doi: 10.1016/j.yjoc.2021.100002

Sawyer, R. K. (2022). The dialogue of creativity: Teaching the creative process by animating student work as a collaborating creative agent. *Cogn. Instruct.* 40, 459–487. doi: 10.1080/07370008.2021.1958219

Sedova, K., Sedlacek, M., Svaricek, R., Majcik, M., Navratilova, J., Drexlerova, A., et al. (2019). Do those who talk more learn more? the relationship between student classroom talk and student achievement. *Learn. Instruct.* 63, 101217. doi: 10.1016/j.learninstruc.2019.101217

Shrivastava, D., Ahmed, C. G. S., Laha, A., and Sankaranarayanan, K. (2017). A machine learning approach for evaluating creative artifacts. *ArXiv abs/1707.05499.*

Simpson, E., Do Dinh, E.-L., Miller, T., and Gurevych, I. (2019). "Predicting humorousness and metaphor novelty with gaussian process preference learning," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 5716–5728. doi: 10.18653/v1/P19-1572

Song, K., Tan, X., Qin, T., Lu, J., and Liu, T.-Y. (2020). Mpnet: Masked and permuted pre-training for language understanding. *Adv. Neural Inf. Proc. Syst.* 33, 16857–16867. doi: 10.48550/arXiv.2004.09297

Stella, M., and Kenett, Y. N. (2019). Viability in multiplex lexical networks and machine learning characterizes human creativity. *Big Data Cogn. Comput.* 3, 45. doi: 10.3390/bdcc3030045

Sung, Y.-T., Cheng, H.-H., Tseng, H.-C., Chang, K.-E., and Lin, S.-Y. (2022). "Construction and validation of a computerized creativity assessment tool with automated scoring based on deep-learning techniques," in *Psychology of Aesthetics, Creativity, and the Arts.* doi: 10.1037/aca0000450

Toma, J. D. (2011). "Approaching rigor in applied qualitative," in *The SAGE Handbook for Research in Education: Pursuing Ideas as the Keystone of Exemplary Inquiry* 263–281. doi: 10.4135/9781483351377.n17

Torrance, E. P. (2008). *The Torrance Tests of Creative Thinking Norms—Technical Manual Figural (Streamlined) Forms a and b. 1998.* Bensenville, IL: Scholastic Testing Service.

Tuzcu, A. (2021). The impact of google translate on creativity in writing activities. *Lang. Educ. Technol.* 1, 40–52.

Vartanian, O., Smith, I., Lam, T. K., King, K., Lam, Q., and Beatty, E. L. (2020). The relationship between methods of scoring the alternate uses task and the neural correlates of divergent thinking: Evidence from voxel-based morphometry. *NeuroImage* 223, 117325. doi: 10.1016/j.neuroimage.2020.117325

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems* 30.

Vo, H., and Asojo, A. (2018). Feedback responsiveness and students' creativity. *Acad. Exch. Quart.* 1, 53–57.

Wagire, A. A., Rathore, A., and Jain, R. (2020). Analysis and synthesis of industry 4.0 research landscape: Using latent semantic analysis approach. *J. Manuf. Technol. Manag.* 31, 31–51. doi: 10.1108/JMTM-10-2018-0349

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., et al. (2019). "Superglue: A stickier benchmark for general-purpose language understanding systems," in *Advances in neural Information Processing Systems* 32.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. R. (2018). Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461.*

Williams, F. (1980). *Creativity Assessment Packet (CAP).* Buffalo, NY: D. O. K. Publishers Inc.

Yang, Z., Zhu, C., and Chen, W. (2018). Parameter-free sentence embedding via orthogonal basis. *arXiv preprint arXiv:1810.00438.*

Zhang, W., Sjoerds, Z., and Hommel, B. (2020). Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage* 210, 116572. doi: 10.1016/j.neuroimage.2020.116572

Zuñiga, D., Amido, T., and Camargo, J. (2017). "Communications in computer and information science," in *Colombian Conference on Computing* (Cham: Springer).

*CORRESPONDENCE
Ying Wang
✉ wangying0635@163.com

Enhancing English reading skills and self-regulated learning through online collaborative flipped classroom: a comparative study

Ying Wang*

Department of Foreign Language, Liaocheng University Dongchang College, Liaocheng, Shandong, China

Introduction: This research investigates the effectiveness of an online collaborative flipped classroom approach in enhancing English reading skills and self-regulated learning among Chinese English learners.

Methods: A total of 71 participants were divided into three instructional groups: traditional instruction (TI) group ($n = 24$), flipped instruction (FI) group ($n = 22$), and online flipped instruction (OFI) group ($n = 25$). The participants' reading comprehension ability was assessed using the reading section of the IELTS exam. Self-Regulated Learning (SRL) strategy use was evaluated using a questionnaire, and weekly online quizzes assessed participants' understanding of course materials. Online learning behaviors were examined by considering online log-on times. The instruction period lasted for 12 weeks, with pre-tests and post-tests conducted to measure progress.

Results: The results indicated that both the FI and OFI groups outperformed the TI group in terms of reading comprehension and self-regulated learning. Furthermore, the OFI students demonstrated superior online learning behaviors and objective performances compared to the FI students.

Discussion: These findings suggest that the integration of flipped and online instruction methods holds promise for improving English reading skills and enhancing self-regulated learning among Chinese English learners.

KEYWORDS

online collaborative flipped classroom, English reading skills, self-regulated learning, online learning behaviors, EFL students

Introduction

Online learning has emerged as an adaptable, accessible, and efficient avenue for second language (L2) acquisition, enabling learners to assume an active role in their language learning journey (Lin et al., 2017; Basilaia and Kvavadze, 2020; Subedi et al., 2020; Pokhrel and Chhetri, 2021). In fact, the surge of online learning and digitalized education has sparked a transformative shift in the educational landscape, ushering in an era of digital transformation within this domain (Fisher, 2006; García-Morales et al., 2021; Zarifis and Efthymiou, 2022; Fathi et al., 2023; Widayanti and Meria, 2023). Amidst this evolving educational paradigm, a prominent innovative strategy has garnered widespread recognition for its student-centered ethos–the flipped

classroom. This approach integrates various pedagogical elements, encompassing cooperative and collaborative learning, peer-based interactions, problem-solving techniques, and dynamic learning methods, effectively crafting an engaging online learning milieu (Slavin, 1991; Topping and Ehly, 1998; Michael, 2006; Bergmann and Sams, 2014; Hung, 2017; Fathi et al., 2021). Through this multifaceted approach, the flipped classroom leverages technology to optimize learning experiences and elevate learner engagement.

Flipped classrooms, a pedagogical approach gaining prominence in the field of English as a foreign language (EFL) instruction, are characterized by a fundamental shift in the traditional classroom paradigm. In the EFL context, a flipped classroom model involves the strategic use of digital resources to invert the conventional sequence of in-class instruction and out-of-class learning activities. This instructional approach redefines the roles of both educators and learners, fostering a more learner-centered and interactive environment. In a typical flipped EFL classroom, educators curate and deliver digital content, often in the form of video lessons, online modules, or multimedia materials, which cover core course topics and language skills. Students are provided access to these resources prior to in-person class sessions (Chen Hsieh et al., 2017). This pre-class exposure to content equips learners with foundational knowledge, enabling them to arrive at in-class sessions prepared and ready for interactive engagement. During in-person class time, the focus shifts from traditional lecturing to collaborative and application-based activities. Educators facilitate discussions, problem-solving exercises, and hands-on language tasks that encourage active participation and deeper comprehension. This approach capitalizes on the concept of "homework in the classroom" and "classwork at home, "allowing students to clarify doubts, seek clarification, and engage in peer-to-peer learning under the guidance of their instructors (Mehring, 2016).

While the flipped classroom has been recognized as a beneficial approach for enhancing EFL learners' linguistic competence (O'Flaherty and Phillips, 2015; Shih and Huang, 2020; Turan and Akdag-Cimen, 2020), limited research exists on its impact on other variables, particularly reading comprehension–an essential skill for academic knowledge acquisition–and self-regulated learning strategies–an important tool for independent language learning (Vitta and Al-Hoorie, 2020; Fathi and Rahimi, 2022). Research suggests that an online flipped classroom model can enhance reading comprehension (Karimi and Hamzavi, 2017; Samiei and Ebadi, 2021). Reading comprehension involves constructing meaning by connecting background knowledge with textual information, making it a vital component of academic contexts (Yapp et al., 2021). Learners need to develop the ability to read independently, even in online or home settings, by engaging with texts at the word, sentence, and text levels, seeking feedback from peers through discussions, accessing resources, and reflecting on their reading practices (Jeon and Yamashita, 2014).

Flipped classrooms, a pedagogical innovation gaining traction, often leverage homework assignments focused on reading materials to cultivate learners' autonomy, motivation, and a positive attitude toward advancing reading comprehension skills (Fulgueras and Bautista, 2020). Additionally, an intrinsic link exists between the flipped classroom model and the cultivation of self-regulated learning (SRL) strategies. Within this paradigm, learners are entrusted with the responsibility of not only acquiring and organizing information but also actively engaging in processes such as monitoring, reflection, and evaluation of their own learning practices (Lai and Hwang, 2016;

Theobald, 2021). Crucially, flipped EFL classrooms prioritize the learner's autonomy and self-regulated learning. Learners are encouraged to take ownership of their learning process, make informed decisions regarding their study pace, and employ self-regulation strategies to enhance their language acquisition process (Lai and Hwang, 2016). This instructional method harnesses technology to create a dynamic and adaptable learning ecosystem that empowers students to assume agency over their language learning.

The inclusion of self-regulated learning as a variable of interest alongside reading comprehension is guided by the understanding that these two aspects might be intertwined in a symbiotic relationship. Self-regulated learning encompasses a spectrum of behaviors, motivations, and metacognitive functions, all of which converge as students plan learning tasks, set attainable goals, track their progress, and engage in thoughtful reflection on their learning journey (Nilson, 2023). The strategic employment of self-regulated learning strategies, especially within the context of online collaborative flipped classrooms, is postulated to synergistically enhance reading comprehension abilities. This study endeavors to unravel the interplay between self-regulated learning and reading comprehension, shedding light on how the deliberate cultivation of metacognitive strategies through the flipped classroom model can potentially influence learners' abilities to comprehend and engage with English text.

Despite separate investigations on the flipped classroom, reading comprehension, and self-regulated learning strategies, further research is needed to understand how online flipped classrooms can influence reading comprehension and self-regulated strategies. To fill this research gap, this study aims to compare the effects of online flipped instruction and traditional flipped instruction on L2 reading comprehension and self-regulated learning among Chinese EFL learners. Additionally, the study seeks to explore differences in online learning behaviors between the two instructional groups. By saturating and confirming the existing literature and generating context-based findings, this study contributes to the expanding body of research on the effectiveness of the flipped classroom model in L2 instruction. Moreover, it provides valuable insights into the impact of flipped instruction on self-regulated learning in an online setting. The findings of this study may have practical implications for language teachers and curriculum developers interested in incorporating the flipped classroom model into their language instruction.

Literature review

Flipped classroom

The numerous contributions of digital learning to motivate students and make students active language learners were due to its approachability, convenience, collaboration, and proximity of digital devices that could enhance autonomy and add variations to the learning process (Prensky, 2005; Murdock and Williams, 2011; Zarifis and Efthymiou, 2022). The same was advocated in Asian countries since students widely used technological features to communicate through text, video calls, and other features that could help them interact and engage (Sweeny, 2010). Enhancing student-centered approaches in the online environment, teachers can employ flipped classroom model (FCM) to change traditional class activities. FCM brings rich chances for learners, adds flexibility and adaptability

(Bergmann and Sams, 2014; Shih and Huang, 2020), and offers practical tasks during class. Insights have arisen from diverse fields, spanning social sciences (Wanner and Palmer, 2015; Lee and Wallace, 2018), engineering (Karabulut-Ilgu et al., 2018), and education (Zainuddin and Attaran, 2016; Sommer and Ritzhaupt, 2018), all of which increasingly advocate the effectiveness of the flipped classroom in enhancing learners' educational outcomes (Çakıroğlu and Öztürk, 2017; Liu et al., 2019). Furthermore, several defining characteristics have been proposed for the flipped classroom, including interactive learning (Crouch et al., 2007), real-time engagement (Novak, 2011), inverted instruction (Davis, 2013), and the flipped learning model (Bergmann and Sams, 2014).

Participating in online flipped classrooms might empower EFL learners to cultivate autonomy in their decision-making and actions, fostering a sense of ownership and control over their reading experiences (Mehring, 2016; Fulgueras and Bautista, 2020). This newfound autonomy motivates learners to proactively adapt and refine their reading strategies to meet the demands of comprehension. Moreover, learners develop a positive attitude toward the challenges encountered during the reading process, embracing them as opportunities for growth and deeper understanding (Jia et al., 2023). Furthermore, online flipped classrooms offer EFL learners avenues for improving their vocabulary and grammatical knowledge (Turan and Akdag-Cimen, 2020). Prior to, during, or after reading a text, learners can leverage various techniques to enhance their language proficiency (Jiang et al., 2022). These include consulting dictionaries to clarify unfamiliar words, utilizing contextual clues to predict and deduce meanings, engaging in discussions with peers to elicit insights, and employing effective organizational strategies such as rehearsal, rereading, and summarization (Mohammaddokht and Fathi, 2022). By employing a range of cognitive and metacognitive strategies, learners optimize their reading experience and foster a deeper understanding of the text (Kintsch, 2012; Fischer and Yang, 2022). Through the interactive and collaborative nature of online flipped classrooms, EFL learners engage in a multifaceted approach to reading. They not only improve their linguistic competencies but also develop critical thinking skills, cultivate effective study habits, and foster a reflective stance toward their reading practices (Fulgueras and Bautista, 2020; Samiei and Ebadi, 2021). This comprehensive approach to reading instruction nurtures learners' confidence, self-efficacy, and motivation, positioning them for success in their language learning journey.

The flipped classroom can indirectly use technology, mobile, and computers outside of the classroom by watching videos of lectures, working with multimedia with other peers to gain knowledge and information (Kiernan and Aizawa, 2004; Stockwell, 2013; Amer, 2014), and have more considerable engagement in problem-solving, knowledge-sharing, and information-exchange communicative activities which are meaningful, along with personalized feedback in the classroom (Kim et al., 2017; Zarrinabadi and Ebrahimi, 2018). The benefits of experiencing flipped classroom can be positive perceptions of being actively involved, having more engagement, enhancing autonomy and critical thinking (Critz and Knight, 2013), and reaching greater achievement (Butt, 2014). The empirical findings also add novel findings to the literature. Accordingly, after investigating 66 pre-service English language teachers, Gok et al. (2021) found that there could be a considerable decline in FL classroom anxiety and reading anxiety during the flipped classroom. In another study, Jiang

et al. (2021) revealed that learners' demeanor and others' assistance could moderate the significance of preparation to be motivated and involved in an online flipped classroom.

Numerous investigations have explored the influence of flipped classroom methodologies on the reading proficiencies of EFL students and associated variables. In their study, Mohammaddokht and Fathi (2022) noted that flipped instruction produced substantial enhancements in EFL reading capabilities while concurrently alleviating reading-related apprehension. These findings imply the potential utility of flipped instruction in the context of EFL reading courses. Correspondingly, Fulgueras and Bautista (2020) scrutinized the repercussions of flipped classrooms on the development of critical thinking abilities and reading comprehension among senior high school ESL learners in the Philippines. The outcomes indicated advancements in critical thinking and reading comprehension proficiencies for both the flipped and conventional lecture-discussion pedagogies. However, the flipped learning approach unequivocally outperformed its traditional counterpart, underscoring its effectiveness in fortifying these competencies.

Examining student viewpoints on the implementation of flipped classrooms in EFL reading classes during the Covid-19 pandemic, Nursyahdiyah et al. (2022) conducted a case study that unveiled the efficacy of the flipped classroom strategy in enhancing the caliber of EFL learning. Furthermore, it fostered greater autonomy among students in their learning endeavors and positively influenced the role of technology in the realm of education. Yulian's research (2021) established that the adoption of the flipped classroom paradigm led to enhancements in critical thinking skills pertinent to critical reading. Students expressed favorable perceptions of this approach, placing emphasis on self-guided learning as a principal advantage. Likewise, Maharsi et al. (2021) scrutinized the integration of the flipped classroom approach within an EFL private university in Indonesia. The findings underscored that conventional classrooms exhibited augmented post-test scores in comparison to their flipped classroom counterparts, potentially attributed to teacher-centric instructional methods and technology-related variables. Nevertheless, a significant portion of students perceived flipped classrooms as catalysts for self-reliant and dynamic learning experiences, with recognition of both the merits and demerits associated with this approach.

In addition, Li et al. (2022) delved into the repercussions of employing the flipped classroom paradigm within the sphere of EFL instruction. Their inquiry strategically probed the manner in which the flipped methodology can augment the acquisition of students' communicative competence. The outcomes of their investigation cast a revealing light upon the potential advantages associated with integrating the flipped pedagogical approach into the domain of EFL instruction, thus furnishing insights into the realms of inventive language learning methodologies. In an exploration conducted by Liu et al. (2022), salient revelations emerged regarding the efficacy of the flipped framework in amplifying both writing prowess and the utilization of metacognitive strategies. Their inquiry makes a notable contribution to the comprehension of how the flipped classroom configuration can positively influence not only writing proficiency but also the maturation of metacognitive faculties in the context of collaborative writing. Similarly, Shih and Huang (2020) centered their inquiry on the adept application of metacognitive strategies among college students in an EFL flipped classroom milieu. Through an intricate analysis of students' deliberate utilization of metacognitive

strategies, their research enriches the understanding of the intricate dynamics underlying students' cognitive processes and strategic approaches within the flipped learning milieu.

Su Ping et al. (2020) also embarked on a scholarly exploration of the trajectory undertaken by EFL students within the framework of a flipped classroom, with a specific focus on a writing-intensive class. Hailing from the educational landscape of Malaysia, their research presents a distinctive vantage point that brings to light the array of experiences and outcomes that unfold for EFL learners immersed in a writing-centered flipped classroom setting. Engaging in a methodical appraisal, Turan and Akdag-Cimen (2020) executed a comprehensive dissection of the implementation of the flipped classroom methodology in the context of English language instruction. Their amalgamation of research findings furnishes substantial revelations into the overarching efficacy and repercussions of the flipped pedagogical approach in the domain of language learning. This synthesis substantially enriches the broader comprehension of the multi-faceted dimensions inherent in the implementation of the flipped approach. Delving into the intersection of pedagogy and technology, Jiang et al. (2021) undertake an investigative expedition into the amalgamation of automatic speech recognition technology within the structure of a flipped classroom. Their inquiry unveils the latent potential of technology to elevate the complexity of EFL learners' oral language capabilities, effectively interweaving modern technological advancements with the foundations of the flipped learning milieu. In another study, Karjanto and Simon (2019) explored the application of the flipped classroom methodology in a Calculus course situated within a cultural context influenced by Confucian heritage. They designed a theoretical framework that integrated elements such as Bloom's taxonomy, English-medium instruction, and the incorporation of technology. Their instructional design encompassed four distinct approaches, including variations of the flipped classroom model. The quantitative analysis yielded notable findings, revealing a significant discrepancy in examination scores, particularly evident when comparing fully-flipped instruction to single-topic flipped instruction. Furthermore, their qualitative investigations underscored positive enhancements in student engagement and interactions with instructors. Nevertheless, they also unearthed challenges linked to language, cultural factors, competition, and the adaptation to technological tools.

Taken together, these studies contribute to the literature by examining the efficacy of the flipped classroom approach in enhancing various dimensions of EFL learning, including communication, writing, metacognition, critical thinking, and oral language proficiency. Their insights resonate with the evolving landscape of language education, offering valuable guidance for educators seeking to embrace innovative pedagogical strategies to meet the diverse needs of language learners.

Reading comprehension

Reading comprehension is a multifaceted skill that involves various cognitive processes and strategies. According to Kintsch (2012), reading comprehension entails the ability to connect existing knowledge (schema) with the information presented in the text, summarize key elements, draw appropriate conclusions, and enhance understanding by posing probing questions. It encompasses the process of constructing meaning from written texts, ranging from recognizing individual symbols and linguistic units to synthesizing and integrating information within a meaningful framework, thereby engaging higher-order thinking skills (Kendeou et al., 2014; Zhang and Zhang, 2022).

Comprehending a written text is a complex cognitive activity that relies on several interconnected factors. It necessitates the activation of prior knowledge, fluency in reading, relevant past experiences, the utilization of cognitive and metacognitive strategies, a strong grasp of lexical and grammatical knowledge, the ability to organize information, make judgments, and engage in reflective evaluation (Syatriana, 2011). Consequently, reading comprehension is recognized as a challenging skill in internationally recognized tests such as the International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) (Pellegrino and Hilton, 2012; Hung, 2015). Furthermore, several other variables can influence learners' reading comprehension abilities. These variables encompass reading types, individuals' attitudes toward reading, the methods employed during reading activities, adaptability to different text genres, and the strategies utilized by learners to comprehend the texts effectively (Jeon and Yamashita, 2014; Zhang and Zhang, 2022). These factors interact and contribute to learners' overall reading comprehension performance (Jeon and Yamashita, 2014). Given the complexity of reading comprehension and its significance in academic and language proficiency assessments, it is imperative to further explore various instructional procedures and their impact on learners' comprehension abilities (Yapp et al., 2021). Also, understanding the variables that affect reading comprehension can inform instructional practices, curriculum design, and the development of effective strategies to enhance learners' reading skills.

Reading comprehension, a challenging process that contains components, procedures, and aspects with the desire to discover great ways of accelerating it, is an integrated process of generating meanings from a reading section (Meniado, 2016). Besides, there appeared several ways to improve EFL learners' reading comprehension, for instance, by incorporating online flipped classrooms, as supported in the literature (Öztürk and Çakıroğlu, 2021; Samiei and Ebadi, 2021; Fischer and Yang, 2022; Hasan et al., 2022; Mohammaddokht and Fathi, 2022). According to Samiei and Ebadi (2021), WebQuest-based flipped classroom significantly enhances learners' inferential reading comprehension as revealed via the data analysis. In a similar study, Hashemifardnia et al. (2018) examined how flipped classroom affects junior high school students' reading comprehension in EFL context. They stated that online flipped classrooms could substantially affect reading comprehension. Although the significance of flipped classrooms in enhancing reading comprehension has received limited exploration, the objective of this study is to contribute to the existing literature by investigating the impact of online flipped classrooms on the reading comprehension of EFL learners.

Self-regulated learning

In educational psychology, Self-Regulated Learning (SRL) stands as a foundational construct, intrinsically linked with Zimmerman's theoretical framework (Zimmerman, 2000). Zimmerman posits that

human regulatory skill, or the lack thereof, holds a pivotal role in shaping our perception of personal agency, which, in turn, forms the very core of our self-concept (Zimmerman, 2000; Zimmerman and Schunk, 2001). The development of this regulatory capability, encompassing its subcomponents and functional aspects, has remained a central focus of social cognitive theory and research (Zimmerman, 2000; Zimmerman and Schunk, 2001).

Zimmerman's comprehensive framework extends its purview to elucidate common dysfunctions observed in self-regulatory functioning, including phenomena such as biased self-monitoring, self-blaming judgments, and defensive self-reactions (Zimmerman, 2000). In seeking to provide a holistic perspective on self-regulation, Zimmerman's framework addresses a multitude of facets. These include delving into the structural elements of self-regulatory systems, discerning the influences of social and physical environmental contexts on self-regulation, investigating dysfunctions that may arise within the realm of self-regulation, and exploring the developmental trajectory of self-regulation (Zimmerman, 2000).

Within the realm of SRL, learners engage in a multifaceted set of strategies that empower them to meticulously plan, closely monitor, and critically evaluate their learning activities (Zimmerman and Schunk, 2001). These strategies, deeply ingrained in Zimmerman's model, serve as the scaffolding upon which learners construct their self-regulated learning processes. They assume control over their learning endeavors, establish meaningful goals, evaluate their progress, and adapt their strategies judiciously to optimize learning outcomes (Zimmerman and Schunk, 2001).

These SRL strategies seamlessly align with three distinct phases: Planning, Monitoring, and Evaluating. In the *Planning* phase, learners undertake activities that lay the groundwork for effective reading comprehension. This phase encompasses actions such as previewing reading tasks, setting clear learning objectives, and formulating goals before immersing themselves in the reading materials.

The *Monitoring* phase, on the other hand, hinges on learners' adeptness at overseeing their reading progress and performance. Strategies like self-checking comprehension, identifying challenging sections, and making real-time adjustments during the reading process epitomize this phase.

Lastly, the *Evaluating* phase revolves around the critical assessment of the efficacy of one's learning strategies and the attainment of learning objectives. In this phase, learners engage in reflection on their reading experiences, conduct a thorough analysis of the success of their approaches, and contemplate adjustments for future learning endeavors.

In the specific context of English reading comprehension, the application of SRL strategies assumes paramount importance. Learners can substantially enhance their reading skills by proactively employing SRL strategies that seamlessly align with Zimmerman's model. These strategies, which traverse the planning, monitoring, and evaluating phases, enable learners to not only navigate the intricate landscape of reading comprehension effectively but also become architects of their own learning experiences. In the planning phase, learners prelude their reading journeys by engaging in activities such as previewing reading tasks and crystallizing their learning objectives. Subsequently, the monitoring phase calls for ongoing self-assessment and vigilant tracking of progress during the reading process. Finally, the evaluating phase encourages learners to engage in a comprehensive assessment of their comprehension, scrutinize the

efficacy of their chosen strategies, and craft a roadmap for continued learning success.

Self-regulated learning strategies hold significant importance in empowering students to assume agency over their learning process and actively participate in their educational journey (Zimmerman, 2002). Self-regulation encompasses the development of a practical understanding of one's own abilities, enabling students to make informed decisions and take appropriate actions to enhance their learning experiences (Zimmerman, 2000; Pajares, 2009). Students who possess a high level of self-regulation acquire the capacity to exert control over their learning processes, actively constructing meaning, establishing goals, making deliberate choices regarding the strategies they employ, and assuming leadership in directing their own learning (Zimmerman and Schunk, 2001; Pintrich, 2004). Moreover, they effectively integrate contextual and personal factors into their learning experiences, recognizing the interplay between these elements.

The utilization of self-regulated learning strategies results in increased engagement and proactive learning among students (Nilson, 2023). They develop the ability to monitor their progress, adapt their learning strategies as needed, and demonstrate perseverance in the face of challenges. This acquisition of self-regulation empowers students to become autonomous learners, capable of adjusting their approaches to different learning tasks and contexts, thereby enhancing the effectiveness of their learning outcomes (Theobald, 2021). Furthermore, self-regulated learners display metacognitive awareness, engaging in reflection on their learning processes, evaluation of their performance, and identification of areas for improvement (Andrade and Evans, 2012). Through self-reflection, self-evaluation, and self-assessment, they continually refine their learning strategies, optimizing their overall learning outcomes.

Blended teaching can make students autonomous in their language learning process by planning to learn, having more pace for selecting and sequencing the video- and audio-based content, possessing ownership, making decisions, enhancing higher-order learning skills, and observing learning to support self-regulated learning strategies (Lai and Hwang, 2016; Tan et al., 2017; Van Laer and Elen, 2017; Lee and Choi, 2019; Shih and Huang, 2020; Fathi et al., 2023). Moreover, an online flipped classroom, as a blended learning strategy, can offer authentic, meaningful, and personal materials, offer learners control and provide sufficient scaffolding and opportunities for interaction, reflection, and cooperation (Van Laer and Elen, 2017).

Within the context of online flipped classrooms, students are called upon to employ a spectrum of self-regulated learning strategies, encompassing cognitive, metacognitive, and behavioral dimensions, to effectively navigate their pre-class tasks and subsequently participate in in-class sessions (Geduld, 2016). The cognitive facet entails strategies such as rehearsing, organizing, transforming, and expanding knowledge, while metacognitive strategies involve goal setting and performance monitoring. Additionally, behavioral aspects encompass time and resource management as well as note-taking practices (Karlen, 2016). These multifaceted strategies converge to form a cohesive skill set crucial for successful EFL learning within the dynamic landscape of online flipped classrooms.

Our focus on the interaction between flipped classrooms and self-regulated learning not only acknowledges the evolving demands placed upon learners but also sheds light on the symbiotic

relationship between instructional methodology and cognitive autonomy. While prior research has indeed identified predictive links between self-regulated strategies and flipped classrooms (Van Alten et al., 2020; Öztürk and Çakıroğlu, 2021), our study seeks to extend this understanding by validating and strengthening these insights within a distinct educational setting. To this end, this study aims to examine how online flipped classrooms can influence EFL learners' reading comprehension and self-regulated learning strategies quantitatively and explore its role in students' perception qualitatively.

The present study

Although previous studies have explored the effectiveness of flipped instruction and online learning in improving language skills, there remains a need for more empirical research, particularly in the form of comparative studies that directly compare different instructional methods. This current study aims to address this research gap by investigating the effectiveness of three teaching methods–Online Flipped Instruction (OFI), Flipped Instruction (FI), and Traditional Instruction (TI)–in enhancing L2 reading comprehension performance and self-regulation among students.

Through the examination of online collaborative flipped instruction in the OFI group, traditional flipped instruction in the FI group, and conventional instruction in the TI group, this study will assess the impact of each method on students' L2 reading comprehension performance and self-regulation. Via directly comparing these three approaches, the study will generate empirical evidence to identify the most effective teaching method for enhancing students' language learning outcomes. Against this backdrop, this study aims to answer the following research questions:

1. What are the comparative effects of the OFI, FI, and TI methods on L2 reading comprehension performance?
2. What are the comparative effects of the OFI, FI, and TI methods on L2 reading self-regulation?
3. Are there significant differences in online learning behaviors and objective performances between students in the OFI and FI groups?

By answering these research questions, this study aims to contribute to the existing literature by providing empirical evidence on the comparative effectiveness of different instructional methods in promoting L2 reading comprehension and self-regulated learning of Chinese EFL learners. The rationale behind the emphasis on this specific student cohort (i.e., EFL learners) is rooted in the recognition that diverse factors, including cultural, linguistic, and educational backgrounds, can shape the implementation and outcomes of instructional methodologies. The intricacies of English language acquisition for Chinese learners (Wu, 2001), coupled with the demands of academic reading skills, render this population particularly intriguing for investigation. By delving into the experiences and responses of Chinese English learners within the realm of online collaborative flipped classrooms, this study aims to unearth insights that can inform tailored pedagogical strategies, curriculum design, and instructional support.

Method

Participants

The participants in this research were 71 EFL students, aged between 18 to 30 years old, who were enrolled in an English language course at a large language institute in mainland China. The majority of the participants were females ($n = 45$, 63.4%) and the rest of the students were males ($n = 26$, 36.6%). The participants had varying educational backgrounds, with most of them holding a high school diploma or equivalent ($n = 54$, 76.1%) and the remaining participants had a bachelor's degree ($n = 17$, 23.9%). Participants' proficiency level was determined based on the standardized placement test, Test of English for International Communication (TOEIC; Woodford, 1982), which is widely employed to evaluate the English proficiency of non-native speakers. Participants with an intermediate proficiency level (score range between 550 and 780) were included in the study. The participants were divided into three classes who were randomly assigned to a traditional instruction (TI) group ($n = 24$), and two experimental groups: Flipped Instruction (FI) group ($n = 22$) and Online Flipped Instruction (OFI) group ($n = 25$).

While addressing potential concerns regarding the sample size, insights were drawn from the recommendations of American Council on the Teaching of Foreign Languages (2010), which suggest that an optimal class size of approximately 15 students is advisable for facilitating collaborative learning activities effectively, especially within student-centered educational contexts. However, it's worth noting that collaborative learning can still be implemented with larger class sizes, such as the 22 students in our study. To accommodate the larger class size while adhering to the principles of collaborative learning, we strategically designed the flipped instruction and online flipped instruction approaches, ensuring that they were conducive to group interactions and active engagement. This approach aimed to maintain the quality and effectiveness of collaborative learning, even with a larger number of students, thereby enhancing the reliability and validity of the research findings.

Measures

Reading comprehension

The participants' reading comprehension ability was measured using the IELTS reading test (University of Cambridge ESOL Examinations, 2011). The IELTS reading test consists of three sections, each containing one long reading passage with increasing difficulty, followed by a set of multiple-choice questions. The IELTS Academic Reading test is a standardized assessment that consists of 40 questions and is administered within a strict time limit of 60 min. The test aims to assess the participants' ability to comprehend and analyze academic English texts. The IELTS reading test has been extensively employed in L2 research, and has demonstrated good reliability and validity in measuring reading comprehension ability.

According to Weir and O'Sullivan (2017), the post-1989 evolution of IELTS primarily involves the transformation of the initial ELTS into

a legitimate, psychometrically sound, high-stakes assessment with the capacity for widespread global administration on a large scale. Moreover, current data from the IELTS website (IELTS, 2021) provides comprehensive statistics for the test forms administered in 2019. Specifically, for the Reading section, the reported reliability coefficient stands at a robust value of 0.92, within a confidence interval of 0.90 to 0.93.

Self-regulated learning strategy use questionnaire

The Self-Regulated Learning (SRL) questionnaire utilized in this study was developed by Tse et al. (2022) and was based on Zimmerman's (2000) cyclical phases model. The questionnaire comprised 13 statements that assessed the employment of SRL strategies in English reading. The statements were formulated according to three categories of SRL strategies: planning, monitoring, and evaluating. Planning involved activities such as previewing reading tasks and setting goals prior to reading, whereas monitoring referred to checking and monitoring one's reading progress and performance. Evaluating concerned the assessment of learning outcomes and the effectiveness of strategies. Participants rated all items using a 4-point Likert scale ranging from 1 (never or almost never) to 4 (every day or almost every day). The SRL strategy use questionnaire exhibited robust internal consistency, attaining a commendable coefficient of reliability at $r = 0.86$ within the context of this study.

Online quizzes

To thoroughly assess the participants' engagement with the course materials and their understanding of the content, a series of online quizzes were administered on a weekly basis to both the OFI and FI groups. These quizzes were thoughtfully designed to serve multiple purposes: to aid participants in their preparation for the subsequent in-class activities, to reinforce their comprehension of the assigned video lessons, and to evaluate their grasp of the core concepts relevant to the course.

Each weekly quiz comprised a carefully curated set of questions, ranging from 15 to 25 in number. These questions encompassed a variety of formats, including multiple-choice and short-answer questions, and were meticulously aligned with the main topics covered in the video lessons. Importantly, the questions were intentionally tailored to bridge the gap between the video content and the overarching theme of English reading comprehension. By incorporating elements of reading analysis, interpretation, and application, these quizzes aimed to foster a deeper understanding of the materials presented in the videos and to promote critical thinking in the context of L2 reading.

Throughout the 12-week course duration, a total of 10 quizzes were administered. The quizzes were strategically spaced to correspond with the course's progression and to ensure that participants had the opportunity to revisit and consolidate their knowledge on a regular basis. The mean score obtained by each participant across these quizzes was computed as an additional metric of achievement. This mean score served to provide insights into the

participants' consistent performance and their evolving comprehension of the course materials. It is noteworthy that the quiz scores contributed significantly to the participants' final grades, reflecting their competence in grasping the course content. Specifically, the final grades were calculated based on a comprehensive evaluation framework, which included various components. These components encompassed the overall participation score (50%), comprising class attendance (20%), quiz scores (20%), and assignments (10%). Additionally, the final grades considered the midterm test score (20%) and the final test (post-test) score (30%). This multifaceted approach to assessment aimed to holistically gage the participants' progress, engagement, and achievement throughout the course.

Online learning behavior

Following Fischer and Yang (2022), the present study examined three distinct online learning behaviors, encompassing regular online log-on time, group video-watching time, and total online log-on times. Each of these dimensions warrants attention and consideration within the context of our research. Firstly, regular online log-on time signifies the temporal commitment learners invest in engaging with the weekly assigned video lessons online. It reflects the extent to which students actively participate in the preparatory phase of flipped classrooms, where they access and assimilate instructional content independently before class. This dimension directly intersects with the flipped classroom model, as it measures the conscientiousness with which students approach their pre-class learning activities. Secondly, group video-watching time stands as a critical component of online learning behavior, capturing the duration during which participants in the experimental groups jointly engage in watching video lessons within small online groups. This dimension encapsulates the collaborative aspect of the flipped classroom approach, emphasizing the value of peer interaction and shared learning experiences. It is an essential component that furthers our understanding of how students engage with instructional materials and with each other, highlighting the interpersonal dimension of online learning. Lastly, total online log-on times amalgamate individual regular online log-on times with the time spent in small group video-watching sessions. This cumulative measure offers a comprehensive perspective on students' overall engagement with online learning materials and activities. It underscores the holistic nature of online learning behavior, recognizing that effective learning in the digital realm encompasses both independent and collaborative dimensions.

Importantly, these dimensions of online learning behavior closely align with the tenets of self-regulated learning. SRL involves learners taking charge of their learning processes, which includes planning, monitoring, and evaluating their learning activities. The temporal commitment demonstrated through regular online log-on times resonates with the planning phase of SRL, where learners proactively engage with course materials and set the stage for effective learning. Group video-watching time correlates with the monitoring phase, as it reflects learners' active involvement in tracking their progress through collaborative engagement. Lastly, the cumulative measure of total online log-on times speaks to the evaluation phase, wherein learners assess their learning strategies and the effectiveness of their collaborative endeavors.

Group video-watching time, on the other hand, exclusively captured the amount of time that the experimental group participants spent watching the video lessons in small online groups. Lastly, total online log-on times represented the cumulative measure, incorporating the individual regular online log-on times and the small group video-watching times of the participants.

To examine participants' online learning behaviors, weekly video lessons and accompanying quizzes were uploaded onto the LMS used in this study. Both groups (i.e., FI and OFI) participants were instructed to watch the assigned videos on the LMS and complete the quizzes before attending class. The LMS automatically tracked each student's log-on time for video viewing on a weekly basis. These log-on times were then aggregated to calculate the overall regular online log-on time for each participant throughout the course. Additionally, the students' online activities, including video-watching and collaborative quiz sessions, were recorded. The duration of each online group session was calculated, resulting in the group video-watching time. This analysis provided insights into the students' collective time spent watching the assigned videos and collaborating within the virtual environment, contributing to a comprehensive understanding of their online learning behaviors. To explore participants' online learning behaviors, weekly video lessons and accompanying quizzes were uploaded to the LMS employed in this study. Both the Flipped Instruction and Online OFI groups were directed to view the designated videos on the LMS and take the quizzes prior to coming to class. The LMS automatically recorded students' log-on times for video viewing on a weekly basis, which were then combined to determine their overall online log-on time throughout the course.

Furthermore, students' online engagement, encompassing video-watching and collaborative quiz sessions, was meticulously logged. The duration of each group session was calculated, resulting in the compilation of group video-watching time. This analysis furnished valuable insights into the collective time students invested in viewing the assigned videos and engaging in collaboration within the virtual learning environment. Consequently, it contributed to a comprehensive comprehension of their online learning behaviors.

Procedure

All three groups in the study were instructed by the same teacher, who had professional training in teaching English, specifically in academic reading skills. The course materials utilized in the study were focused on academic English reading skills, with particular emphasis on the learners' proficiency in the IELTS academic reading test. The course lasted for 12 weeks, during which time all three groups received 180 min of instruction per week, distributed over two 90-min sessions. Pre-tests and post-tests were administered during the first and last week of the course to measure the progress of the learners.

In the traditional instruction (TI) group, the instructor employed the traditional 'sage on the stage' method to deliver the course materials and facilitate group exercises and activities. The classroom sessions primarily involved teacher-led instruction, group exercises, and activities. Pre- and post-test assessments were conducted by the teacher. In the flipped instruction (FI) and online flipped instruction (OFI) groups, the teacher played a more dynamic role, both inside and

outside of the classroom. In the classroom, the instructor designed the curriculum, coordinated and supervised group activities, provided feedback and answered questions, as well as conducted pre- and post-test assessments. Beyond the confines of the traditional classroom setting, the educator meticulously curated and presented pre-designed video lessons, which were thoughtfully crafted to integrate reading comprehension exercises, textual analysis, and targeted reading strategies directly into the video content. Additionally, the teacher assumed responsibility for conceptualizing, producing, recording, and refining these tailored video lessons to ensure a comprehensive and effective approach to teaching reading skills and fostering self-regulated learning. Furthermore, the instructor efficiently oversaw the organization and administration of course materials within the educational institution's designated learning management system. To facilitate online interactions, the teacher harnessed the capabilities of the Zoom Webinar video meeting web application. Moreover, during the OFI group sessions conducted virtually, the instructor delivered periodic feedback to foster an optimal learning environment.

In the weekly in-class sessions, the instructor began by providing brief announcements and instructions, as well as answering students' questions. Then, for the remaining class time, the instructor divided students into groups and engaged them in various exercises and tasks to practice the weekly lesson content, which focused on reading skills and strategies. All learners were exposed to the identical in-class instruction. As part of their extracurricular learning, the FI students engaged in independent viewing of the designated course video lessons and subsequently undertook brief quizzes aligned with each instructional video.

In contrast to the other groups, the OFI students were purposefully grouped into small cohorts, consisting of no more than four individuals. These smaller groups fostered a conducive environment for collaborative learning as they collectively watched the assigned weekly video lesson online. Throughout these sessions, the OFI students actively engaged in collaborative discussions, exchanging ideas, and jointly completing the video quizzes, thereby reinforcing their understanding of the material as a unified entity. To facilitate seamless communication and interaction during their online video sessions, the small OFI groups effectively utilized Google Hangouts, leveraging its features for synchronous video conferencing and real-time collaboration. Consequently, their sessions were recorded and securely submitted to the instructor for meticulous review and insightful feedback.

It is noteworthy that the OFI students were explicitly informed of the significance of their weekly online video sessions. They were apprised that these sessions would be thoughtfully reviewed and deliberated upon during the subsequent in-person sessions held on a weekly basis. This ensured that the insights, discussions, and collaborative efforts from the online environment seamlessly integrated with the face-to-face instructional setting, fostering continuity and cohesion in the students' learning experiences. In contrast, the TI students followed a different approach during their weekly lessons. These sessions encompassed a combination of in-class instruction focused on delivering the course content and a variety of engaging activities designed to enhance student participation and comprehension. Commencing each session, the instructor dedicated the initial half to comprehensive coverage of the lesson materials, providing necessary explanations and clarifications to support student learning. Subsequently, the latter

TABLE 1 Main features of the reading intervention across experimental groups.

Instructional features	Traditional Instruction (TI)	Flipped Instruction (FI)	Online Flipped Instruction (OFI)
Teacher's Role	'Sage on the stage' method	Dynamic role	Dynamic role
Curriculum design	Yes	Yes	Yes
Group activities	Yes	Yes	Yes
Feedback and Q&A	Yes	Yes	Yes
Pre- and post-tests	Yes	Yes	Yes
Video lessons	No	Yes	Yes
Online quizzes	No	Yes	Yes
Specialized video lessons	No	Some	Yes
Learning management system	No	No	Yes
Zoom webinar	No	No	Yes
Collaborative online sessions	No	No	Yes
Google hangouts integration	No	No	Yes
Recorded sessions review	No	No	Yes

TABLE 2 Results of the paired-sample *t*-test for reading comprehension

	Pre-test		Post-test		*t*	*p*
	M	SD	*M*	SD		
TI	5.12	0.66	5.66	0.56	3.79**	0.005
FI	5.18	0.72	6.08	0.72	5.09***	0.000
OFI	4.99	0.83	6.41	0.69	6.74***	0.000

**$p < 0.01$, ***$p < 0.001$. TI, Traditional instruction; FI, Flipped Instruction; OFI, Online flipped instruction.

portion of the session was dedicated to a range of interactive group activities and tasks, specifically designed to align with the lesson topics and skills covered in the assigned readings. This format emulated the structure and format adopted in the weekly in-class sessions of the FI and OFI groups, ensuring consistency and alignment across the instructional approaches employed throughout the study (Table 1).

Prior to the commencement of the study, all participants, regardless of their assigned experimental group, underwent pre-testing to establish baseline measurements of their reading comprehension abilities and SRL strategies. The pre-tests were conducted during the initial week of the course. Subsequently, upon completion of the 12-week course, post-tests were administered during the final week to assess the participants' progress. The SRL strategy use questionnaire was administered at the beginning and end of the course to gage any

changes in self-regulated learning behaviors and strategies. These assessments were conducted in a controlled classroom environment to ensure consistent testing conditions for all participants. Participants were instructed to respond to the pre- and post-tests and the SRL surveys with their utmost attention and sincerity, as their responses played a crucial role in evaluating the effectiveness of the instructional approaches employed in this study. The test and survey data were collected and analyzed to provide valuable insights into the impact of the different instructional methods on reading comprehension and self-regulated learning outcomes.

Data analysis

To analyze the data collected for this study, the researchers used statistical software SPSS. The first research question was explored using a paired-sample t-test to compare the mean scores of the pre- and post-tests for the three groups (OFI, FI, and TI). Subsequently, to examine potential distinctions in the post-test scores among the groups, ANOVA was employed, supplemented by Fisher's LSD post-hoc analysis. Regarding the third research question, an independent t-test was conducted to explore possible disparities in the online learning behaviors and objective performances, specifically the average online quiz scores and final course grades, between the two experimental flipped groups (OFI and FI).

Results

The first research question aimed to investigate which teaching method–Traditional Instruction (TI), Flipped Instruction (FI), or Online Flipped Instruction (OFI)–yielded the most significant results in terms of the students' L2 reading comprehension performance. To address this question, a paired-sample t-test was conducted to compare the pre-test and post-test scores of the students in each group. As seen in Table 2, the results revealed that all three groups showed significant improvement in their reading comprehension performance from pre-test to post-test. The mean scores of the students in the FI and OFI groups increased from pre-test to post-test, while the mean score of the TI group slightly decreased.

The paired-sample t-test results showed that all three groups showed significant improvement in their reading comprehension performance from pre-test to post-test (TI: $t = 3.79$, $p = 0.005$; FI: $t = 5.09$, $p = 0.000$; OFI: $t = 6.74$, $p = 0.000$). The OFI group showed the highest improvement, followed by the FI group, and then the TI group.

To further examine which teaching method had the most significant effect on the students' reading comprehension performance, a one-way ANOVA was conducted. Table 3 presents the results of the ANOVA for reading comprehension. The results indicated a significant difference between the groups in terms of their reading comprehension performance [$F(2, 68) = 3.61$, $p = 0.034$].

Post hoc tests using the LSD method were conducted to determine which groups were significantly different from each other. Table 4 presents the results of the *post hoc* tests. The results showed that there was a significant difference in reading comprehension performance between the OFI and TI groups ($p = 0.008$), indicating that the OFI group had a significantly higher mean score than the TI group. Additionally, the results revealed a significant difference between the FI and OFI groups

TABLE 3 Results of ANOVA for reading comprehension.

	SS	df	MS	F	P
Between groups	3.02	2	1.51	3.61*	0.034
Within groups	21.32	68	0.43		
Total	24.35	70			

*$p < 0.05$.

TABLE 4 Results of *post hoc* LSD.

					95% CI	
		Mean difference	SE	*p*	Lower	Upper
(I)	(J)	(I–J)				
TI	FI	0.42	0.191	0.023	0.042	0.798
OFI	TI	0.75	0.189	0.008	0.378	1.122
OFI	FI	0.33	0.194	0.031	−0.048	0.708

TABLE 5 Results of the paired-sample t-test self-regulated learning.

	Pre-test		Post-test		*t*	*p*
	M	SD	M	SD		
TI	3.28	0.71	3.41	0.75	1.58	0.123
FI	3.21	0.89	3.93	0.83	4.30**	0.006
OFI	3.15	0.68	4.35	0.67	7.32***	0.000

**$p < 0.01$, ***$p < 0.001$.

TABLE 6 Results of ANOVA self-regulated learning in L2 reading.

	SS	df	MS	F	P
Between groups	4.22	2	2.11	5.67*	0.012
Within groups	36.75	68	0.64		
Total	40.97	70			

*$p < 0.05$.

($p = 0.031$), indicating that the OFI group had a slightly higher mean score than the FI group, although this difference was not as big as the difference between the OFI and TI groups. Finally, there was also a significant difference between the TI and FI groups ($p = 0.023$).

The second research question examined the differences in self-regulated learning strategies among students in the three teaching methods. Based on the results of the paired-sample t-test for self-regulated learning (Table 5), the two flipped teaching methods showed an increase in self-regulated learning from pre-test to post-test. However, the largest increase was observed in the OFI group ($M = 3.15$, SD = 0.68 to $M = 4.35$, SD = 0.67) with a significant t-value of 7.32 ($p < 0.001$). The FI group also showed a significant increase in self-regulated learning ($M = 3.21$, SD = 0.89 to $M = 3.93$, SD = 0.83) with a t-value of 4.30 ($p < 0.01$). The TI group, on the other hand, showed a non-significant increase in self-regulated learning ($M = 3.28$, SD = 0.71 to $M = 3.41$, SD = 0.75) with a t-value of 1.58 ($p = 0.123$).

Also, the results of the ANOVA for self-regulated learning in L2 reading (Table 6) indicate that there was a significant difference between the teaching methods in terms of their effect on self-regulated learning ($F = 5.67$, $p = 0.012$). The *post hoc* LSD analysis (Table 7) revealed that the OFI group had a significantly higher mean self-regulated learning score ($M = 4.35$, SE = 0.19) than both the FI group ($M = 3.93$, SE = 0.19) and the TI group ($M = 3.41$, SE = 0.19) with mean differences of 0.42 ($p = 0.035$) and 0.94 ($p < 0.001$), respectively. Additionally, the FI group had a significantly higher mean self-regulated learning score than the TI group with a mean difference of 0.52 ($p = 0.015$). Overall, it was found that both the FI and OFI teaching methods resulted in a greater increase in self-regulated learning in L2 reading compared to traditional instruction. Furthermore, the OFI method yielded the highest increase in self-regulated learning, suggesting that online flipped classrooms can be a beneficial approach for enhancing students' self-regulation in L2 reading.

The third research question addressed in this study was whether there were significant differences in the online learning behaviors and objective performances of OFI and FI students. Based on the results of the independent-samples *t*-tests (see Table 8) provided, there were significant differences in the total online log-on time, online quiz score, and final score between the FI and OFI students. The OFI students had a significantly higher total online log-on time ($M = 9.23$ h, SD = 3.27) than the FI students ($M = 2.96$ h, SD = 0.79), $t = −5.234$, $p < 0.001$. The OFI students also had a significantly higher online quiz score ($M = 89$, SD = 16.54) than the FI students ($M = 64$, SD = 14.23), $t = −4.751$, p < 0.001, and a significantly higher final score ($M = 88$, SD = 14.97) compared to the FI students ($M = 69$, SD = 11.29), $t = −3.272$, $p = 0.008$. These results suggest that the OFI students had better objective performances than the FI students in terms of online quiz and final score, and also spent more time online overall.

Discussions

The arrival of student-centered language education shed some light on the significance of implementing blended learning in EFL classes. One such developmental move was an online flipped classroom that could foster collaboration and cooperation, scaffolding (Topping and Ehly, 1998) and problem-solving (Barrows, 1996), and flexibility (Michael, 2006). Thus, the present study aimed to compare the effectiveness of traditional instruction (TI), flipped instruction (FI), and online flipped instruction (OFI) on the reading comprehension and self-regulated learning strategies of EFL learners. The findings of the study revealed that both FI and OFI instructional methods were more effective in improving students' reading comprehension scores compared to traditional instruction. Furthermore, the OFI method showed significant improvements in students' SRL strategy use compared to the other two groups. These results suggest that incorporating flipped and online flipped instruction into EFL reading instruction may enhance students' learning outcomes.

One of the noteworthy findings of this study was the positive impact of online flipped classrooms on the reading comprehension abilities of EFL learners. The success of this approach can be attributed to a range of effective strategies employed. These encompass assigning videos and reading tasks for completion outside the traditional

TABLE 7 Results of *post hoc* LSD.

(I)	(J)	Mean difference	SE	p	95% CI	
					Lower	Upper
(I)	(J)	(I–J)				
TI	FI	0.52	0.192	0.015	0.097	0.943
OFI	TI	0.94	0.191	0.000	0.516	1.364
OFI	FI	0.42	0.196	0.035	0.026	0.814

TABLE 8 Results of the independent-samples *t*-test.

	Group	M	SD	t	p
Total online log-on time (hours)	FI	2.96	0.79	−5.234***	0.000
	OFI	9.23	3.27		
Online quiz score	FI	64	14.23	−4.751***	0.000
	OFI	89	16.54		
Final score	FI	69	11.29	−3.272**	0.008
	OFI	88	14.97		

$**p < 0.01$, $***p < 0.001$.

classroom, facilitating collaborative reading tasks through online communication platforms, and fostering reflection and self-evaluation of reading comprehension skills. These strategies have been supported by prior research, demonstrating increased motivation, engagement, and productivity within the classroom (Strelan et al., 2020; Vitta and Al-Hoorie, 2020).

These results can be attributed to a combination of well-founded strategies that EFL learners can effectively employ to enhance their reading performance in the context of flipped classrooms. Firstly, the practice of assigning videos and reading tasks for completion outside of the traditional classroom plays a pivotal role. This approach encourages learners to take ownership of their learning process, fostering independence and self-directed study habits. By engaging with course materials independently, students have the opportunity to delve deeper into the content, preparing them for more meaningful in-class discussions and activities. Secondly, the utilization of collaborative reading tasks facilitated through online communication platforms enhances the learning experience. It promotes peer interaction and shared exploration of texts, enabling learners to benefit from diverse perspectives and insights. This collaborative dimension not only enriches their understanding of the reading materials but also cultivates crucial communication skills in an online context, aligning with the demands of the digital age. Furthermore, the practice of facilitating peer feedback exchange during reading activities fosters a constructive learning environment. Learners actively contribute to each other's growth by providing valuable insights and critiques. This not only bolsters their comprehension but also encourages a culture of continuous improvement and mutual support. In addition, individual and collective reflection on reading practices is integral to the success of flipped classrooms. Encouraging learners to assess their own reading comprehension skills and engage in group discussions about their strategies encourages metacognitive awareness. It empowers them to

adapt and refine their approaches to reading, leading to more effective comprehension. Moreover, the promotion of self-evaluation of reading comprehension skills empowers learners to take charge of their progress. By regularly assessing their own understanding and identifying areas for improvement, students become more self-aware and accountable for their learning outcomes.

Finally, providing opportunities for communicative activities beyond the confines of the classroom solidifies the positive outcomes of online flipped classrooms. These activities allow students to apply their reading comprehension skills in real-world contexts, reinforcing their practical utility. As a result, they become more motivated, engaged, reflective, and productive within the classroom environment. These findings align with the research of Strelan et al. (2020) and Vitta and Al-Hoorie (2020), underscoring the effectiveness of these instructional approaches. By implementing these strategies, educators can harness the power of online flipped classrooms to elevate student motivation, engagement, reflection, and productivity. Ultimately, learners actively participate in critical discussions, engage in the negotiation of meaning, and emerge with a deeper understanding of the reading materials.

The effectiveness of these strategies is substantiated by prior research (Guo, 2019; Fulgueras and Bautista, 2020; Mohammad Hosseini et al., 2020; Shih and Huang, 2020; Samiei and Ebadi, 2021; Yulian, 2021; Fischer and Yang, 2022) which has consistently demonstrated their efficacy in improving EFL learners' skills in general and L2 reading in particular. Also, the results of the study are consistent with previous research that has found that flipped instruction can lead to increased student engagement and achievement (Strayer, 2012; Bergmann and Sams, 2014; Stöhr et al., 2020; Turan and Akdag-Cimen, 2020). Flipped instruction allows students to take control of their learning by providing them with access to course materials outside of class and enabling them to review and study at their own pace. The results of the current study also support the findings of previous research that has found that online learning can be an effective method of instruction (Bernard et al., 2009).

It was also revealed that online flipped instruction had a more significant effect on L2 reading comprehension of the EFL participants than flipped instruction. This difference can be attributed to several key factors. Firstly, the essence of the flipped classroom model, as conceptualized by Bergmann and Sams (2014) and Su Ping et al. (2020), revolves around providing students with pre-class access to instructional materials. This empowers them to progress through the content at their own pace, ensuring a foundational understanding before engaging in more dynamic and participatory in-class activities. This advantage is further magnified when executed through online platforms (Stöhr et al., 2020), where students can efficiently allocate their in-class time for focused practice and interactive learning experiences (Hew et al., 2020).

Additionally, a body of prior research underscores the potency of technology-based instruction, particularly the use of online materials and multimedia resources, in augmenting language learning outcomes (González-Lloret, 2019; Jain et al., 2023). The dynamic and interactive nature of technology-supported learning environments can substantially contribute to language skill development (Yang et al., 2021). By providing our students with accessible online materials tailored to their individual learning pace, the online flipped instruction approach likely fostered a deeper engagement with the content and a more effective honing of their reading skills.

Furthermore, the integration of multimedia resources within our online materials played a pivotal role in sustaining student engagement and motivation (Shin et al., 2020). This multimedia-rich environment not only catered to diverse learning preferences but also injected an element of interactivity, capturing and sustaining student interest throughout the learning process.

Previous research also supported that reading comprehension could be enhanced by engaging learners in an online flipped classroom model (Karimi and Hamzavi, 2017; Hashemifardnia et al., 2018; Samiei and Ebadi, 2021). As reading comprehension involves associating the background knowledge with the reading text, EFL learners who are connected to the internet and can search around any topic become able to create the background knowledge before starting their reading due to their time flexibility outside of class. Participating in online flipped classrooms empowers EFL learners, providing them with the agency and autonomy to make informed decisions and take purposeful actions in their reading practices (Fulgueras and Bautista, 2020). This pedagogical approach fosters a sense of ownership and responsibility in learners, motivating them to adapt and refine their reading strategies, and explore novel approaches and techniques (Fischer and Yang, 2022). Furthermore, online flipped classrooms cultivate a positive attitude toward the inherent challenges of comprehending texts, encouraging learners to perceive these challenges as opportunities for personal growth and deeper understanding (Samiei and Ebadi, 2021).

Moreover, EFL learners derive significant benefits from the opportunity to enhance their vocabulary and grammatical knowledge through engaging with various activities before, during, and after reading texts (Turan and Akdag-Cimen, 2020). They actively employ a range of strategies to strengthen their language proficiency, including the use of dictionaries to clarify unfamiliar words, the application of contextual cues to predict and infer meaning, engaging in peer discussions to seek clarification and deepen understanding, and employing effective organizational strategies such as rehearsal, rereading, and summarization. By skillfully utilizing these cognitive and metacognitive strategies, learners cultivate an enriched reading process that facilitates comprehensive comprehension and fosters critical thinking abilities (Karimi and Hamzavi, 2017; Fulgueras and Bautista, 2020; Samiei and Ebadi, 2021; Yulian, 2021; Li et al., 2022; Liu et al., 2022). Furthermore, EFL learners can reach opportunities to discuss and negotiate the reading points, ideas, difficulties, and strategies and engage in problem-solving and information-exchange communicative activities while they are in the classes (Kim et al., 2017; Zarrinabadi and Ebrahimi, 2018). Additionally, Gok et al. (2021) discovered that EFL learners' classroom anxiety and foreign language reading anxiety could be deceased during the flipped classroom, and they could be more motivated to comprehend reading text (Jiang et al., 2021). In the same vein, Hashemifardnia et al. (2018) studied how flipped classrooms predicted junior high school students' reading comprehension. They identified that online flipped classrooms could considerably affect reading comprehension.

Another finding of this study was the significance of online flipped classrooms on self-regulated learning strategies of learners since learners can gradually learn how to acquire, organize, reflect, and appraise their own learning practices (Lai and Hwang, 2016). EFL learners need to plan their tasks and exercises outside of class, set aims and objectives and develop their understanding of their styles of learning (metacognitive strategy), recognize and boost their efforts, consolidate content and input (behavioral and motivational), observe

and evaluate their learning (self-evaluation) (Paris and Paris, 2001; Artino and Stephens, 2009; Zimmerman and Moylan, 2009; Kistner et al., 2010; Kramarski et al., 2013; Panadero and Alonso-Tapia, 2014). Also, as learners need to be actively involved in flipped classrooms to make decisions, become autonomous, take ownership, know their competence, regulate their learning process, perceive the intended meanings, and link the personal and environmental factors, they can reach more incredible self-regulated learning strategies (Zimmerman, 2000; Pintrich, 2004; Pajares, 2009). Thus, online flipped classrooms can benefit EFL learners' reading comprehension and self-regulated strategies. More specifically, online collaborative learning can enhance SRL strategies by providing students with opportunities to work together, reflect on their learning, and engage in metacognitive processes (Zimmerman, 2000; Dabbagh and Kitsantas, 2012; Kim et al., 2017; Lei et al., 2022). online instruction provides learners with greater autonomy and control over their learning, allowing them to engage in self-regulated learning activities such as setting goals, monitoring progress, and adapting strategies as needed (Dabbagh and Kitsantas, 2012). In contrast, traditional classroom-based instruction can limit learner autonomy and control, as learners may be subject to the pace and structure of the lesson set by the teacher.

Another finding of this study was the significance of online flipped classrooms on self-regulated learning strategies of EFL learners. This finding can be attributed to several key factors, each underpinned by sound pedagogical principles. To begin with, the framework of online flipped classrooms inherently promotes the gradual development of SRL skills among learners (Lai and Hwang, 2016). This gradual evolution is facilitated as learners are encouraged to acquire, organize, reflect upon, and appraise their own learning practices over time. In the context of EFL learning, this process necessitates students to meticulously plan their tasks and exercises outside the traditional class setting, establish clear aims and objectives, and cultivate a nuanced understanding of their unique learning styles–a core facet of metacognitive strategy (Paris and Paris, 2001; Artino and Stephens, 2009; Zimmerman and Moylan, 2009; Kistner et al., 2010; Kramarski et al., 2013; Panadero and Alonso-Tapia, 2014).

Moreover, active participation in online flipped classrooms empowers learners to make informed decisions, fostering autonomy, ownership of their learning journey, and an acute awareness of their competence–all central tenets of SRL (Zimmerman, 2000; Pintrich, 2004; Pajares, 2009). Within this dynamic educational context, students are encouraged to regulate their learning processes, decipher intended meanings, and establish connections between personal and environmental factors, thereby further enriching their repertoire of SRL strategies. The advantages of online collaborative learning, a key component of our online flipped classrooms, further bolster the enhancement of SRL strategies. Collaborative learning environments provide students with invaluable opportunities to work together, engage in reflective practices, and participate in metacognitive processes (Zimmerman, 2000; Dabbagh and Kitsantas, 2012; Kim et al., 2017; Lei et al., 2022). Through collaborative endeavors, students gain insights from their peers, engage in discussions that require them to reflect on their learning, and employ metacognitive strategies to evaluate their comprehension and learning progress.

Additionally, online instruction, in general, affords learners greater autonomy and control over their learning experiences. This increased agency empowers students to engage in various SRL activities, including setting meaningful goals, monitoring their progress, and adapting

strategies as needed (Dabbagh and Kitsantas, 2012). In contrast, traditional classroom-based instruction may inadvertently curtail learner autonomy and control, as students often find themselves subject to the predetermined pace and structure of lessons set by the teacher.

Additionally, the flipped instruction approach can provide learners with opportunities to engage in self-regulated learning activities such as previewing, reviewing, and reflecting on materials before and after class (Bergmann and Sams, 2014). However, in a traditional flipped instruction approach, learners may be limited to engaging in these activities outside the classroom, which may not be conducive to developing self-regulated learning skills. In contrast, online flipped instruction can provide learners with ongoing access to materials, resources, and feedback, allowing them to engage in self-regulated learning activities both inside and outside the classroom (Galway et al., 2014; Jia et al., 2023).

Finally, the findings revealed that the OFI students demonstrated superior online learning behaviors and objective performances than the FI students. One key difference was the amount of time the students spent engaging with the online course materials. These differences can be attributed to several key factors, each contributing to the enhanced performance of the OFI students. First and foremost, the notable disparity in the amount of time dedicated to engaging with online course materials between the OFI and FI groups is a pivotal finding. OFI students invested significantly more time in their learning endeavors compared to the FI students. This disparity suggests that the inclusion of the online component within the OFI group acted as a motivating factor, encouraging students to spend more time interacting with course materials. The online collaborative flipped classroom approach inherently promotes active learning, fostering a sense of responsibility and autonomy among students (Burke and Fedorek, 2017; Tang et al., 2023). This encourages learners to delve deeper into the subject matter and allocate more time to their studies.

Moreover, the higher scores achieved by the OFI students on online quizzes provide additional insights. This outcome underscores the effectiveness of the online collaborative flipped classroom approach in the OFI group, as it enabled students to grasp the subject matter more profoundly. The combination of pre-learning through online resources and the collaborative online component in the OFI group likely facilitated a more comprehensive understanding of the material (Ng et al., 2022). This approach allowed students to review content at their own pace, seek clarifications as needed, and actively engage in collaborative activities with their peers–factors that collectively contributed to improved quiz performance. Furthermore, the superior overall final scores attained by the OFI students compared to the FI students signify not only their proficiency in online quizzes but also their enhanced performance on the final assessment. This result suggests that the benefits observed in the online component of the OFI approach transcended the realm of quizzes and extended to broader assessments. The amalgamation of flipped instruction and the collaborative online component likely played a synergistic role in fostering improved learning outcomes.

Conclusion and implications

This study investigated the effectiveness of an online collaborative flipped classroom approach in improving English reading skills and self-regulated learning among Chinese EFL learners. The results demonstrated that both the FI and OFI groups outperformed the traditional group in terms of reading comprehension and self-regulated learning. Moreover, the OFI students exhibited superior online learning behaviors and objective performances compared to the FI students, spending more time engaging with the course materials and achieving higher scores on online quizzes and final assessments. These findings highlight the potential benefits of integrating flipped instruction and online collaboration in L2 instruction.

The findings of this study have several implications for educators and researchers in the field of language instruction. Firstly, the incorporation of an online collaborative flipped classroom approach has the potential to enhance students' engagement, learning outcomes, and overall performance in English reading skills. Educators are encouraged to consider adopting similar instructional methods that promote active learning, self-regulated strategies, and collaborative learning opportunities. Secondly, the study highlights the effectiveness of combining online resources with flipped instruction to create a blended learning environment. Integrating online components allows learners to access course materials flexibly, review content at their own pace, and engage in collaborative activities, thereby enhancing their learning experience.

Moreover, the online collaborative flipped classroom model promoted self-regulated learning behaviors, such as time management, goal-setting, and progress monitoring. Educators should provide explicit instruction and support in developing self-regulated learning skills to empower learners to take ownership of their learning process. Furthermore, the study underscores the importance of leveraging technology for language instruction. Online platforms, interactive tools, and multimedia resources facilitate learner engagement, provide opportunities for authentic practice, and enable effective monitoring of students' progress. In terms of future research, it would be valuable to explore the long-term effects of the online collaborative flipped classroom approach on language acquisition and the transferability of skills. Additionally, investigating the impact of this instructional model on other language skills and diverse learner populations would contribute to a more comprehensive understanding of its potential benefits.

Finally, as the educational landscape undergoes profound transformations driven by technological advancements, it is paramount to recognize the sweeping changes shaping modern education. These shifts span diverse domains, encompassing innovative applications of artificial intelligence technology, the adoption of tailored models for digital transformation, and the exploration of how online platforms impact learner satisfaction (García-Morales et al., 2021; Zarifis and Efthymiou, 2022; Jain et al., 2023; Widayanti and Meria, 2023). Embracing this broader perspective, our research findings naturally align within the dynamic realm of this educational transformation, offering insights that resonate with the evolving fabric of contemporary learning.

Despite the valuable insights garnered from this study, it is essential to acknowledge several limitations. Firstly, the sample size utilized in this study was relatively small, comprising participants from a specific educational context in China. It is important to acknowledge that the relatively modest participant count of 71 EFL students may be subject to constraints in statistical power, particularly for detecting smaller effect sizes. While the sample size was chosen to achieve a balance between practical feasibility and the desire for meaningful statistical analyzes, future studies with larger cohorts could provide a more robust foundation for further validating the outcomes observed in this research. Secondly, the study's duration was limited to a 12-week period, potentially

constraining the ability to observe any long-term effects of the instructional intervention. Future research endeavors should consider implementing the intervention over an extended timeframe to thoroughly assess the sustainability of the observed benefits.

Moreover, it is imperative to recognize that this study relied solely on quantitative data and employed self-report measures to evaluate self-regulated learning strategies and online learning behaviors. While self-report measures are commonly employed, it is crucial to acknowledge their inherent subjectivity and susceptibility to response biases. Augmenting the self-report data with objective measures, such as direct observation or behavioral tracking, would enhance the robustness and validity of the findings, enabling a more comprehensive understanding of learners' engagement and behaviors. Furthermore, it is noteworthy that the findings of this study may have been influenced by the specific context in which it was conducted, encompassing factors such as cultural background, educational system, and institutional support. These contextual elements have the potential to impact the efficacy of the employed instructional approaches. Consequently, caution should be exercised when attempting to generalize the findings to diverse contexts. Future research endeavors should strive to investigate the effectiveness of online collaborative flipped instruction in a range of educational settings, thereby enabling a more nuanced understanding of its applicability.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to YW, wangying0635@163.com.

Ethics statement

Author contributions

YW: Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing.

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

References

American Council on the Teaching of Foreign Languages. (2010). *Maximum class size*. Available at: https://www.actfl.org/search/node/class%20size. (Accessed May 2, 2016).

Amer, M. (2014). A study of learners' usage of a mobile learning application for learning idioms and collocations. *CALICO J.* 31, 285–302. doi: 10.11139/cj.31.3.285302

Andrade, M. S., and Evans, N. W. (2012). *Principles and practices for response in second language writing: developing self-regulated learners*. Abingdon: Routledge.

Artino, A. R., and Stephens, J. M. (2009). Academic motivation and self-regulation: a comparative analysis of undergraduate and graduate students learning online. *Internet High. Educ.* 12, 146–151. doi: 10.1016/j.iheduc.2009.02.001

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: a brief overview. *New Dir. Teach. Learn.* 1996, 3–12. doi: 10.1002/tl.37219966804

Basilaia, G., and Kvavadze, D. (2020). Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia. *Pedagog. Res.* 5, 1–9. doi: 10.29333/pr/7937

Bergmann, J., and Sams, A. (2014). *Flipped learning: Gateway to student engagement*. Washington, DC: International Society for Technology in Education.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., et al. (2009). A meta-analysis of three types of interaction treatments in distance education. *Rev. Educ. Res.* 79, 1243–1289. doi: 10.3102/0034654309333844

Burke, A. S., and Fedorek, B. (2017). Does "flipping" promote engagement?: a comparison of a traditional, online, and flipped class. *Act. Learn. High. Educ.* 18, 11–24. doi: 10.1177/1469787417693487

Butt, A. (2014). Student views on the use of a flipped classroom approach: evidence from Australia. *Bus. Educ. Accreditat.* 6, 33–43.

Çakıroğlu, Ü., and Öztürk, M. (2017). Flipped classroom with problem based activities: exploring self-regulated learning in a programming language course. *J. Educ. Technol. Soc.* 20, 337–349.

Chen Hsieh, J. S., Wu, W. C. V., and Marek, M. W. (2017). Using the flipped classroom to enhance EFL learning. *Comput. Assist. Lang. Learn.* 30, 1–21. doi: 10.1080/09588221.2015.1111910

Critz, C. M., and Knight, D. (2013). Using the flipped classroom in graduate nursing education. *Nurse Educ.* 38, 210–213. doi: 10.1097/NNE.0b013e3182a0e56a

Crouch, C. H., Watkins, J., Fagen, A. P., and Mazur, E. (2007). Peer instruction: engaging students one-on-one, all at once. *Res. Based Reform Univ. Phys.* 1, 40–95.

Dabbagh, N., and Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *Internet. High. Educ.* 15, 3–8. doi: 10.1016/j.iheduc.2011.06.002

Davis, C. (2013). "Flipped or inverted learning: strategies for course design" in *Enhancing instruction with visual media: Utilizing video and lecture capture*. eds. E. G. Smyth and J. X. Volker (Hershey, PA: IGI Global), 241–265.

Fathi, J., Mohammaddokht, F., and Afzali, M. (2023). Exploring Iranian EFL teachers' attitudes toward the use of learning management systems in English classes. *Íkala Rev. Lenguaje Cult.* 28, 30–48. doi: 10.17533/udea.ikala.v28n1a02

Fathi, J., Naghshbandi, Z., and Mohamadi, P. (2021). The effect of a flipped writing classroom on writing performance and self-regulation of Iranian EFL learners. *Lang. Relat. Res.* 12, 627–659. doi: 10.29252/LRR.12.4.20

Fathi, J., and Rahimi, M. (2022). Examining the impact of flipped classroom on writing complexity, accuracy, and fluency: a case of EFL students. *Comput. Assist. Lang. Learn.* 35, 1668–1706. doi: 10.1080/09588221.2020.1825097

Fathi, J., Rahimi, M., and Liu, G. Z. (2023). A preliminary study on flipping an English as a foreign language collaborative writing course with video clips: its impact on writing skills and writing motivation. *J. Comput. Assist. Learn.* 39, 659–675. doi: 10.1111/jcal.12772

Fischer, I. D., and Yang, J. C. (2022). Flipping the flipped class: using online collaboration to enhance EFL students' oral learning skills. *Int. J. Educ. Technol. High. Educ.* 19:15. doi: 10.1186/s41239-022-00320-2

Fisher, T. (2006). Educational transformation: is it, like 'beauty', in the eye of the beholder, or will we know it when we see it? *Educ. Inf. Technol.* 11, 293–303. doi: 10.1007/s10639-006-9009-1

Fulgueras, M. J., and Bautista, J. (2020). Flipped classroom: its effects on ESL learners' critical thinking and Reading comprehension levels. *Int. J. Lang. Lit. Stud.* 2, 257–270. doi: 10.36892/ijlls.v2i3.228

Galway, L. P., Corbett, K. K., Takaro, T. K., Tairyan, K., and Frank, E. (2014). A novel integration of online and flipped classroom instructional models in public health higher education. *BMC Med. Educ.* 14, 1–9. doi: 10.1186/1472-6920-14-181

García-Morales, V. J., Garrido-Moreno, A., and Martín-Rojas, R. (2021). The transformation of higher education after the COVID disruption: emerging challenges in an online learning scenario. *Front. Psychol.* 12:616059. doi: 10.3389/fpsyg.2021.616059

Geduld, B. (2016). Exploring differences between self-regulated learning strategies of high and low achievers in open distance learning. *Afr. Educ. Rev.* 13, 164–181. doi: 10.1080/18146627.2016.1182739

Gok, D., Bozoglan, H., and Bozoglan, B. (2021). Effects of online flipped classroom on foreign language classroom anxiety and reading anxiety. *Comput. Assist. Lang. Learn.* 36, 840–860. doi: 10.1080/09588221.2021.1950191

González-Lloret, M. (2019). Technology and L2 pragmatics learning. *Annu. Rev. Appl. Linguist.* 39, 113–127. doi: 10.1017/S0267190519000047

Guo, J. J. (2019). The use of an extended flipped classroom model in improving students' learning in an undergraduate course. *J. Comput. High. Educ.* 31, 362–390. doi: 10.1007/s12528-019-09224-z

Hasan, M. K., Ibna Seraj, P. M., Fakih, A.-H., and Klimova, B. (2022). Conceptions and viewpoints of English as a foreign language undergraduate students towards flipped instructed classroom. *Educ. Res. Int.* 2022:6140246. doi: 10.1155/2022/6140246

Hashemifardnia, A., Namaziandost, E., and Shafiee, S. (2018). The effect of implementing flipped classrooms on Iranian junior high school students' reading comprehension. *Theory Pract. Lang. Stud.* 8, 665–673. doi: 10.17507/tpls.0806.17

Hew, K. F., Jia, C., Gonda, D. E., and Bai, S. (2020). Transitioning to the "new normal" of learning in unpredictable times: pedagogical practices and learning performance in fully online flipped classrooms. *Int. J. Educ. Technol. High. Educ.* 17, 57–22. doi: 10.1186/s41239-020-00234-x

Hung, H. T. (2015). Flipping the classroom for English language learners to foster active learning. *Comput. Assist. Lang. Learn.* 28, 81–96. doi: 10.1080/09588221.2014.967701

Hung, H. T. (2017). Design-based research: redesign of an English language course using a flipped classroom approach. *TESOL Q.* 51, 180–192. doi: 10.1002/tesq.328

IELTS. (2021). *Test performance*. Available at: https://www.ielts.org/en-us/for-researchers/test-statistics/test-performance.

Jain, A., Sharma, P., and Meher, J. R. (2023). Effects of online platforms on learner's satisfaction: a serial mediation analysis with instructor presence and student engagement. *Int. J. Inf. Learn. Technol.* 2023:17. doi: 10.1108/IJILT-02-2023-0017

Jeon, E. H., and Yamashita, J. (2014). L2 reading comprehension and its correlates: a meta-analysis. *Lang. Learn.* 64, 160–212. doi: 10.1111/lang.12034

Jia, C., Hew, K. F., Jiahui, D., and Liuyufeng, L. (2023). Towards a fully online flipped classroom model to support student learning outcomes and engagement: a 2-year design-based study. *Internet High. Educ.* 56:100878. doi: 10.1016/j.iheduc.2022.100878

Jiang, L., Meng, H., and Zhou, N. (2021). English learners' readiness for online flipped learning: interrelationships with motivation and engagement, attitude, and support. *Lang. Teach. Res.* 2021:27459. doi: 10.1177/13621688211027459

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., Liu, K. S. X., and Park, M. (2022). A scoping review on flipped classroom approach in language education: challenges, implications and an interaction model. *Comput. Assist. Lang. Learn.* 35, 1218–1249. doi: 10.1080/09588221.2020.1789171

Karabulut-Ilgu, A., Jaramillo Cherrez, N., and Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *Br. J. Educ. Technol.* 49, 398–411. doi: 10.1111/bjet.12548

Karimi, M., and Hamzavi, R. (2017). The effect of flipped model of instruction on EFL learners' reading comprehension: learners' attitudes in focus. *Adv. Lang. Literary Stud.* 8, 95–103. doi: 10.7575/aiac.alls.v.8n.1p.95

Karjanto, N., and Simon, L. (2019). English-medium instruction Calculus in Confucian-heritage culture: flipping the class or overriding the culture? *Stud. Educ. Eval.* 63, 122–135. doi: 10.1016/j.stueduc.2019.07.002

Karlen, Y. (2016). Differences in students' metacognitive strategy knowledge, motivation, and strategy use: a typology of self-regulated learners. *J. Educ. Res.* 109, 253–265. doi: 10.1080/00220671.2014.942895

Kendeou, P., Van den Broek, P., Helder, A., and Karlsson, J. (2014). A cognitive view of reading comprehension: implications for reading difficulties. *Learn. Disabil. Res. Pract.* 29, 10–16. doi: 10.1111/ldrp.12025

Kiernan, P. J., and Aizawa, K. (2004). Cell phones in task based learning are cell phones useful language learning tools? *ReCALL* 16, 71–84. doi: 10.1017/S0958344004000618

Kim, J., Park, H., Jang, M., and Nam, H. (2017). Exploring flipped classroom effects on second language learners' cognitive processing. *Foreign Lang. Ann.* 50, 260–284. doi: 10.1111/flan.12260

Kintsch, W. (2012). "Psychological models of reading comprehension and their implications for assessment" in *Measuring up: Advances in how we assess reading ability*. eds. J. P. Sabatini, E. R. Albro and T. O'Reilly (Rowman and Littlefield Education: Plymouth)

Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., Büttner, G., and Klieme, E. (2010). Promotion of self-regulated learning in classrooms: investigating frequency, quality, and consequences for student performance. *Metacogn. Learn.* 5, 157–171. doi: 10.1007/s11409-010-9055-3

Kramarski, B., Weiss, I., and Sharon, S. (2013). Generic versus context-specific prompts for supporting self-regulation in mathematical problem solving among students with low or high prior knowledge. *J. Cogn. Educ. Psychol.* 12, 197–214. doi: 10.1891/1945-8959.12.2.197

Lai, C. L., and Hwang, G. J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Comput. Educ.* 100, 126–140. doi: 10.1016/j.compedu.2016.05.006

Lee, G., and Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: outcomes and perceptions. *TESOL Q.* 52, 62–84. doi: 10.1002/tesq.372

Lee, J., and Choi, H. (2019). Rethinking the flipped learning pre-class: its influence on the success of flipped learning and related factors. *Br. J. Educ. Technol.* 50, 934–945. doi: 10.1111/bjet.12618

Lei, X., Fathi, J., Noorbakhsh, S., and Rahimi, M. (2022). The impact of mobile-assisted language learning on English as a foreign language learners' vocabulary learning attitudes and self-regulatory capacity. *Front. Psychol.* 13:872922. doi: 10.3389/fpsyg.2022.872922

Lin, C. H., Zhang, Y., and Zheng, B. (2017). The roles of learning strategies and motivation in online language learning: a structural equation modeling analysis. *Comput. Educ.* 113, 75–85. doi: 10.1016/j.compedu.2017.05.014

Li, S., He, J., Tao, Y., and Liu, X. (2022). The effects of flipped classroom approach in EFL teaching: can we strategically use the flipped method to acquire communicative competence? *Lang. Teach. Res.* 2022:1575. doi: 10.1177/13621688221081575

Liu, C., Sands-Meyer, S., and Audran, J. (2019). The effectiveness of the student response system (SRS) in English grammar learning in a flipped English as a foreign language (EFL) class. *Interact. Learn. Environ.* 27, 1178–1191. doi: 10.1080/10494820.2018.1528283

Liu, G. Z., Rahimi, M., and Fathi, J. (2022). Flipping writing metacognitive strategies and writing skills in an English as a foreign language collaborative writing context: a mixed-methods study. *J. Comput. Assist. Learn.* 38, 1730–1751. doi: 10.1111/jcal.12707

Maharsi, I., Wijayanti, Y. R., and Astari, T. R. (2021). Evaluating flipped classroom approach in EFL students reading classes. *LLT J.* 24, 92–102. doi: 10.24071/llt.v24i1.2768

Mehring, J. (2016). Present research on the flipped classroom and potential tools for the EFL classroom. *Comput. Sch.* 33, 1–10. doi: 10.1080/07380569.2016.1139912

Meniado, J. C. (2016). Metacognitive Reading strategies, motivation, and Reading comprehension performance of Saudi EFL students. *Engl. Lang. Teach.* 9, 117–129. doi: 10.5539/elt.v9n3p117

Michael, J. (2006). Where's the evidence that active learning works? *Adv. Physiol. Educ.* 30, 159–167. doi: 10.1152/advan.00053.2006

Mohammaddokht, F., and Fathi, J. (2022). An investigation of flipping an English reading course: focus on reading gains and anxiety. *Educ. Res. Int.* 2022:62983. doi: 10.1155/2022/2262983

Mohammad Hosseini, H., Ejtehadi, A., and Mohammad Hosseini, M. (2020). Flipping microlearning-based EFL classroom to enhance learners' self-regulation. *Lang. Teach. Res. Q.* 20, 43–59. doi: 10.32038/ltrq.2020.20.03

Murdock, J. L., and Williams, A. M. (2011). Creating an online learning community: is it possible? *Innov. High. Educ.* 36, 305–315. doi: 10.1007/s10755-011-9188-6

Ng, D. T., Ng, E. H., and Chu, S. K. (2022). Engaging students in creative music making with musical instrument application in an online flipped classroom. *Educ. Inf. Technol.* 27, 45–64. doi: 10.1007/s10639-021-10568-2

Nilson, L. B. (2023). *Creating self-regulated learners: strategies to strengthen students' self-awareness and learning skills*. Abingdon: Taylor and Francis.

Novak, G. M. (2011). Just-in-time teaching. *New Dir. Teach. Learn.* 2011, 63–73. doi: 10.1002/tl.469

Nursyahdiyah, N., Dalimunte, A. A., and Daulay, S. H. (2022). The implementation of flipped classroom in EFL reading during Covid-19 pandemic: Indonesian EFL students' voices. *Engl. Franca* 6, 325–340. doi: 10.29240/ef.v6i2.5329

O'Flaherty, J., and Phillips, C. (2015). The use of flipped classrooms in higher education: a scoping review. *Intern High. Educ.* 25, 85–95. doi: 10.1016/j.iheduc.2015.02.002

Öztürk, M., and Çakıroğlu, Ü. (2021). Flipped learning design in EFL classrooms: implementing self-regulated learning strategies to develop language skills. *Smart Learn. Environ.* 8:2. doi: 10.1186/s40561-021-00146-x

Pajares, F. (2009). "Motivational role of self-efficacy beliefs in self-regulated learning" in *Motivation and self-regulated learning: Theory, research, and applications*. eds. D. H. Schunk and B. J. Zimmerman (New York: Routledge), 111–139.

Panadero, E., and Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Anal. Psicol.* 30, 450–462. doi: 10.6018/analesps.30.2.167221

Paris, S. G., and Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educ. Psychol.* 36, 89–101. doi: 10.1207/S15326985EP3602

Pellegrino, J., and Hilton, M. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: National Academies Press.

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

Pokhrel, S., and Chhetri, R. (2021). A literature review on impact of COVID-19 pandemic on teaching and learning. *Higher Educ. Future* 8, 133–141. doi: 10.1177/2347631120983481

Prensky, M. (2005). In digital games for education, complexity matters. *Educ. Technol.* 45, 22–28.

Samiei, F., and Ebadi, S. (2021). Exploring EFL learners' inferential reading comprehension skills through a flipped classroom. *Res. Pract. Technol. Enhanc. Learn.* 16:12. doi: 10.1186/s41039-021-00157-9

Shih, H. C. J., and Huang, S. H. C. (2020). College students' metacognitive strategy use in an EFL flipped classroom. *Comput. Assist. Lang. Learn.* 33, 755–784. doi: 10.1080/09588221.2019.1590420

Shin, D. S., Cimasko, T., and Yi, Y. (2020). Development of metalanguage for multimodal composing: a case study of an L2 writer's design of multimedia texts. *J. Second. Lang. Writ.* 47:100714. doi: 10.1016/j.jslw.2020.100714

Slavin, R. E. (1991). Synthesis of research of cooperative learning. *Educ. Leadersh.* 48, 71–82.

Sommer, M., and Ritzhaupt, A. (2018). Impact of the flipped classroom on learner achievement and satisfaction in an undergraduate technology literacy course. *J. Inf. Technol. Educ. Res.* 17, 159–182. doi: 10.28945/4059

Stockwell, G. (2013). "Tracking learner usage of mobile phones for language learning outside of the classroom" in *Learner-computer interaction in language education: A festschrift in honor of Robert Fischer. CALICO monograph series*. eds. P. Hubbard, M. Schulze and B. Smith (CALICO: San Marcos, TX), 118–136.

Stöhr, C., Demazière, C., and Adawi, T. (2020). The polarizing effect of the online flipped classroom. *Comput. Educ.* 147:103789. doi: 10.1016/j.compedu.2019.103789

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learn. Environ. Res.* 15, 171–193. doi: 10.1007/s10984-012-9108-4

Strelan, P., Osborn, A., and Palmer, E. (2020). The flipped classroom: a meta-analysis of effects on student performance across disciplines and education levels. *Educ. Res. Rev.* 30:100314. doi: 10.1016/j.edurev.2020.100314

Subedi, S., Nayaju, S., Subedi, S., Shah, S. K., and Shah, J. M. (2020). Impact of E-learning during COVID-19 pandemic among nursing students and teachers of Nepal. *Int. J. Sci. Healthc. Res.* 5, 68–76.

Su Ping, R. L., Verezub, E., Adi Badiozaman, I. F. B., and Chen, W. S. (2020). Tracing EFL students' flipped classroom journey in a writing class: lessons from Malaysia. *Innov. Educ. Teach. Int.* 57, 305–316. doi: 10.1080/14703297.2019.1574597

Sweeny, S. M. (2010). Writing for the instant messaging and text messaging generation: using new literacies to support writing instruction. *J. Adolesc. Adult. Lit.* 54, 121–130. doi: 10.1598/JAAL.54.2.4

Syatriana, E. (2011). *Developing the students' reading comprehension through cognitive reading strategies of the first year students of SMAN 16 Makassar (unpublished master's thesis)*. Indonesia.

Tan, C., Yue, W.-G., and Fu, Y. (2017). Effectiveness of flipped classrooms in nursing education: systematic review and meta-analysis. *Chin. Nurs. Res.* 4, 192–200. doi: 10.1016/J.CNRE.2017.10.006

Tang, T., Abuhmaid, A. M., Olaimat, M., Oudat, D. M., Aldhaeebi, M., and Bamanger, E. (2023). Efficiency of flipped classroom with online-based teaching under COVID-19. *Interact. Learn. Environ.* 31, 1077–1088. doi: 10.1080/10494820.2020.1817761

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: a meta-analysis. *Contemp. Educ. Psychol.* 66:101976. doi: 10.1016/j.cedpsych.2021.101976

Topping, K., and Ehly, S. (1998). *Peer assisted learning*. Oxford: Routledge.

Tse, S. K., Lin, L., and Ng, R. H. W. (2022). Self-regulated learning strategies and reading comprehension among bilingual primary school students in Hong Kong. *Int. J. Biling. Educ. Biling.* 25, 3258–3273. doi: 10.1080/13670050.2022.2049686

Turan, Z., and Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: a systematic review. *Comput. Assist. Lang. Learn.* 33, 590–606. doi: 10.1080/09588221.2019.1584117

University of Cambridge ESOL Examinations. (2011). *Score processing, reporting and interpretation*. http://www.ielts.org/researchers/score_processing_and_reporting.aspx. (Accessed November 11, 2011).

Van Alten, D. C., Phielix, C., Janssen, J., and Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Comput. Educ.* 158:104000. doi: 10.1016/j.compedu.2020.104000

Van Laer, S., and Elen, J. (2017). In search of attributes that support self-regulation in blended learning environments. *Educ. Inf. Technol.* 22, 1395–1454. doi: 10.1007/s10639-016-9505-x

Vitta, J. P., and Al-Hoorie, A. H. (2020). The flipped classroom in second language learning: a meta-analysis. *Lang. Teach. Res.* 27:1403. doi: 10.1177/1362168820981403

Wanner, T., and Palmer, E. (2015). Personalising learning: exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Comput. Educ.* 88, 354–369. doi: 10.1016/j.compedu.2015.07.008

Weir, C. J., and O'Sullivan, B. (2017). *Assessing English on the global stage: The British Council and English language testing, 1941—2016*. Mumbai: Equinox.

Widayanti, R., and Meria, L. (2023). Business modeling innovation using artificial intelligence technology. *Int. Trans. Educ. Technol.* 1, 95–104. doi: 10.33050/itee.v1i2.270

Woodford, P. E. (1982). *An introduction to TOEIC: The initial validity study*. Princeton, NJ: Educational Testing Service. Available at: https://www.ets.org/Media/Research/pdf/TOEIC-RS-00.pdf.

Wu, Y. A. (2001). English language teaching in China: trends and challenges. *TESOL Q.* 35, 191–194. doi: 10.2307/3587867

Yang, X., Kuo, L. J., Eslami, Z. R., and Moody, S. M. (2021). Theoretical trends of research on technology and L2 vocabulary learning: a systematic review. *J. Comput. Educ.* 8, 465–483. doi: 10.1007/s40692-021-00187-8

Yapp, D. J., de Graaff, R., and van den Bergh, H. (2021). Improving second language reading comprehension through reading strategies: a meta-analysis of L2 reading strategy interventions. *J. Sec. Lang. Stud.* 4, 154–192. doi: 10.1075/jsls.19013.yap

Yulian, R. (2021). The flipped classroom: improving critical thinking for critical reading of EFL learners in higher education. *Stud. Engl. Lang. Educ.* 8, 508–522. doi: 10.24815/siele.v8i2.18366

Zainuddin, Z., and Attaran, M. (2016). Malaysian students' perceptions of flipped classroom: a case study. *Innov. Educ. Teach. Int.* 53, 660–670. doi: 10.1080/14703297.2015.1102079

Zarifis, A., and Efthymiou, L. (2022). *The four business models for AI adoption in education: giving leaders a destination for the digital transformation journey*. In *2022 IEEE global engineering education conference (EDUCON)*. IEEE, pp. 1868–1872.

Zarrinabadi, N., and Ebrahimi, A. (2018). Increasing peer collaborative dialogue using a flipped classroom strategy. *Innov. Lang. Learn. Teach.* 13, 267–276. doi: 10.1080/17501229.2018.1455688

Zhang, S., and Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: a meta-analysis. *Lang. Teach. Res.* 26, 696–725. doi: 10.1177/1362168820913998

Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory Pract.* 41, 64–70. doi: 10.1207/s15430421tip4102_2

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, CA: Academic Press), 13–39.

Zimmerman, B. J., and Moylan, A. R. (2009). "Self-regulation: where metacognition and motivation intersect" in *Handbook of metacognition in education*. eds. D. J. Hacker, J. Dunlosky and A. C. Graesser (New York: Routledge), 299–315.

Zimmerman, B. J., and Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Abingdon: Routledge.

Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning

Ling Wei*

College of Foreign Languages, Chongqing College of Mobile Communication, Chongqing, China

Introduction: This mixed methods study examines the effects of AI-mediated language instruction on English learning achievement, L2 motivation, and self-regulated learning among English as a Foreign Language (EFL) learners. It addresses the increasing interest in AI-driven educational technologies and their potential to revolutionize language instruction.

Methods: Two intact classes, consisting of a total of 60 university students, participated in this study. The experimental group received AI-mediated instruction, while the control group received traditional language instruction. Pre-tests and post-tests were administered to evaluate English learning achievement across various domains, including grammar, vocabulary, reading comprehension, and writing skills. Additionally, self-report questionnaires were employed to assess L2 motivation and self-regulated learning.

Results: Quantitative analysis revealed that the experimental group achieved significantly higher English learning outcomes in all assessed areas compared to the control group. Furthermore, they exhibited greater L2 motivation and more extensive utilization of self-regulated learning strategies. These results suggest that AI-mediated instruction positively impacts English learning achievement, L2 motivation, and self-regulated learning.

Discussion: Qualitative analysis of semi-structured interviews with 14 students from the experimental group shed light on the transformative effects of the AI platform. It was found to enhance engagement and offer personalized learning experiences, ultimately boosting motivation and fostering self-regulated learning. These findings emphasize the potential of AI-mediated language instruction to improve language learning outcomes, motivate learners, and promote autonomy.

Conclusion: This study contributes to evidence-based language pedagogy, offering valuable insights to educators and researchers interested in incorporating AI-powered platforms into language classrooms. The results support the notion that AI-mediated language instruction holds promise in revolutionizing language learning, and it highlights the positive impact of AI-driven educational technologies in the realm of language education.

KEYWORDS

AI-mediated language instruction, English learning achievement, L2 motivation, self-regulated learning, EFL learners, mixed-methods approach

1. Introduction

The application of information technology in language learning and teaching has garnered considerable attention from language researchers in recent years (Ahmadi, 2018; Shadiev and Yang, 2020; Lai et al., 2022; Lei et al., 2022; Soleimani et al., 2022; Shadiev et al., 2023). Leveraging information technology in language education can enhance the learning experience for learners by enabling personalized, interactive, and communicative learning processes (Rodinadze and Zarbazoia, 2012; Chun et al., 2016; Shatri, 2020; Rahimi and Fathi, 2022). Language educators have adopted information technology to create virtual language learning environments that actively engage learners and facilitate language learning processes (Fathi and Rahimi, 2022; Loncar et al., 2023; Nguyen and Le, 2023). Among the information technology programs, artificial intelligence (AI) has emerged as a promising tool applied in language learning and teaching to enhance learners' learning achievements (Haristiani, 2019; Pedro et al., 2019; Knox, 2020; Pikhart, 2020; Huang et al., 2023).

In the context of computer programs, AI is purposefully designed to interpret and respond to human queries, operating as a platform that relies on human intelligence to furnish pertinent information (Buzko et al., 2016; Spector and Ma, 2019; Devi et al., 2022; Khosravi et al., 2022; Nemorin et al., 2023). For example, ChatGPT, an AI-equipped tool, can efficiently supply users with requested information based on their queries (Fitria, 2023; Yan, 2023). In fact, the emergence of AI has ushered in transformative changes across various industries, and the field of education and language learning is no exception (Spiro et al., 2017; Balyen and Peto, 2019; Su et al., 2023). AI's potential to revolutionize traditional teaching and learning methods has captured the attention of educators, researchers, and policymakers worldwide (Michalski et al., 2013; Ilkka, 2018). With its ability to process vast amounts of data, analyze complex patterns, and deliver personalized insights, AI offers new possibilities for enhancing educational practices and student outcomes (e.g., Roll and Wylie, 2016; Ouyang and Jiao, 2021). Within this direction, educators have integrated AI-assisted language learning tools into education to support learners in enhancing their language skills (Lu, 2018; Tafazoli et al., 2019). ChatGPT, as an AI-assisted language learning tool, offers potential benefits for language learners' language skills and subskills (Baskara and Mukarto, 2023; Hong, 2023; Kohnke et al., 2023). It provides learners with writing ideas, suggests alternative sentences to improve their writing performance, and contributes to their language learning achievements (Su et al., 2023; Yan, 2023). AI-supported language learning tools are known for creating immersive and engaging learning environments, allowing learners to conveniently undertake language learning tasks and improve their overall language proficiency (Divekar et al., 2022).

Several research studies have explored the influence of AI-assisted language learning tools on English language learners' overall learning achievement and specific language skills and sub-skills (Kim, 2019; Junaidi, 2020; Zheng et al., 2021; Xu et al., 2022; Hsu et al., 2023; Yan, 2023). For instance, Xu et al. (2022) conducted a study investigating the impact of AI-powered language learning tools on English language learners' overall learning achievement and found a positive contribution of AI-assisted language learning tools to learners' achievement. Another study by Hsu et al. (2023) examined the effects of AI-assisted language learning tools on EFL learners' vocabulary knowledge and revealed that learners utilizing AI tools demonstrated significant improvement and outperformed their peers in vocabulary knowledge. Moreover, Junaidi (2020) investigated the role of AI-assisted language learning tools in enhancing EFL learners' speaking skills and found that AI learners outperformed non-AI learners in speaking proficiency.

Despite these valuable insights, prior research has not sufficiently delved into the impact of AI-supported instruction on English language learners' language learning achievement, particularly in the context of English as a foreign language (EFL). Therefore, undertaking a comprehensive inquiry into the impact of AI-assisted language learning tools on the language learning achievements of EFL learners would constitute a significant and valuable contribution to the current body of literature. Furthermore, exploring the role of AI-assisted instruction in enhancing EFL learners' second language (L2) motivation and self-regulated learning, both of which are crucial aspects of English language learning, would provide further depth to the literature. L2 motivation significantly influences English language learners' engagement and efforts to achieve language proficiency (Boo et al., 2015), while self-regulatory learning pertains to learners' ability to autonomously plan, monitor, and evaluate their language learning progress (Moyer, 2014).

The present study contributes to the existing literature in several ways. First and foremost, it addresses a notable gap by providing empirical evidence on the specific impact of AI-assisted language learning tools on EFL learners' language learning achievement, L2 motivation, and self-regulated learning. Although prior research has explored the effectiveness of AI-driven language learning tools, our study narrows its scope to EFL learners, providing valuable insights tailored to this specific context. Second, this research extends the theoretical understanding of the role of AI in language instruction. While the literature has recognized the favorable impacts of AI on language learning, we have adopted a mixed methods research design to delve more profoundly into the specific ways in which AI-mediated instruction contributes to the improvement of language proficiency and the emotional and motivational aspects of EFL learners.

The theoretical contribution of this study extends to the broader understanding of the interplay between technology-driven instruction and pedagogical strategies. The investigation aimed to elucidate whether the observed differences between AI and traditional instruction conditions are attributed to the technology used or the instructional strategies employed. By scrutinizing these elements, the study seeks to provide insights into the distinctive advantages that AI-supported instruction may bring to the language learning landscape, shedding light on whether its impact is rooted in technological novelty or pedagogical innovation.

2. Literature review

2.1. Theoretical framework

The theoretical framework underpinning this study draws on Vygotsky's (1984) influential contributions to social constructivist theories of learning. Vygotsky's work provided a crucial foundation for the development of social constructivism, which emphasizes the significant role of social interactions and collaborative learning experiences and cognitive growth. In this perspective, less proficient learners engage in collaborative learning activities with more

proficient individuals, including instructors or, in contemporary settings, computer programs. These interactions serve as cognitive scaffolding, supporting less proficient learners in the acquisition and development of their knowledge (Vygotsky and Cole, 1978; Vygotsky, 1984).

A fundamental concept within social constructivist theory is the Zone of Proximal Development (ZPD), which is proposed in a broader developmental perspective, encompassing various domains beyond learning activities (Vygotsky et al., 1997). The ZPD comprises two distinct levels: the actual level, reflecting a learner's demonstrated abilities in independent tasks, and the potential level, representing their untapped capacity that can be realized through active engagement in interactive learning activities with peers or more proficient individuals. The ZPD illuminates the dynamic interplay between a learner's current state of development and the scaffolding provided by the learning environment. It serves as a compass guiding educator to facilitate learning experiences that are optimally challenging, promoting growth and skill acquisition.

In the context of this study, both the AI-assisted and non-AI-assisted groups engage in interactive language learning activities, firmly aligned with Vygotsky's social constructivist approach (Swain et al., 2015). In the control group, learners interact with their peers, contributing to each other's ZPD by collaboratively assisting in their language learning journey. Conversely, in the experimental group, learners engage with an AI-assisted language learning tool, which serves as a collaborative partner in the language learning process. Through these interactions, learners harness AI technology to regulate their language learning and advance toward their ZPD, underpinning the core of our research exploration.

In alignment with Vygotsky's principles, this study incorporates contemporary theories related to AI and computer-assisted learning, recognizing the transformative impact of technology on language instruction. The integration of AI technology introduces additional layers of collaborative learning (Zheng et al., 2021; Weng and Chiu, 2023), where learners interact with AI-assisted language learning tools to regulate their language learning experiences and advance toward their ZPD. These interactions, which parallel the core principles of Vygotsky's social constructivist theory, emphasize the role of technology as a collaborative partner in the language learning process.

Also, it is worth noting that collaborative abilities emerge as a pivotal aspect that significantly influences the effectiveness of AI-based educational tools in the context of AI-mediated instruction and learning (Akata et al., 2020; Chen et al., 2022). Collaborative abilities refer to learners' proficiency in actively engaging with AI systems, instructors, or peers to collectively enhance their learning skills and achieve optimal learning outcomes (Hwang et al., 2020). These abilities encompass a spectrum of skills and strategies employed by learners to interact effectively within AI-mediated environments.

Collaborative abilities also play a fundamental role in shaping the dynamics of AI-assisted language learning (Fitria, 2023). Learners who exhibit strong collaborative abilities are adept at harnessing AI tools to engage in meaningful interactions, seek clarification, and co-construct knowledge. Moreover, collaborative abilities are closely intertwined with the utilization of feedback mechanisms within AI systems (Hsu et al., 2023). The quality and impact of feedback received by learners depend on their collaborative abilities to effectively interpret and apply feedback to their language learning practices.

Overall, via integrating Vygotsky's foundational principles with contemporary theories in AI and computer-assisted learning, this study seeks to explore the evolving dynamics of language learning in the digital age. The collaborative interactions between learners and AI technology might represent a convergence of established educational theories and emerging technological paradigms, contributing to a deeper understanding of language learning achievement, L2 motivation, and self-regulated learning in the context of AI-mediated language instruction.

2.2. Artificial intelligence

The rapid advancement of AI has revolutionized various domains, including education, with profound implications for teaching and learning practices (Chen et al., 2020). AI, as a branch of computer science, enables machines to simulate human intelligence, learn from experiences, and perform tasks that typically require human cognitive abilities. In the education contexts, AI technologies hold immense potential to transform traditional instructional methods, providing personalized learning experiences tailored to individual needs and preferences (Hwang et al., 2020). From intelligent tutoring systems and language learning applications to adaptive learning platforms, AI's integration in education has garnered considerable attention from researchers, educators, and policymakers worldwide (Ilkka, 2018; Kim et al., 2019; Chen et al., 2022; Huang and Tan, 2023).

According to Aldosari (2020), AI is characterized as an intelligent program capable of executing diverse tasks. For instance, individuals can seek assistance from AI-powered tools for academic inquiries, and these tools promptly provide the required information. AI finds applications in educational settings, making intelligent decisions akin to human decision-making. Additionally, AI is widely employed in language learning to enhance learners' language skills and sub-skills (Zhang and Zou, 2020; Xia et al., 2022a). Numerous AI-assisted language learning tools are accessible on computers and mobile devices, facilitating language learners in their language learning endeavors. These tools offer valuable support in improving various language learning skills.

For instance, ChatGPT, which is an AI-assisted tool, might be utilized in language learning settings to help learners develop their language learning skills and sub-skills (Fang et al., 2023; Fitria, 2023; Kim, 2023; Schmidt-Fajlik, 2023; Su et al., 2023; Yan, 2023). ChatGPT can provide language learners with the required feedback and comments on different language learning skills and sub-skills issues in order to contribute to learners' language achievement in general. ChatGPT is able to provide grammatically sound sentences which can help learners produce well-organized texts. This AI-assisted language learning tool can also understand human inquiries and provide the best possible answers (Huang and Tan, 2023).

A substantial body of studies has been carried out examining the effects of AI-assisted language learning tools on English language learners' language achievement (Suryana et al., 2020; Divekar et al., 2022; Fitria, 2023). For instance, Zheng et al. (2021) carried out a meta-analysis of the impact of AI on learning achievement and learning perception. Twenty-four papers, including 2,908 participants, from 2001 to 2020 were analyzed. The results indicated more significant effects of AI on learning achievement in comparison to its effects on learning perception. This means that most of the studies

revolved around the great impact of AI on learning achievement followed by its impact on learner perceptions. Xu et al. (2022) examined the effects of AI-assisted language learning on English language learner speech and interaction. The findings indicated that the AI-assisted language learning tool using a speech recognition feature improved the learners' language learning achievement and engaged them in interactive language learning activities.

Ebadi and Amini (2022) investigated the impact of AI-assisted language learning on EFL learners' language learning engagement. The data were collected via motivation, social presence, and human-likeness questionnaires and recording the interaction of the EFL learners with the AI tool. The findings revealed that the AI tool had a significant impact on the learners' learning motivation and engagement. In a similar vein, Carpio Cañada et al. (2015) examined the effects of an AI-powered language learning approach on language learners' motivation and learning achievement. The results indicated that the AI-assisted language learning approach contributed to the learners' language learning motivation. The learners' motivation in language learning activities also resulted in the learners' better learning achievement. Similarly, Ali et al. (2023) explored the impact of an AI tool (i.e., ChatGPT) on English language learners' and teachers' motivation. The results demonstrated that the AI-assisted language learning tool had a great contribution to the learners' writing and reading skills, while it had neutral influences on their speaking and listening skills. In much the same vein, Schmidt-Fajlik (2023) compared the differences between ChatGPT, Grammarly, and ProWritingAid tools in checking, understanding, and developing EFL learners' English grammar. The findings indicated that ChatGPT was more effective than the other AI tools in detecting and improving the EFL learners' English grammar.

As for language learning sub-skills, Hsu et al. (2023), for example, examined the impact of AI-assisted image recognition technologies on EFL learners' vocabulary knowledge, self-regulation, and anxiety. Utilizing an experimental research design, the learners in the experimental group received AI-assisted image recognition technologies to demonstrate images along with the related vocabulary to develop the learners' vocabulary knowledge. The findings showed that the experimental learners developed their vocabulary knowledge and self-regulation and decreased their language learning anxiety. However, the experimental learners only outperformed their control counterparts in vocabulary knowledge. That is, there were no significant differences between the experimental and control groups regarding their self-regulation and language learning anxiety. Kim (2019) also investigated the influence of AI-supported language learning instruction on EFL university students' grammar skills. Using an experimental research design, one group of the students received the AI tool while the other group followed the conventional instruction of grammar skills. The results showed that the experimental students who received the AI-based instruction outperformed those who did not receive the AI tool, signifying that the AI-assisted language learning instruction had a great contribution to the students' grammar skill development.

With regard to language learning skills, Yan (2023), for instance, examined the role of an AI-assisted language learning tool on EFL students' writing performance. The findings revealed that the AI-assisted language learning tool developed the writing performance of the EFL learners. Yan also claimed that the EFL learners, using the AI-assisted language learning tool, could

accomplish their writing tasks more swiftly than ever, findings which capitalized on the efficient role of the AI-assisted language learning tool for the writing tasks. Similarly, Utami et al. (2023) investigated the impact of AI-powered language learning on the academic research writing of three Indonesian EFL learners. Following a case study research design, the required data were gathered via questionnaires and interviews. The results showed that the AI-powered language learning approach enhanced the students' academic research writing by giving them the required feedback, comments, and alternative sentences. Moreover, the AI-powered language learning approach was found to enhance the writing engagement of the EFL learners. Lee et al. (2023) also explored the effects of an AI-assisted language learning tool on EFL learners' reading enjoyment. A group of EFL learners received the AI tool to generate reading topics for the learners based on their interests, while the other group were engaged in conventional reading comprehension activities. The findings revealed that the AI-powered language learning tool had a substantial contribution to the learners' reading enjoyment.

El Shazly (2021) investigated the impact of AI-powered language learning tools on Egyptian EFL learners' speaking performance and foreign language anxiety. The learners' speaking performance was examined through roleplaying activities and evaluated via an IELTS-speaking test, and their foreign language anxiety was examined through a questionnaire. The findings revealed that the use of the AI-assisted language applications in the learners' speaking activities did not diminish the learners' foreign language anxiety. However, it was revealed that the utilization of AI-assisted language learning tools increased the learners' speaking performance and interactive speaking activities. In much the same vein, Junaidi (2020) investigated the impact of AI-assisted language learning tools on EFL learners' speaking skills. Adopting a quasi-experimental research design, both experimental and control groups' speaking skills, including fluency, grammar, vocabulary, and pronunciation, were assessed via pre- and post-tests to check the learners' improvements in their speaking skills on the one hand and to examine the differences between the experimental and control learners' speaking skills on the other hand. The findings demonstrated that the learners' speaking skills were developed in both groups; however, the experimental learners outperformed their control counterparts in speaking skills, which corroborated the significant roles of AI-powered language learning tools in developing EFL learners' speaking skills.

The literature review presented herein illuminates the multifaceted impact of AI-assisted language learning tools on various language skills and sub-skills, underscoring their potential to enhance vocabulary knowledge, grammar proficiency, writing skills, reading enjoyment, speaking performance, as well as learners' motivation and engagement. Although previous research has probed the effects of AI on specific facets of language learning, a noteworthy gap exists in the exploration of its holistic influence on overall language achievement, particularly in the context of EFL instruction for Chinese learners. Moreover, the dimensions of L2 motivation and self-regulated learning, integral components of language proficiency and autonomous learning, have received limited attention within the realm of AI-supported language instruction. This discernible gap in the literature paves the way for the current study, which not only aims to bridge this void but also to elucidate whether the observed differences

between AI and traditional instruction conditions stem from technological novelty or pedagogical innovation. Thus, the present research seeks to contribute to the broader understanding of technology-driven language instruction and its multifaceted implications, shedding light on AI's potential to reshape language learning paradigms.

2.3. The present study

As supported by the existing literature, AI-assisted language learning applications have demonstrated a significant role in enhancing English language learners' overall language learning achievement and specific language skills and sub-skills (Kim, 2019; El Shazly, 2021; Xu et al., 2022; Lee et al., 2023; Schmidt-Fajlik, 2023; Yan, 2023). Moreover, these studies have shown that AI-powered language learning tools contribute to improved self-regulated learning and increased motivation among English language learners (Carpio Cañada et al., 2015; Ebadi and Amini, 2022; Hsu et al., 2023).

Despite the positive findings from the literature, there remains a dearth of research specifically investigating the impact of AI-assisted language learning tools on EFL learners' English learning achievement, L2 motivation, and self-regulated learning. To address this gap, the present study aimed to quantitatively examine the effects of the AI-assisted language learning approach on these key aspects of EFL learners' language development. Additionally, in order to gain deeper insights into the learners' experiences, perceptions, and attitudes toward the utilization of AI-powered language learning tools, the study also conducted a qualitative exploration.

The research questions guiding this study are as follows:

1. Are there any significant differences between AI and non-AI-assisted language learning instruction in developing English learning achievement, L2 motivation, and self-regulated learning of EFL learners?
2. What are the perceptions of EFL learners toward the effects of AI-assisted language learning on their language learning achievement?

3. Method

3.1. Participants

This study was conducted at a university in mainland China, involving a carefully selected group of EFL learners. Two intact classes ($n = 60$) were included, and participants were randomly assigned to either the experimental or control group to ensure a fair distribution of learners' characteristics. The Experimental Group consisted of 30 participants, comprising 17 females and 13 males, all aged between 19 and 23 years, with intermediate English proficiency levels, as assessed through a standardized English proficiency test. Similarly, the Control Group consisted of 30 participants, including 19 females and 11 males, aged between 20 and 24 years, with intermediate English proficiency levels evaluated through the same standardized test.

To be included in the study, participants had to meet certain criteria, including being undergraduate students majoring in various fields at the university, having no prior experience with AI-mediated language instruction, and not being diagnosed with any learning disabilities that could significantly affect their language learning abilities.

Throughout the study, ethical considerations were of utmost importance to ensure the well-being and rights of the participants. Before the study began, informed consent was obtained from all participants, ensuring they were fully informed about the purpose and procedures of the research, and they had the right to withdraw at any time. The researchers also maintained strict confidentiality and data anonymity to protect the participants' privacy. The study adhered strictly to ethical guidelines for research involving human participants, highlighting the researchers' commitment to prioritize the participants' well-being and rights at every stage of the study.

3.2. Instruments

3.2.1. English achievement test

To gauge participants' initial level of English achievement, a comprehensive English achievement test was administered to both the experimental and control groups during the pre-test phase. The English achievement test was meticulously designed by a panel of experienced teachers with extensive expertise in language teaching and testing, ensuring its reliability and relevance to the study's objectives. The test was thoughtfully tailored to align with the learning objectives and curriculum of the university's EFL program, making it a suitable tool to assess participants' language proficiency levels accurately. It consisted of multiple sections, each evaluating essential language skills, including grammar, vocabulary, reading comprehension, and writing. To ensure the quality and validity of the English achievement test, the assessment underwent a thorough evaluation by three domain experts. Their valuable input and critical analysis confirmed its face and content validity, ensuring that the test effectively measured the targeted language competencies. The English achievement test demonstrated high reliability with a Cronbach alpha coefficient of 0.87, indicating strong internal consistency among its items.

3.2.2. L2 motivation scale

In this study, students were given the 16-item motivation scale designed by Mehdiyev et al. (2017) at both the beginning and end of the procedure process. The scale utilizes a five-point Likert scale, ranging from 'Totally agree' to 'Absolutely disagree'. It comprises three distinct factors: attitude (4 items, "I find learning English more challenging compared to others."), self-confidence (5 items, "The process of learning English gives me happiness."), and personal use (7 items, "I aim to acquire English skills to stay informed about global developments."). The Cronbach alpha values for each factor were found to be 0.81, 0.79, and 0.82, respectively, with an overall reliability coefficient of 0.82.

3.2.3. Self-regulation questionnaire (SRQ)

The Self-Regulation Questionnaire (SRQ) utilized in this study was developed by Brown et al. (1999). It consists of 63 items designed to assess seven subprocesses, forming seven rationally derived

subscales of the SRQ. Each subscale comprises nine items, and the total sum score is recommended to estimate the participants' self-regulation ability. To respond to the items, students in the study used a Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). A sample item of the scale is "I can effectively manage my study time to meet my learning goals." In the pilot study conducted for this research, the SRQ exhibited a reliability index of 0.83, further confirming its consistency and accuracy.

3.2.4. Semi-structured interview

To collect qualitative data, we conducted semi-structured interviews with 14 participants from the experimental group. We employed a purposive sampling strategy to ensure a well-rounded representation of both gender and age groups within the experimental cohort, aiming to capture a diverse range of experiences and perspectives for a richer qualitative analysis. The demographic profile of these participants is as follows: Six of them were female, while eight were male, with ages ranging from 20 to 24 years. Their English proficiency levels were assessed as intermediate through standardized tests conducted at the study's outset. Furthermore, the participants brought diverse academic backgrounds, spanning various disciplines across the university's academic spectrum.

The semi-structured interviews were thoughtfully designed to offer participants a flexible yet focused platform to share their insights and perceptions regarding AI-mediated language instruction. These interviews explored a wide array of aspects related to their language learning journey with the Duolingo platform. Topics encompassed their emotions and attitudes toward interactive learning activities, the influence of personalized feedback on language skill development, and their levels of motivation and autonomy throughout the learning process. Additionally, participants were encouraged to reflect on any challenges they encountered and articulate the advantages they perceived from their engagement with AI-mediated instruction. Conducted in a one-on-one format, these interviews fostered a comfortable and open atmosphere that facilitated participants in freely expressing their thoughts. With their consent, each interview session was audio-recorded, supplemented by detailed notes capturing both verbal and non-verbal cues, as well as contextual information. Subsequently, we transcribed the interview transcripts verbatim and subjected them to a rigorous thematic analysis, a systematic method used to identify and analyze meaningful patterns and themes within the qualitative data.

3.2.5. AI-mediated language learning platform

For the experimental group, the AI-mediated language learning platform, Duolingo, was utilized to deliver interactive and personalized English language lessons. Duolingo's effectiveness in providing adaptive and engaging language instruction has been widely recognized, making it a suitable platform for this study. Learners in the experimental group had access to a variety of interactive exercises, online quizzes, language games, and real-time feedback tailored to their individual performance on the platform. The control group received traditional language instruction using standard teaching materials, including textbooks, lectures, and classroom activities. Teacher-led discussions, grammar drills, reading passages, and writing exercises were incorporated into the curriculum to facilitate language learning in the control group.

3.2.6. Data collection procedure

Before implementing the intervention, a pre-test assessment was conducted for both the experimental and control groups to measure

their initial English learning achievement. The pre-test evaluated participants' proficiency in grammar, vocabulary, reading comprehension, and writing skills. Additionally, participants completed self-report questionnaires to gauge their L2 motivation and self-regulated learning strategies, providing valuable insights into their attitudes and approaches toward language learning. Once the pre-test phase was completed, random assignment was used to divide participants into either the experimental group or the control group. This random assignment ensured that any individual differences or biases between the two groups were evenly distributed, enhancing the study's reliability.

The experimental group was exposed to AI-mediated language instruction using the renowned AI-powered language learning platform, Duolingo. The 10-week intervention period included interactive and personalized English language lessons. Participants engaged in various activities, such as interactive exercises, online quizzes, language games, and received real-time feedback tailored to their performance. In-class activities involved group discussions, language exercises, and practice sessions with Duolingo, all facilitated by the instructor. To reinforce learning, participants were encouraged to use Duolingo outside of class, dedicating at least 2 h per week to supplementary practice.

In parallel, the control group received traditional language instruction for the same 10-week duration. This instruction emphasized conventional teaching methods and materials, including textbooks, lectures, and classroom activities. Sample activities included teacher-led discussions, grammar drills, reading passages, and writing exercises. In-class sessions were teacher-centered, with the instructor leading the lectures and activities. Outside of class, participants were encouraged to practice through assigned homework, dedicating a minimum of 2 h per week to further develop their language skills.

After the 10-week intervention period, both groups underwent a post-test assessment mirroring the pre-test evaluation. The post-test measured English learning achievement, L2 motivation, and self-regulated learning, enabling a comparative analysis of the two groups' progress and the impact of their respective interventions.

Throughout the data collection process, researchers diligently monitored the implementation of interventions and participant engagement to ensure accuracy and validity. Ethical considerations, including obtaining informed consent and maintaining confidentiality, were strictly adhered to throughout the study to protect the participants' rights and well-being. The data collected from the pre-test and post-test assessments, as well as the self-report questionnaires, were subsequently subjected to appropriate statistical analyses to determine the effects of AI-mediated language instruction using Duolingo on English learning achievement, L2 motivation, and self-regulated learning when compared to traditional language instruction.

3.3. Data analysis

3.3.1. Quantitative data analysis

For the quantitative phase, descriptive statistics and inferential analyses examined the effect of AI-mediated language instruction on English learning achievement, L2 motivation, and self-regulated learning of EFL learners. Statistical Package for the Social Sciences (SPSS) version 26 was used for all analyses. Descriptive statistics

summarized pre-test and post-test scores for both groups. A mixed-design analysis of variance (ANOVA) was conducted to determine the intervention effect, assessing main effects of group (experimental vs. control), time (pre-test vs. post-test), and the interaction between group and time on L2 achievement, L2 motivation, and self-regulation. The main effect of group compared post-test scores, indicating AI-mediated instruction's impact on these variables compared to traditional instruction. The interaction effect between group and time explored differences in impact over the intervention period.

3.3.2. Qualitative data analysis

For the qualitative phase, thematic analysis examined semi-structured interview data from 14 students in the experimental group. Transcripts were transcribed verbatim and subjected to open coding, identifying initial codes capturing participants' experiences with AI-mediated instruction. Codes were organized into potential themes, reflecting meaningful patterns related to the impact on English learning achievement, L2 motivation, and self-regulated learning. Thematic saturation was achieved through iterative coding, and member checking and peer debriefing ensured trustworthiness (Williams and Morrow, 2009). Integrating qualitative findings with quantitative results in the discussion section provided a comprehensive understanding of AI-mediated language instruction's impact. Triangulation enriched the analysis, enhancing validity and reliability.

4. Results

4.1. Impact on English learning achievement, L2 motivation, and self-regulated learning

Prior to conducting the mixed-design ANOVA, we assessed the assumptions for robust and reliable inferential statistics. Normality assumption was examined through histograms, Q-Q plots, and Shapiro–Wilk tests, which confirmed that the data were approximately normally distributed within each group and time point ($p > 0.05$). The assumption of homogeneity of variance was met as indicated by Levene's test and Brown-Forsythe test ($p > 0.05$) for all conditions. Sphericity assumption was violated for the repeated measures factor (Time) in the analysis of L2 achievement ($p < 0.05$); thus, the Greenhouse–Geisser correction was applied. The assumption of independence was satisfied through random assignment of participants to experimental and control groups, ensuring data collection without external influence.

Table 1 presents the descriptive statistics of the pre- and post-test scores of L2 achievement, L2 motivation, and self-regulation for both the experimental and control groups. Prior to the intervention, there were no significant differences between the two groups in terms of their pre-test scores for L2 achievement (Experimental: $M = 43.21$, $SD = 12.69$; Control: $M = 44.39$, $SD = 13.21$), L2 motivation (Experimental: $M = 3.12$, $SD = 1.6$; Control: $M = 3.04$, $SD = 1.4$), and self-regulation (Experimental: $M = 2.89$, $SD = 0.54$; Control: $M = 3.01$, $SD = 0.69$). These similar baseline scores ensured a balanced starting point for the intervention.

Following the 10-week intervention period, the experimental group ($M = 73.86$, $SD = 15.26$) demonstrated a significant improvement in their post-test scores on L2 achievement compared

to the control group ($M = 61.11$, $SD = 14.97$). The mixed-design ANOVA results in Table 2 revealed a significant main effect for Group [$F(1, 58) = 94.35$, $p < 0.001$, $\eta^2 = 0.81$], indicating that the AI-mediated language instruction had a substantial impact on enhancing English learning achievement among EFL learners. Moreover, there was a significant main effect for Time [$F(1, 58) = 10.89$, $p < 0.001$, $\eta^2 = 0.27$], suggesting that both groups experienced an overall improvement in their L2 achievement scores over the 10-week period. Additionally, the interaction effect between Group and Time was significant [$F(1, 58) = 8.37$, $p < 0.001$, $\eta^2 = 0.23$], indicating that the improvement in L2 achievement significantly differed between the experimental and control groups as a result of AI-mediated instruction.

Regarding L2 motivation, the experimental group ($M = 3.89$, $SD = 1.8$) exhibited a significant increase in their post-test scores compared to the control group ($M = 3.35$, $SD = 1.5$) after the intervention. The mixed-design ANOVA results in Table 2 indicated a significant main effect for Time [$F(1, 58) = 34.10$, $p < 0.001$, $\eta^2 = 0.44$], suggesting that both groups experienced an overall improvement in their motivation to learn English over the 10-week period. Nonetheless, there was a significant interaction effect between Group and Time [$F(1, 58) = 7.43$, $p = 0.04$, $\eta^2 = 0.11$], indicating that that the improvement in L2 motivation significantly differed between the experimental and control groups due to the AI-mediated instruction.

For self-regulated learning, the experimental group ($M = 3.94$, $SD = 0.73$) demonstrated a significant increase in their post-test scores, whereas the control group ($M = 3.37$, $SD = 0.76$) exhibited a smaller improvement. The mixed-design ANOVA results in Table 2 showed a significant main effect for Group [$F(1, 58) = 11.36$, $p < 0.001$, $\eta^2 = 0.64$], indicating that the AI-mediated language instruction had a substantial impact on enhancing self-regulated learning among EFL learners. Moreover, there was a significant main effect for Time [$F(1, 48) = 8.94$, $p < 0.001$, $\eta^2 = 0.24$], suggesting that both groups experienced an overall improvement in their self-regulation strategies over the 10-week period. Furthermore, the interaction effect between Group and Time was significant [$F(1, 58) = 7.68$, $p < 0.001$, $\eta^2 = 0.38$], indicating that the enhancement in self-regulated learning substantially differed between the experimental and control groups as a result of AI-mediated instruction.

Overall, the findings of this study demonstrate that the AI-mediated language instruction significantly enhanced English learning achievement, L2 motivation, and self-regulated learning among EFL learners. Although both groups experienced an increase in L2 achievement, motivation, and self-regulation over time, the

TABLE 1 Descriptive statistics of pre- and post-test scores of L2 achievement, L2 motivation, and self-regulation in both experimental and control groups.

		Pre-test		Post-test	
	Group	M	SD	M	SD
L2 Achievement	Experimental	43.21	12.69	73.86	15.26
	Control	44.39	13.21	61.11	14.97
L2 Motivation	Experimental	3.12	1.6	3.89	1.8
	Control	3.04	1.4	3.35	1.5
Self-regulation	Experimental	2.89	0.54	3.94	0.73
	Control	3.01	0.69	3.37	0.76

TABLE 2 Mixed-design ANOVAs, investigating the differences between the two groups' post-test scores on L2 achievement, L2 motivation, and self-regulation.

Source		Type III sum of squares	df	Mean square	F	Sig.	Partial eta squared
L2 achievement	Group	836.18	1	836.18	94.35	0.00	0.81
	Time	73.12	1	73.12	10.89	0.00	0.27
	Group × Time	56.54	1	56.54	8.37	0.00	0.23
L2 Motivation	Group	16.56	1	16.56	9.76	0.00	0.18
	Time	53.12	1	53.12	34.10	0.00	0.44
	Group × Time	10.91	1	10.91	7.43	0.04	0.11
Self-regulation	Group	82.54	1	82.54	11.36	0.00	0.64
	Time	21.12	1	21.12	8.94	0.00	0.24
	Group × Time	13.85	1	13.85	7.68	0.00	0.38

AI-mediated instruction showed a significant advantage over traditional instruction in terms of the dependent variables. These results highlight the potential of AI-based language learning platforms, such as the one used in the experimental group, to positively impact language learning outcomes, motivation, and foster self-regulated learning strategies among EFL learners.

4.2. Learners' perceptions

To gain a comprehensive understanding of the experiences and perceptions of students who participated in the AI-mediated language instruction, a qualitative phase was integrated into the study. Semi-structured interviews were conducted with 14 carefully selected students from the experimental group, ensuring a diverse range of perspectives and experiences. These interviews aimed to delve deeper into the nuances and subjective aspects of the students' engagement with the AI-mediated instruction, shedding light on the underlying mechanisms and contextual factors that influenced their English learning achievement, L2 motivation, and self-regulated learning.

4.2.1. Thematic analysis: emergent themes

Thematic analysis was employed to identify and analyze the emergent themes that encapsulated the rich insights shared by the interview participants. The themes that emerged from the interviews revealed profound insights into the impact of AI-mediated language instruction on the students' language learning journey. The following themes were identified.

4.2.2. Engaging and immersive learning experience

Participants were notably enthusiastic about the impact of AI-mediated language instruction, often describing it as a refreshing and captivating learning experience. They conveyed how the interactive nature of the AI platform's exercises, combined with dynamic language games and immediate feedback, created an immersive atmosphere that held their attention and made learning enjoyable. One participant shared, "Using the AI platform felt like I was actively participating in my learning. The interactive exercises were like puzzles I wanted to solve, and the real-time feedback kept me engaged."

The platform's ability to transform language learning into an interactive adventure was frequently highlighted. Students expressed that the AI platform's engaging features turned the learning process into a captivating journey, making it an experience they looked forward to. One participant noted, "Learning with the AI platform was exciting. It wasn't just about reading textbooks; it was like diving into a world of English where I could explore and interact."

4.2.3. Tailored and personalized learning pathway

Students recognized and valued the AI platform's capability to tailor the learning content to their individual strengths, weaknesses, and proficiency levels. They emphasized that the platform's personalized approach offered them a customized learning pathway that focused on their specific needs. This tailored instruction enabled them to address their challenges more effectively and capitalize on their strengths. One participant elaborated, "The AI platform did not treat everyone the same. It understood where I needed improvement and guided me accordingly."

This personalized learning experience contributed to participants' sense of autonomy and confidence in their language learning journey. They expressed how the platform's recognition of their learning style and preferences created an environment where they felt in control of their progress. As one other student pointed out, "The AI platform adapted to my pace and style. It pushed me when I was ready for a challenge and provided support when I needed it. It felt like I had a personalized tutor."

4.2.4. Significant improvement in language proficiency

Participants attributed their significant advancements in English language proficiency to the AI-mediated instruction. They highlighted how the interactive and adaptive features of the AI platform contributed to their enhanced skills across multiple language dimensions, including grammar, vocabulary, reading comprehension, and writing. Several students shared personal anecdotes of their progress, illustrating how their newfound language abilities positively impacted various aspects of their academic and personal lives.

The students' stories underscored how their improved language proficiency enabled them to excel in coursework that required English language skills, leading to higher grades and increased academic confidence. One participant recounted, "I used to struggle

with writing essays, but after using the AI platform, my writing skills improved. This semester, I received one of my highest essay scores."

4.2.5. Intrinsic motivation and active engagement

The AI-mediated instruction played a transformative role in cultivating intrinsic motivation among students and driving active engagement with the English language. Participants described a heightened sense of curiosity and enthusiasm that extended beyond the classroom setting. Many of them took proactive measures to seek out supplementary language resources and opportunities for practice. The AI platform, by fostering a genuine interest in English, inspired a passion for continuous exploration and learning.

The impact of this intrinsic motivation was evident in the participants' dedication to continuous language practice and their willingness to invest extra effort in their language learning journey. One participant expressed, "The AI platform ignited a spark in me. I wasn't just learning because I had to; I genuinely wanted to improve and discover more about the language."

4.2.6. Empowerment in self-regulated learning

Participants reported a notable sense of empowerment in their self-regulated learning practices, largely attributed to the guidance provided by the AI platform. They shared how the platform facilitated their use of various self-regulation strategies, such as setting clear learning goals, planning study sessions, and regularly monitoring their progress. The AI platform acted as a dependable companion, offering structure and tools that empowered participants to navigate their learning paths with confidence.

The participants' testimonials highlighted how the AI platform's guidance transformed them into more autonomous learners. One student reflected, "With the AI platform, I learned how to set specific goals and manage my learning time effectively. I felt more in control of my progress." Another student noted, "The AI platform helped me become a more independent learner. It taught me how to set goals and take ownership of my learning process."

4.2.7. Positive learning environment and alleviated anxiety

Participants felt that the AI-mediated instruction played a crucial role in cultivating an environment conducive to positive learning experiences. They emphasized how the platform's design encouraged them to step out of their comfort zones, take risks, and embrace mistakes without fear of judgment. This inclusive atmosphere was described as a catalyst for enhanced learning. One participant shared, "I used to hesitate to speak English, fearing I'd say something wrong. But with the AI platform's support, I felt safe to express myself. It really helped boost my confidence."

Moreover, the non-judgmental and constructive feedback provided by the AI platform was highlighted as a key factor in mitigating anxiety. Participants revealed that the platform's feedback not only pointed out their errors but also offered explanations and suggestions for improvement. This approach allowed them to learn from their mistakes and develop resilience in their language learning process. As one participant noted, "When I got something wrong, the AI would explain why and show me how to get it right. It made me feel like mistakes were just part of the learning process."

4.2.8. Flexibility and convenience for language practice

Participants enthusiastically highlighted the flexibility and convenience offered by the AI platform for practicing English. They appreciated the freedom to engage in language learning at their own pace, regardless of time or location. Many mentioned that they could seamlessly integrate language practice into their daily routines, such as during commutes or breaks. One participant shared, "I could practice English whenever I had a spare moment. It made learning feel like a part of my life, not an extra task."

The AI platform's accessibility beyond the traditional classroom setting was noted as a significant advantage. Participants emphasized that this flexibility allowed them to maintain consistent language practice even outside of formal study hours. They praised the platform for enabling them to stay engaged with English without the constraints of physical space. A participant explained, "With the AI platform, I did not have to wait for a class or find a specific place to study. It gave me the freedom to learn wherever I was."

The integration of the qualitative phase, with its in-depth interviews and thematic analysis, complemented the quantitative findings by providing rich, nuanced insights into the students' experiences with AI-mediated language instruction. The emergent themes highlighted the positive impact of the AI platform on their learning journey, emphasizing the importance of engaging and personalized instruction, intrinsic motivation, self-regulated learning, positive learning environments, and flexible language practice. These qualitative findings enhance the overall understanding of the effectiveness and benefits of AI-mediated language instruction for EFL learners, contributing to a more comprehensive and nuanced understanding of the phenomenon.

5. Discussion

This research investigated the efficacy of AI-assisted language learning instruction in enhancing English learning achievement, L2 motivation, and self-regulated learning among Chinese EFL learners. Drawing on Vygotsky's (1984) social constructivist theory of learning, a mixed-methods approach was employed for comprehensive data collection and analysis. The quantitative findings demonstrated a significant positive impact of AI-powered language learning instruction on the EFL learners' English learning achievement, L2 motivation, and self-regulated learning. Specifically, the AI-powered instruction was found to effectively improve all aspects of language learning achievement, including grammar, vocabulary, reading comprehension, and writing skills.

These quantitative results align with previous research conducted by Xu et al. (2022) and Zheng et al. (2021), who also reported positive outcomes of AI-powered language learning instruction on EFL learners' learning achievement. Additionally, the current study's findings are consistent with the research conducted by Hsu et al. (2023) and Utami et al. (2023), which highlighted the significant impact of AI-assisted language learning on language learners' vocabulary knowledge and writing performance, respectively.

In accordance with Vygotsky's (1984) social constructivist theory of learning, the AI learners were initially immersed in collaborative language learning activities facilitated by the AI-assisted language learning tool, with the primary goal of enhancing their language

learning achievement in grammar, vocabulary, reading comprehension, and writing skills. These collaborative activities, which encompassed elements of other-regulation, played a pivotal role in assisting learners in regulating their cognitive abilities related to language achievement, leading to the gradual internalization of their learning capabilities (Swain et al., 2015; Kim et al., 2018). In other words, the AI tool served as a catalyst, engaging learners in collaborative language learning activities that ultimately fostered the internalization of these skills (Zheng et al., 2021; Xu et al., 2022). While learners in the non-AI class also internalized their language learning abilities through interactions with peers in collaborative activities, the AI learners demonstrated superior performance due to their utilization of the AI-assisted language learning tool. The AI system expedited the transition from other-regulation to self-regulation of language learning abilities, enabling AI learners to reach their ZPD or potential level of functioning earlier than anticipated (Davis and Miyake, 2018). These findings align with the qualitative insights from the study, as learners corroborated the positive impact of the AI-assisted language learning tool on their independent functioning in language learning activities (Liang et al., 2021). The collaborative aspect of these activities, intertwined with the AI tool's support, played a crucial role in this transition.

Moreover, the student-centered nature of language learning activities adopted in both classes likely contributed to the observed outcomes. Both groups engaged in group-work activities, while the teacher assumed the role of facilitator, encouraging students to partake in collaborative language learning endeavors. These student-centered activities positively influenced the language achievement of EFL learners in both groups (Levine, 2004). However, the integration of the AI-assisted language learning tool in the AI learners' collaborative activities appeared to be more effective in enhancing their learning achievement, synergizing the benefits of collaborative activities with the adaptive supports provided by the AI tool. This can be attributed to the AI tool's ability to provide personalized feedback and comments on learners' language skills, which significantly impacted their language achievement (Kim et al., 2019; Liang et al., 2021; Chen et al., 2022). The AI system offered more effective and efficient language learning feedback, comments, and alternatives compared to other language learners, thus making a substantial contribution to learners' language achievement (Dodigovic, 2007; Kim et al., 2019). Although collaborative activities played a crucial role in language learning, it was the combination of collaborative activities and the adaptive supports provided by the AI tool that seemed to have the most significant impact on learners' language achievement. The AI learners also benefited from receiving more immediate and specific feedback tailored to their own written texts. These findings further align with the qualitative insights, where learners emphasized that the AI tool provided relevant feedback and facilitated their language skill improvement at their individual pace.

The high level of L2 motivation observed in AI learners may be attributed to their active engagement in interactive language learning activities facilitated by the AI system, enabling them to improve their skills at their own pace (Hwang et al., 2020; Liang et al., 2021). This finding aligns with previous research by Utami et al. (2023), who also reported strong motivation among language learners using AI-powered language learning activities. The qualitative findings of our study further corroborate this aspect, as participants emphasized the motivating language learning activities provided by

the AI tool. The user-friendly environment created by the AI tool not only reduced language learning anxiety but also enhanced learners' L2 motivation (Spiro et al., 2017; Hsu et al., 2023). By offering a personalized language learning environment free from time and space restrictions, the AI tool fostered a sense of autonomy and enthusiasm in learners, thereby contributing to heightened motivation (Carpio Cañada et al., 2015).

In addition, one other significant outcome of this study is the evidence supporting the enhancement of self-regulated learning among Chinese EFL learners through the utilization of AI-assisted language learning instruction. Self-regulated learning, characterized by learners' ability to independently plan, monitor, and evaluate their learning process, plays a pivotal role in fostering autonomy and lifelong learning skills (Zimmerman, 2000). The integration of AI-mediated instruction has proven to be instrumental in transforming the SRL landscape, offering learners adaptive tools that empower them to take charge of their own learning process (Xia et al., 2022b; Jin et al., 2023; Wang and Lin, 2023).

The qualitative and quantitative findings of this study converge to illuminate the positive impact of AI-assisted instruction on learners' self-regulatory abilities. Through the personalized nature of the AI platform, learners were presented with tailored learning content, real-time feedback, and opportunities for self-assessment. These features not only catered to learners' individual strengths and areas requiring improvement but also facilitated the development of their metacognitive and self-regulatory skills (Zimmerman, 1989). The AI tool acted as a dynamic guide, assisting learners in setting realistic goals, strategizing their study sessions, and reflecting on their progress, thereby cultivating their self-regulation capacities.

The thematic analysis of the qualitative data further underscores the enhancement of self-regulated learning attributed to the AI-assisted instruction. Participants vividly expressed how the AI platform empowered them to assume greater control over their learning process. They articulated using goal-setting techniques, tracking their own advancement, and adjusting their learning strategies based on the AI feedback. This increased sense of autonomy and ownership resonates with the core principles of self-regulated learning, as learners transitioned from being passive recipients to active agents in their language learning endeavors (Boekaerts and Corno, 2005; Fathi et al., 2021).

The findings of this study align with the broader educational research highlighting the interplay between technology-enhanced learning environments and self-regulated learning (Steffens, 2006; Johnson and Davies, 2014; Fathi et al., 2019; Lau and Jong, 2022). AI-mediated instruction acts as a scaffold, supporting learners as they navigate through the intricacies of self-regulation. The AI system provides learners with timely insights into their progress and performance, enabling them to make informed decisions regarding their learning strategies and goals (Jin et al., 2023; Wang and Lin, 2023).

The overall findings can be attributed to the efficacy and efficiency of the data generation by the AI tool compared to traditional language teachers' instruction. AI's ability to generate immediate and informative feedback proved invaluable to English language learners, addressing their weaknesses and providing timely guidance (Divekar et al., 2022). This prompt feedback mechanism not only supplied learners with essential information but also facilitated the improvement of their language learning skills. Additionally, the availability of written texts and materials for learners' reference further

consolidated their language learning progress, while the flexibility of accessing the materials anytime and anywhere amplified their learning experience (Junaidi, 2020; Hsu et al., 2023).

Also, the qualitative phase of this study, comprising semi-structured interviews with 11 students from the experimental group, provided a deeper understanding of the students' experiences and perceptions regarding AI-mediated language instruction. Thematic analysis revealed emergent themes that shed light on the multifaceted impact of the AI platform on English learning achievement, L2 motivation, and self-regulated learning among EFL learners. Overall, the findings from the semi-structured interviews align with the quantitative results, reinforcing the positive impact of AI-mediated instruction on language learning. The students' overwhelming enthusiasm for the engaging and interactive learning experience offered by the AI platform resonates with the observed higher levels of L2 motivation in the experimental group. The platform's ability to adapt and personalize learning content according to each student's needs and proficiency level aligns with the significant improvement in language proficiency demonstrated by the experimental group compared to the control group.

The emergent themes collectively emphasize the transformative nature of AI-mediated language instruction (Kim et al., 2019), supporting the existing literature regarding the potential advantages of AI in education in general (e.g., Malik et al., 2019; Chen et al., 2020; Luan et al., 2020; Ouyang and Jiao, 2021). The AI platform acted as more than just a language-learning tool; it became a catalyst for intrinsic motivation, empowerment in self-regulated learning, and the creation of a positive and supportive learning environment. Students reported feeling more at ease taking risks and making mistakes, fostering a growth mindset that further fueled their language learning journey. The interviews also shed light on the convenience and flexibility offered by the AI platform, allowing students to engage in consistent language practice beyond the confines of traditional classroom settings. This continuous and meaningful language practice is reflected in the experimental group's overall progress and higher English learning achievement compared to the control group.

The integration of qualitative insights enriches the discussion and strengthens the study's validity by offering a comprehensive view of the students' experiences with AI-mediated language instruction. The qualitative phase provides nuance and depth, capturing the subjective aspects of the learning journey that quantitative measures may not fully capture. Moreover, the integration of both qualitative and quantitative data enhances the study's credibility and contributes to a holistic understanding of the phenomenon.

Overall, the study's findings underscore the significant role of AI-mediated language instruction in enhancing L2 motivation and language learning skills. The interactive and personalized learning environment created by the AI tool empowers learners, reduces anxiety, and fosters intrinsic motivation. Moreover, the prompt feedback mechanism and easy access to learning materials contribute to the overall effectiveness and efficiency of AI-based language learning, making it a valuable asset in modern language classrooms.

The findings have significant implications for both EFL learners and teachers, shedding light on the potential impact of AI-assisted language learning tools in language classrooms.

For EFL learners, the study recommends embracing AI-powered language learning tools as valuable resources to enhance their language learning journey. These tools offer a more engaging and personalized learning experience, leading to increased motivation and enthusiasm for further language learning tasks. With AI-powered platforms like Duolingo, learners can access language practice anytime and anywhere, empowering them to take control of their learning process and fostering a sense of autonomy and self-regulation.

EFL teachers can also benefit from integrating AI-assisted language learning tools into their instructional strategies. These platforms provide valuable data and insights into each student's progress and areas of improvement, allowing teachers to tailor their instruction more effectively. By identifying specific language learning challenges, teachers can offer targeted support and guidance to individual students, leading to more efficient and effective language learning outcomes. The AI-supported instruction enables educators to differentiate their teaching approaches, catering to the diverse needs and abilities of their learners.

The study further highlights the potential of AI-powered language learning tools to revolutionize EFL classrooms by creating a personalized and adaptive learning environment. Learners engaging with interactive AI platforms receive instant feedback, comments, and alternative sentences, fostering continuous improvement and instilling confidence in their language abilities. This real-time support creates a supportive and encouraging atmosphere, promoting a positive learning experience. In addition, the integration of AI in language learning opens up exciting avenues for research and development in language pedagogy. As AI technology advances, educators and researchers can explore innovative ways to leverage its potential for improving language learning outcomes and instructional practices. Future studies may delve into the long-term effects of AI-assisted language learning, its applications in different language contexts, and the development of more sophisticated AI tools tailored to specific language learning needs.

6. Conclusion

The current study, following Vygotsky's social constructivism as its theoretical underpinning, examined the role of the AI-powered language learning instruction on Chinese EFL learners' English learning achievement, L2 motivation, and self-regulated learning. It was found that the AI-assisted language learning tool had a significant impact on the EFL learners' English learning achievement, L2 motivation, and self-regulated learning on the one hand and that the AI learners outperformed their non-AI counterparts in all the measures. The EFL learners were found to have positive perceptions toward the application of the AI-powered language learning tool in developing their English learning achievement, L2 motivation, and self-regulated learning. In general, the findings might be related to the effectiveness and efficiency of the AI tool in providing the EFL learners with immediate feedback and helping them personalize their language learning environment to be further engaged in the language learning activities.

Although this study has contributed valuable insights into the integration of AI in language instruction and its impact on English learning achievement, L2 motivation, and self-regulated learning, it is imperative to acknowledge several inherent limitations that should be taken into account when interpreting the results. Firstly, it is important to recognize that the research was conducted within a controlled educational setting, specifically focusing on a select group

of EFL learners from a single university. Consequently, caution must be exercised when attempting to generalize the findings to broader language learning contexts and more diverse learner populations. In order to enhance the external validity of the outcomes, future research endeavors should strive to encompass a more varied and representative sample of EFL learners. By doing so, the robustness and applicability of the results could be confirmed and extended across various educational environments.

Furthermore, the scope of this study primarily centered around the short-term effects of AI-mediated instruction, dictated by the constraints of the limited timeframe in which the research was conducted. To attain a more comprehensive and holistic understanding of the phenomenon, there is a clear imperative for the implementation of long-term follow-up studies. These extended studies would facilitate an assessment of the sustainability of the observed enhancements in language learning achievement, motivation, and self-regulated learning. The insights garnered from such investigations would be invaluable in illuminating the long-term efficacy of AI-powered language instruction and elucidating its potential implications for long-lasting language learning outcomes.

Additionally, it is noteworthy that the mixed-methods approach employed in this study, while yielding valuable insights into learners' experiences with AI-mediated instruction, was accompanied by a limitation related to the size of the qualitative sample. The qualitative phase of the study, involving a relatively small number of participants, presents an opportunity for improvement. A more comprehensive and nuanced understanding could be achieved by expanding the qualitative data collection process to encompass a larger and more diversified group of participants. This expansion would undoubtedly enrich the spectrum of perspectives and experiences related to the utilization of AI tools in language learning. The inclusion of a broader range of voices would provide researchers with an enhanced panorama of insights, ultimately contributing to a more robust and exhaustive analysis.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to LW, Email: Willing5216@163.com.

Ethics statement

Author contributions

LW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

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

References

Ahmadi, D. M. R. (2018). The use of technology in English language learning: a literature review. *Int. J. Res. English Educ.* 3, 115–125. doi: 10.29252/ijree.3.2.115

Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., et al. (2020). A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer* 53, 18–28. doi: 10.1109/MC.2020.2996587

Aldosari, S. A. M. (2020). The future of higher education in the light of artificial intelligence transformations. *Int. J. Higher Educ.* 9, 145–151. doi: 10.5430/ijhe.v9n3p145

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., and Mohammed, A. A. (2023). Impact of chat GPT on learning motivation: teachers and students' voices. *J. English Stud. Arabia Felix* 2, 41–49. doi: 10.56540/jesaf.v2i1.51

Balyen, L., and Peto, T. (2019). Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. *Asia Pac. J. Ophthalmol.* 8, 417–272. doi: 10.1097/01.APO.0000586388.81551.d0

Baskara, F. R., and Mukarto, F. X. (2023). Exploring the implications of Chat GPT for language learning in higher education. *Indonesian J. English Lang. Teach. Appl. Linguist.* 7, 343–358.

Boekaerts, M., and Corno, L. (2005). Self-regulation in the classroom: a perspective on assessment and intervention. *Appl. Psychol.* 54, 199–231. doi: 10.1111/j.1464-0597.2005.00205.x

Boo, Z., Dörnyei, Z., and Ryan, S. (2015). L2 motivation research 2005–2014: understanding a publication surge and a changing landscape. *System* 55, 145–157. doi: 10.1016/j.system.2015.10.006

Brown, J. M., Miller, W. R., and Lawendowski, L. A. (1999). "The self-regulation questionnaire" in *Innovations in clinical practice: A source book*. eds. L. Vande Creek and T. L. Jackson, vol. *17* (Sarasota, FL: Professional Resource Press/Professional Resource Exchange), 281–292.

Buzko, I., Dyachenko, Y., Petrova, M., Nenkov, N., Tuleninova, D., and Koeva, K. (2016). Artificial intelligence technologies in human resource development. *Comput. Model. New Technol.* 20, 26–29.

Carpio Cañada, J., Mateo Sanguino, T. J., Merelo Guervós, J. J., and Rivas Santos, V. M. (2015). Open classroom: enhancing student achievement on artificial intelligence through an international online competition. *J. Comput. Assist. Learn.* 31, 14–31. doi: 10.1111/jcal.12075

Chen, X., Xie, H., Zou, D., and Hwang, G. J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Comput. Educ. Artif. Intell.* 1:100002. doi: 10.1016/j.caeai.2020.100002

Chen, X., Zou, D., Xie, H., Cheng, G., and Liu, C. (2022). Two decades of artificial intelligence in education: contributors, collaborations, research topics, challenges, and future directions. *Educ. Technol. Soc.* 28, 1323–1358. doi: 10.1007/s10639-022-11209-y

Chun, D., Kern, R., and Smith, B. (2016). Technology in language use, language teaching, and language learning. *Mod. Lang. J.* 100, 64–80. doi: 10.1111/modl.12302

Davis, E. A., and Miyake, N. (2018). "Explorations of scaffolding in complex classroom systems" in *Scaffolding* (New York: Psychology Press), 265–272.

Devi, J. S., Sreedhar, M. B., Arulprakash, P., Kazi, K., and Radhakrishnan, R. (2022). A path towards child-centric artificial intelligence based education. *Int. J. Early Childhood* 14, 9915–9922. doi: 10.9756/INT-JECSE/V14I3.1145

Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., et al. (2022). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Comput. Assist. Lang. Learn.* 35, 2332–2360. doi: 10.1080/09588221.2021.1879162

Dodigovic, M. (2007). Artificial intelligence and second language learning: an efficient approach to error remediation. *Lang. Aware.* 16, 99–113. doi: 10.2167/la416.0

Ebadi, S., and Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: a case of CSIEC chatbot. *Interact. Learn. Environ.* 2022, 1–19. doi: 10.1080/10494820.2022.2096638

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: a case study. *Expert. Syst.* 38:e12667. doi: 10.1111/exsy.12667

Fang, T., Yang, S., Lan, K., Wong, D. F., Hu, J., Chao, L. S., et al. (2023). Is chatgpt a highly fluent grammatical error correction system? A comprehensive evaluation. *ar Xiv*. doi: 10.48550/arXiv.2304.01746

Fathi, J., Ahmadnejad, M., and Yousofi, N. (2019). Effects of blog-mediated writing instruction on L2 writing motivation, self-efficacy, and self-regulation: a mixed methods study. *J. Res. Appl. Linguist.* 10, 159–181. doi: 10.22055/RALS.2019.14722

Fathi, J., Arabani, A. S., and Mohamadi, P. (2021). The effect of collaborative writing using Google Docs on EFL learners' writing performance and writing self-regulation. *Lang. Relat. Res.* 12, 333–359. doi: 10.29252/LRR.12.5.12

Fathi, J., and Rahimi, M. (2022). Electronic writing portfolio in a collaborative writing environment: its impact on EFL students' writing performance. *Comput. Assist. Lang. Learn.* 2022, 1–39. doi: 10.1080/09588221.2022.2097697

Fitria, T. N. (2023). Artificial intelligence (AI) technology in Open AI Chat GPT application: a review of Chat GPT in writing English essay. *ELT J. English Lang. Teach.* 12, 44–58. doi: 10.15294/elt.v12i1.64069

Haristiani, N. (2019). Artificial intelligence (AI) chatbot as language learning medium: an inquiry. *J. Phys.* 1387:012020. doi: 10.1088/1742-6596/1387/1/012020

Hong, W. C. H. (2023). The impact of Chat GPT on foreign language teaching and learning: opportunities in education and research. *J. Educ. Technol. Innov.* 5, 37–45.

Hsu, T. C., Chang, C., and Jen, T. H. (2023). Artificial intelligence image recognition using self-regulation learning strategies: effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interact. Learn. Environ.* 1–19. doi: 10.1080/10494820.2023.2165508

Huang, J., and Tan, M. (2023). The role of Chat GPT in scientific communication: writing better scientific review articles. *Am. J. Cancer Res.* 13, 1148–1154.

Huang, X., Zou, D., Cheng, G., Chen, X., and Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educ. Technol. Soc.* 26, 112–131.

Hwang, G. J., Xie, H., Wah, B. W., and Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Comput. Educ.* 1:100001. doi: 10.1016/j.caeai.2020.100001

Ilkka, T. (2018). *The impact of artificial intelligence on learning, teaching, and education*. Maastricht: European Union.

Jin, S. H., Im, K., Yoo, M., Roll, I., and Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. *Int. J. Educ. Technol. High. Educ.* 20, 1–21. doi: 10.1186/s41239-023-00406-5

Johnson, G., and Davies, S. (2014). Self-regulated learning in digital environments: theory, research, praxis. *Br. J. Res.* 1, 1–14.

Junaidi, J. (2020). Artificial intelligence in EFL context: rising students' speaking performance with Lyra virtual assistance. *Int. J. Adv. Sci. Technol. Rehabil.* 29, 6735–6741.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., et al. (2022). Explainable artificial intelligence in education. *Comput. Educ. Artif. Intell.* 3:100074. doi: 10.1016/j.caeai.2022.100074

Kim, N. Y. (2019). A study on the use of artificial intelligence Chatbots for improving English grammar skills. *J. Digital Converg.* 17, 89–110.

Kim, S. G. (2023). Using chat GPT for language editing in scientific articles. *Maxillofacial Plastic Reconstr. Surgery* 45:13. doi: 10.1186/s40902-023-00381-x

Kim, N. J., Belland, B. R., and Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educ. Psychol. Rev.* 30, 397–429. doi: 10.1007/s10648-017-9419-1

Kim, N. Y., Cha, Y., and Kim, H. S. (2019). Future English learning: Chatbots and artificial intelligence. *Multimedia Assisted Lang. Learn.* 22, 32–53.

Knox, J. (2020). Artificial intelligence and education in China. *Learn. Media Technol.* 45, 298–311. doi: 10.1080/17439884.2020.1754236

Kohnke, L., Moorhouse, B. L., and Zou, D. (2023). Chat GPT for language teaching and learning. *RELC J.* 54, 537–550. doi: 10.1177/00336882231162868

Lai, Y., Saab, N., and Admiraal, W. (2022). University students' use of mobile technology in self-directed language learning: using the integrative model of behavior prediction. *Comput. Educ.* 179:104413. doi: 10.1016/j.compedu.2021.104413

Lau, K. L., and Jong, M. S. Y. (2022). Acceptance of and self-regulatory practices in online learning and their effects on the participation of Hong Kong secondary school students in online learning. *Educ. Inf. Technol.* doi: 10.1007/s10639-022-11546-y, [Online ahead of print]

Lee, J. H., Shin, D., and Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC J.* 54, 508–516. doi: 10.1177/00336882231165060

Lei, X., Fathi, J., Noorbakhsh, S., and Rahimi, M. (2022). The impact of mobile-assisted language learning on English as a foreign language learners' vocabulary learning attitudes and self-regulatory capacity. *Front. Psychol.* 13:872922. doi: 10.3389/fpsyg.2022.872922

Levine, G. S. (2004). Global simulation: a student-centered, task-based format for intermediate foreign language courses. *Foreign Lang. Ann.* 37, 26–36. doi: 10.1111/j.1944-9720.2004.tb02170.x

Liang, J. C., Hwang, G. J., Chen, M. R. A., and Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interact. Learn. Environ.* 31, 4270–4296. doi: 10.1080/10494820.2021.1958348

Loncar, M., Schams, W., and Liang, J. S. (2023). Multiple technologies, multiple sources: trends and analyses of the literature on technology-mediated feedback for L2 English writing published from 2015-2019. *Comput. Assist. Lang. Learn.* 36, 722–784. doi: 10.1080/09588221.2021.1943452

Lu, X. (2018). "Natural language processing and intelligent computer-assisted language learning (ICALL)" in *The TESOL encyclopedia of English language teaching*. ed. J. I. Liontas (Chichester: Wiley Blackwell), 1–6.

Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J., Ogata, H., et al. (2020). Challenges and future directions of big data and artificial intelligence in education. *Front. Psychol.* 11:580820. doi: 10.3389/fpsyg.2020.580820

Malik, G., Tayal, D. K., and Vij, S. (2019, 2017). "An analysis of the role of artificial intelligence in education and teaching" in *Recent findings in intelligent computing techniques: Proceedings of the 5th ICACNI*. eds. P. Sa, S. Bakshi, I. Hatzilygeroudis and M. Sahoo, vol. *1* (Singapore: Springer), 407–417.

Mehdiyev, E., Usta, H. G., and Ugurlu, C. (2017). Validity and reliability study: motivation scale in English. *Int. J. Soc. Sci.* 54, 21–37.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (2013). *Machine learning: an artificial intelligence approach*. Berlin: Springer Science & Business Media.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: the critical factors of learner engagement and self-regulation. *Appl. Linguis.* 35, 418–440. doi: 10.1093/applin/amu012

Nemorin, S., Vlachidis, A., Ayerakwa, H. M., and Andriotis, P. (2023). AI hyped? A horizon scan of discourse on artificial intelligence in education (AIED) and development. *Learn. Media Technol.* 48, 38–51. doi: 10.1080/17439884.2022.2095568

Nguyen, L. Q., and Le, H. V. (2023). Enhancing L2 learners' lexical gains via Quizlet learning tool: the role of individual differences. *Educ. Inf. Technol.* 28, 12143–12167. doi: 10.1007/s10639-023-11673-0

Ouyang, F., and Jiao, P. (2021). Artificial intelligence in education: the three paradigms. *Comput. Educ. Artif. Intell.* 2:100020. doi: 10.1016/j.caeai.2021.100020

Pedro, F., Subosa, M., Rivas, A., and Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*. Paris: UNESCO.

Pikhart, M. (2020). Intelligent information processing for language education: the use of artificial intelligence in language learning apps. *Proc. Comput. Sci.* 176, 1412–1419. doi: 10.1016/j.procs.2020.09.151

Rahimi, M., and Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: a mixed methods study. *Comput. Assist. Lang. Learn.* 35, 2627–2674. doi: 10.1080/09588221.2021.1888753

Rodinadze, S., and Zarbazoia, K. (2012). The advantages of information technology in teaching English language. *Front. Lang. Teach.* 3, 271–275.

Roll, I., and Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *Int. J. Artif. Intell. Educ.* 26, 582–599. doi: 10.1007/s40593-016-0110-3

Schmidt-Fajlik, R. (2023). Chat GPT as a grammar checker for Japanese English language learners: a comparison with Grammarly and pro writing aid. *Asia CALL Online J.* 14, 105–119. doi: 10.54855/acoj.231417

Shadiev, R., Wen, Y., Uosaki, N., and Song, Y. (2023). Future language learning with emerging technologies. *J. Comput. Educ.* 10, 463–467. doi: 10.1007/s40692-023-00285-9

Shadiev, R., and Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability* 12:524. doi: 10.3390/su12020524

Shatri, Z. G. (2020). Advantages and disadvantages of using information technology in learning process of students. *J. Turk. Sci. Educ.* 17, 420–428. doi: 10.36681/tused.2020.36

Soleimani, H., Mohammaddokht, F., and Fathi, J. (2022). Exploring the effect of assisted repeated reading on incidental vocabulary learning and vocabulary learning self-efficacy in an EFL context. *Front. Psychol.* 13:851812. doi: 10.3389/fpsyg.2022.851812

Spector, J. M., and Ma, S. (2019). Inquiry and critical thinking skills for the next generation: from artificial intelligence back to human intelligence. *Smart Learn. Environ.* 6, 1–11. doi: 10.1186/s40561-019-0088-z

Spiro, R. J., Bruce, B. C., and Brewer, W. F. (2017). *Theoretical issues in reading comprehension: perspectives from cognitive psychology, linguistics, artificial intelligence and education.* Hillsdale, NJ: Routledge.

Steffens, K. (2006). Self-regulated learning in technology-enhanced learning environments: lessons of a European peer review. *Eur. J. Educ.* 41, 353–379. doi: 10.1111/j.1465-3435.2006.00271.x

Su, Y., Lin, Y., and Lai, C. (2023). Collaborating with chat GPT in argumentative writing classrooms. *Assess. Writ.* 57:100752. doi: 10.1016/j.asw.2023.100752

Suryana, I., Asrianto, A., and Murwantono, D. (2020). Artificial intelligence to master English listening skills for non-English major students. *J. Lang. Lang. Teach.* 8, 48–59. doi: 10.33394/jollt.v8i1.2221

Swain, M., Kinnear, P., and Steinman, L. (2015). *Sociocultural theory in second language education: an introduction through narratives.* Bristol: Multilingual Matters.

Tafazoli, D., María, E. G., and Abril, C. A. H. (2019). Intelligent language tutoring system: integrating intelligent computer-assisted language learning into language education. *Int. J. Inform. Commun. Technol. Educ.* 15, 60–74. doi: 10.4018/IJICTE.2019070105

Utami, S. P. T., Andayani, Winarni, R., and Sumarwati. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive? *Contemp. Educ. Technol.* 15:ep450. doi: 10.30935/cedtech/13419

Vygotsky, L. S. (1984). "The problem of age" in *The collected works of L. S. Vygotsky. Vol. 5. Child psychology.* ed. R. W. Rieber (New York: Plenum)

Vygotsky, L. S., and Cole, M. (1978). *Mind in society: development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Vygotsky, L. S., Rieber, R. W., and Hall, M. J. (1997). *The collected works of LS Vygotsky: the history of the development of higher mental functions.* New York: Plenum Press.

Wang, C. Y., and Lin, J. J. (2023). Utilizing artificial intelligence to support analyzing self-regulated learning: a preliminary mixed-methods evaluation from a human-centered perspective. *Comput. Hum. Behav.* 144:107721. doi: 10.1016/j.chb.2023.107721

Weng, X., and Chiu, T. K. (2023). Instructional design and learning outcomes of intelligent computer assisted language learning: systematic review in the field. *Comput. Educ. Artif. Intell.* 4:100117. doi: 10.1016/j.caeai.2022.100117

Williams, E. N., and Morrow, S. L. (2009). Achieving trustworthiness in qualitative research: a pan-paradigmatic perspective. *Psychother. Res.* 19, 576–582. doi: 10.1080/10503300802702113

Xia, Q., Chiu, T. K., and Chai, C. S. (2022a). The moderating effects of gender and need satisfaction on self-regulated learning through artificial intelligence (AI). *Educ. Inf. Technol.* 28, 8691–8713. doi: 10.1007/s10639-022-11547-x

Xia, Q., Chiu, T. K., Lee, M., Sanusi, I. T., Dai, Y., and Chai, C. S. (2022b). A self-determination theory (SDT) design approach for inclusive and diverse artificial intelligence (AI) education. *Comput. Educ.* 189:104582. doi: 10.1016/j.compedu.2022.104582

Xu, X., Dugdale, D. M., Wei, X., and Mi, W. (2022). Leveraging artificial intelligence to predict young learner online learning engagement. *Am. J. Dist. Educ.* 37, 185–198. doi: 10.1080/08923647.2022.2044663

Yan, D. (2023). Impact of Chat GPT on learners in a L2 writing practicum: an exploratory investigation. *Educ. Inf. Technol.* 1-25, 1–25. doi: 10.1007/s10639-023-11742-4

Zhang, R., and Zou, D. (2020). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Comput. Assist. Lang. Learn.* 35, 696–742. doi: 10.1080/09588221.2020.1744666

Zheng, L., Niu, J., Zhong, L., and Gyasi, J. F. (2021). The effectiveness of artificial intelligence on learning achievement and learning perception: a meta-analysis. *Interact. Learn. Environ.* 1-15, 1–15. doi: 10.1080/10494820.2021.2015693

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *J. Educ. Psychol.* 81, 329–339. doi: 10.1037/0022-0663.81.3.329

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, CA: Academic Press), 13–39.

Check for updates

A structural equation model of online learning: investigating self-efficacy, informal digital learning, self-regulated learning, and course satisfaction

Yi Zheng[1]* and Ao Xiao[2]

[1]School of Foreign Languages and Literatures, Chongqing Normal University, Chongqing, China, [2]NYU Steinhardt Teachers of English to Speakers of Other Languages, New York University, New York, NY, United States

Introduction: English as a Foreign Language (EFL) education increasingly relies on online learning, necessitating a nuanced understanding of crucial factors impacting learning experiences. This research investigates the intricate relationships among online learning self-efficacy, online self-regulated learning, informal digital learning of English (IDLE), and online course satisfaction within the unique context of EFL learners.

Methods: The study involved 563 intermediate college students from various national universities in China. Structural Equation Modeling (SEM) was employed to analyze the data, providing comprehensive insights into the relationships among the identified variables.

Results: The results revealed significant insights. Both online learning self-efficacy and IDLE exhibited direct and positive influences on online course satisfaction. Furthermore, the study uncovered that online self-regulated learning acted as a partial mediator in the connection between online learning self-efficacy and IDLE with online course satisfaction. This mediation implies that learners' self-regulatory behaviors significantly affect how self-efficacy and informal digital language learning experiences impact overall satisfaction with online courses.

Discussion: The findings highlight the pivotal role of nurturing learners' self-efficacy beliefs, fostering IDLE, and promoting effective self-regulated learning strategies in the realm of online language learning. These initiatives are instrumental in enhancing learners' satisfaction and success in online courses.

Conclusion: The implications of these findings for EFL instruction are substantial. By emphasizing the importance of self-efficacy, IDLE, and self-regulated learning strategies, educators can significantly contribute to creating more satisfying and successful online learning experiences for EFL students.

KEYWORDS

online learning self-efficacy, online self-regulated learning, informal digital learning of English (IDLE), online course satisfaction, EFL learners, structural model

1 Introduction

The 21st century has witnessed a swift transition toward digital information and media, significantly enhancing the role of information and communication technology (ICT) across diverse sectors, including education and language learning (Fernández-Gutiérrez et al., 2020; Rahimi and Fathi, 2022; Al-Rahmi et al., 2023; Bolaji and Jimoh, 2023). This surge in technology has become integral to educational frameworks, leading to a substantial rise in online courses and programs, consequently escalating enrollment rates (Arrosagaray et al., 2019; Abraham et al., 2022; Adilbayeva et al., 2022; Fathi et al., 2023).

Online courses play a pivotal role in augmenting learning experiences, particularly in the domain of English as a Foreign Language (EFL), known for providing engaging and authentic learning materials that significantly enhance EFL students' academic performance (Liu et al., 2022; Asratie et al., 2023; Jiang et al., 2023; Yang and Xu, 2023).

Despite widespread integration of e-learning in second language acquisition (SLA), challenges persist regarding students' affective and psychological states in online learning. While some learners find online education as effective as traditional classes (Sagarra and Zapata, 2008), others may not find the online environment satisfactory (Dizon and Thanyawatpokin, 2021). Learner satisfaction serves as a framework for evaluating and enhancing online learning, representing students' perceptions of how well their learning requirements and goals are met (Hew et al., 2020). Acknowledged as a pivotal factor in determining the effectiveness of online courses, student satisfaction correlates with their performance and experience in virtual learning (Elshami et al., 2021; Abdelrady and Akram, 2022). Research indicates that students' satisfaction in online courses significantly correlates with their academic success (Wei and Chou, 2020; Kim and Kim, 2021; She et al., 2021). Particularly in EFL contexts, online course satisfaction links to heightened engagement, determination, and reduced dropout rates, fostering motivation (Chen and Adesope, 2016; Fujii et al., 2022). Online courses differ from face-to-face settings, altering student-instructor and peer interactions (de Jong et al., 2020; Fathi and Ebadi, 2020; Wu et al., 2023). Neglecting students' psychological needs in virtual environments may diminish motivation and confidence in pursuing language learning goals, potentially leading to decreased satisfaction and adverse academic outcomes (Abdullah, 2022; Ji et al., 2022).

In online language learning, several factors significantly influence learners' experiences. Notably, learners' self-efficacy, self-regulation, and engagement in informal digital learning of English (IDLE) play crucial roles (Lee, 2019; Lee and Sylvén, 2021). IDLE, a subset of computer-assisted language learning (CALL), attracts digital-native EFL learners and involves informal language learning through computer-based technologies like smartphones, computers, social media, and blogs (Lai, 2019; Soyoof et al., 2023). These self-directed, unstructured experiences offer authentic language learning opportunities (Liu et al., 2023). Research underscores the positive impact of IDLE on EFL students, enhancing language proficiency, cultural awareness, and communication skills (Lee, 2021; Lee et al., 2022; Liu et al., 2023). Engaging in IDLE activities contributes to improved language learning satisfaction by fostering proficiency and cultural understanding.

Despite growing interest in student satisfaction, little research has probed into its intricacies within L2 online courses, especially in China's unique context (Gyamfi and Sukseemuang, 2018). This dearth of exploration emphasizes the urgency for further investigation in this critical area. Our study takes a pioneering step to uncover the pivotal roles of online learning self-efficacy, online self-regulated learning, and IDLE in predicting satisfaction among Chinese EFL students in the dynamic virtual classroom environment. It fills a void in the field by exploring these multifaceted elements, contributing substantially to the understanding of online language learning. Moreover, it specifically delves into the dynamics of Chinese EFL learners, an area scarcely examined before. Beyond theoretical advancement, this research carries significant practical implications, paving the way for targeted techniques and interventions to enhance learner satisfaction and success in the online language learning realm.

2 Literature review

2.1 Theoretical framework

This study is firmly anchored in the self-determination theory (SDT) proposed by Deci and Ryan (1985). SDT delves into the motivations and self-regulation of individuals across diverse contexts, with a specific emphasis on education. Central to SDT is the notion that individuals possess inherent psychological needs for autonomy, competence, and relatedness, which serve as prime drivers of motivation and exert a profound influence on their behaviors (Ryan and Deci, 2000).

Autonomy addresses the imperative for learners to experience a sense of choice and volition in their learning journey (Deci and Ryan, 2000). In the context of this study, it translates to learners feeling empowered to take charge of their online learning experiences, making informed choices about their learning activities, and setting their own educational objectives (Roca and Gagné, 2008). The *competence* facet of SDT revolves around learners' intrinsic need to perceive themselves as effective and proficient in their learning pursuits (Deci and Ryan, 2015). Within the realm of online learning, learners with a robust sense of competence believe in their capacity to proficiently navigate digital platforms, actively engage in self-regulated learning practices, and ultimately accomplish their learning goals. Furthermore, *relatedness* underscores the importance of social connections and a sense of belonging within the learning environment (Deci and Ryan, 2000). In the online domain, this encompasses learners feeling a genuine connection to their peers, instructors, and the broader learning community.

In the context of online learning, learners' *satisfaction* is intricately tied to the fulfillment of their fundamental psychological needs (Chen and Jang, 2010). When the online learning environment supports their autonomy (granting choice and control over their learning), competence (fostering perceived effectiveness and mastery), and relatedness (facilitating social interaction and connection), learners are more likely to experience satisfaction. Concerning *online self-regulated learning*, SDT provides valuable

insights into the development of self-regulated learning skills within online environments (Hsu et al., 2019). Learners who perceive autonomy in setting their learning goals and strategies, competence in their ability to manage their learning, and relatedness with their peers and instructors are more likely to engage in effective self-regulated learning. In the domain of *online learning self-efficacy*, self-efficacy beliefs, a pivotal component of SDT, are closely intertwined with perceived competence (Chiu, 2022). Learners with a strong sense of self-efficacy in the online learning context are more inclined to believe in their capacity to succeed and are motivated to invest sustained effort and persistence in their learning endeavors. Applying SDT to IDLE, learners engaging in informal digital learning activities do so autonomously, driven by their choice to explore resources that pique their interest (Huang et al., 2019). This can significantly enhance their competence in the language and their sense of relatedness to the global online community of language learners.

By anchoring our study in SDT, we establish a unified theoretical framework that elucidates the intricate interplay between course satisfaction, self-regulated learning, learning self-efficacy, and IDLE among EFL learners in online learning contexts. This approach imbues the study with cohesion and focus, enabling a more comprehensive understanding of the interconnected constructs.

2.2 Online course satisfaction

Satisfaction in education refers to the pleasure derived from achieving desired learning objectives (Elliott and Shin, 2002; Wiers-Jenssen et al., 2002). It represents students' subjective judgment of the learning process, considering how well the learning environment supports their academic success (Pham et al., 2019; Karbakhsh and Ahmadi Safa, 2020; Ngah et al., 2022). Lim et al. (2020) describe satisfaction as the perceived alignment between students' expectations and actual outcomes in their educational experience. Studies such as Pangarso and Setyorini (2023) have shown that higher course satisfaction correlates with better mastery of content. Particularly in EFL learning, student satisfaction correlates with achieving language learning goals, thereby enhancing motivation and performance (Rashidi and Moghadam, 2014; Qutob, 2018; Tsai, 2019).

It would also be worthwhile to examine satisfaction in online environments investigation due to its peculiar significance to the motivations and aspirations of learners. It should be bear in mind that student satisfaction with the online course is an integral factor to be delt with for determining quality and effectiveness of an online learning (Elia et al., 2019; Martin and Bolliger, 2022; Torrado and Blanca, 2022). As Moore (2005) ideated, there are five pillars to assess the effectiveness of online learning, namely faculty satisfaction, access, learning effectiveness and cost effectiveness, and more importantly student satisfaction which is conceptualized to be the most important one. Online course satisfaction has a central role in shaping learners' perceptions regarding instructional quality, as well as their tendencies to participate in future online course learning (Nikou and Maslov, 2023; Paposa and Paposa, 2023). The available evidence suggests that high levels of online satisfaction appear to significantly influence learners' academic

outcomes in the classroom (Abdous, 2019; Daneji et al., 2019; Wu et al., 2022). Unlike students with low online satisfaction, satisfied learners are more likely to be passion and to invest extra effort in online learning. Importantly, high online satisfaction can be significantly conducive to online course completion rates and can encourage learner to be more committed to learning and more motivated to achieve their learning objectives (Kuo et al., 2014; Nashaat et al., 2021; Ranadewa et al., 2021). More particularly, in the context of EFL, Zou et al. (2022) noted that online course satisfaction is positively associated with L2 learners' adjustment to virtual environment, the utility of the platforms, the experienced enjoyment, the ease of use of the platform, and the effectiveness of the platform.

Previous research across various educational contexts, including EFL learning and teaching, highlights the significance of online course satisfaction. For instance, Jiang et al. (2021) found that perceived usefulness and ease of use significantly influenced satisfaction among 928 Chinese students. Bervell et al. (2020) identified positive correlations between different types of interactions—learner-learner, learner-teacher, and learner-material—and learners' satisfaction. In the field of SLA, Shih et al. (2013) observed a relationship between online course satisfaction, online learning motivation, and Big Five personality traits in a study involving 53 tertiary-level EFL learners.

Taken together, albeit the bulk of studies that have examined the potential indicators of learners' online learning satisfaction, relatively fewer studies have touched upon the impact of psychological factors and online learner-related variables on online course satisfaction. Moreover, the exploration of the antecedents of online course satisfaction in the context of EFL is still lagging behind the dynamic sphere and more in-depth qualitative studies are required to increase our understanding of this construct.

2.3 Online self-regulated learning

With roots in psychology, self-regulation has been conceptualized by Bandura (1988) as the form of three attributions of cognitive incentives, namely causal ascriptions, outcome expectancies, and cognized objectives. Appropriately, self-regulation has been gradually barrowed from the realm of psychology and has been extended into the educational context, as a result, the concept of self-regulated learning has been introduced for student learning (Jansen et al., 2019; Wolters and Brady, 2020).

In essence, self-regulation entails learners' kind of process and skills that help them plan and arrange the tasks in the learning context, manage time, ask for support of others (i.e., teachers and peers), and see if the long-term objectives were attained (Posner and Rothbart, 2000; Landrum, 2020). As put forward by Schunk and Zimmerman (2007), self-regulation has to do with learners' capacity to utilize self-managing behaviors and incorporate learning processes, and to merge these with motivation so that learners could be able to use their self-confidence behaviors in the classroom. According to Zimmerman (1989), self-regulated learning pertains to students' effective engagement within the learning process in four processes of cognitive (i.e., involving the kind of procedure used by learners effectively master the content), metacognitive (i.e., referring to students' capability to scheme

plans, strategies or objectives as an attempt to assess and appraise their own performance and learning), motivational (i.e., focusing on the notion whether students are self-motivated and willing to independently be held as accountable for their own successes or failures), and behavioral (i.e., normally conceived as learners' seeking help from peers or teachers to facilitate the learning process).

In online environments, students who are equipped with self-regulatory skills have better chances to achieve their goals, as self-regulated learning can significantly improve students' capacity to exercise self-control, self-monitor and self-evaluate in virtual classrooms (Öztürk and Çakıroğlu, 2021; Lei and Lin, 2022; Saint et al., 2022). To be particular, for EFL learners, self-regulated learning has significant advantages in terms of increasing their self-direction and self-management and encourage them to be independent and autonomous when engaged in online activities (Guo et al., 2021; Teng, 2021; Chen et al., 2023; Shen et al., 2023). As long as L2 learning is considered, given that it can facilitate the dynamic personality of students' interactions, self-regulated learning is often considered to be one of critical learning strategies for blended learning (Xu et al., 2022).

Previous studies have highlighted the pivotal role of self-regulated learning in enhancing satisfaction and academic outcomes (Kuo et al., 2014; Ejubovic and Puška, 2019; Zalli et al., 2019; Lim et al., 2020; Wu et al., 2023). For instance, Zalli et al. (2019) and Lim et al. (2020) observed a positive impact of self-regulated learning on online course satisfaction. Wu et al. (2023) also identified self-regulated learning as a contributing factor to learners' satisfaction with online courses. Ejubovic and Puška (2019) discovered a positive association between self-regulated learning, academic performance, and online course satisfaction. However, Landrum (2020) reported conflicting results, as their study did not find self-regulation strategies to significantly predict online learning satisfaction among students.

Despite the crucial role of self-regulated learning in shaping learners' satisfaction in online environments, there exist scant empirical evidence that delves into association between these two constructs in the realm of SLA. Bearing this in mind, we aimed to investigate the potential role of self-regulated learning in affecting online course satisfaction among EFL learners.

2.4 Online learning self-efficacy

Self-efficacy refers to individuals' beliefs in their ability to perform tasks effectively for desired outcomes (Bandura, 1982). Rooted in social cognitive theory, self-efficacy underscores motivation's role in learning and performance (Schunk, 1995). It emphasizes how learners acquire beliefs, information, and skills within specific contexts (Al-Abyadh and Abdel Azeem, 2022). People's self-efficacy significantly influences their actions; those with high self-efficacy are more resilient, less likely to avoid challenges, and perceive difficulties as growth opportunities (Cassidy, 2015; Supervia et al., 2022). In the realm of language learning, EFL learners with higher self-efficacy levels tend to set and persist in achieving demanding goals (Dong et al., 2022; Derakhshan and Fathi, 2023).

2.4.1 Top of form

Like traditional classrooms, self-efficacy has a paramount role in online environments. It is often argued that self-efficacy can play a central role in students' online learning as it might have great contributions to their success in virtual contexts (Zimmerman and Kulikowich, 2016; Tang et al., 2022). Furthermore, Quang et al. (2022) demonstrated that those EFL learners who are equipped with online self-efficacy tend to be more successful in online environments. More importantly, previous researchers have established that online self-efficacy has been related to students' wellbeing and success (i.e., online course satisfaction) (e.g., Shen et al., 2013; Suryandani and Santosa, 2021; Aldhahi et al., 2022; Derakhshesh et al., 2022). Aldhahi et al. (2022) aimed to evaluate students' online learning satisfaction and to examine whether online-learning self-efficacy correlate with satisfaction in remote learning. Collecting data from a total of 1,226 learners, their results revealed that online learning self-efficacy had a positive role in influencing online course satisfaction. In the same vein, Shen et al. (2013) reported that that online learning self-efficacy positively predicted online learning satisfaction among students. The association between online learning self-efficacy and online course satisfaction was the focus of Derakhshesh's et al. (2022) study. Utilizing structural equation modeling, the authors indicated that online learning self-efficacy could exercise positive effects on online course satisfaction among EFL learners. In a mixed-method research design, Suryandani and Santosa (2021) examined the association between online learning self-efficacy and students' satisfaction with online courses. The results indicated a positive relationship, suggesting that students' satisfaction with online courses could be enhanced by their online learning self-efficacy.

What is found from the literature review is that despite a number of research has been done on the interplay between online self-efficacy and online satisfaction in general education context, the study of the potential correlation between these constructs in the EFL classrooms merits further exploration.

2.5 IDLE

Informal digital learning of English (IDLE) is a phenomenon which is established in the context of CALL and has recently attracted the attention of the digital-native generation worldwide (Lee and Sylvén, 2021). In fact, IDLE is grounded in incidental language learning (Miller and Godfroid, 2020), and informal language learning (Bahrani and Sim, 2012), and is often recognized as a sub-field of CALL in the context of L2 learning and teaching. Researchers often refer to IDLE as an informal English learning concept that provides a self-directed platform and encompasses different types of computer-based technologies and web-based resources like smart phones, computers, social media, and blogs in an informal context (Lai, 2019; Soyoof et al., 2023). In fact, IDLE is recognized as self-directed and situated in the out-of-class due to the fact that students are the ones that first initiate IDLE activities (Liu et al., 2023). IDLE provides a kind of opportunity by which students can learn a language in an unstructured and naturalistic manner and not in a process in the service of certification (Lee et al., 2023).

It is argued that IDLE can be either form-focused (i.e., merely dealing with linguistic elements and the language accuracy) or

meaning-focused (i.e., focusing on the authentic use of language) (Zhang and Liu, 2022). Previous findings from existing literature indicates that IDLE is significantly conducive to contribute L2 learning and teaching. For instance, collecting data from a sample of 1,490 Chinese EFL learners, Liu et al. (2023) revealed that IDLE activities had positive effects on language and culture learning, cultural differences, intercultural communication, and engagement among EFL students. In another study, Lee et al. (2022) found that IDLE could predict students' willingness to communicate in L2. Notwithstanding these researches, the investigation of the relationship between IDLE and online course satisfaction has remained non-examined in the EFL domain.

Despite the recent attention, the research on IDLE is still very limited and further investigation are required to shed light on the role of this construct in language learning. Furthermore, its correlation with other psychological learner-related variables in online environments has not been examined by the L2 researchers. In fact, albeit the array of research foci, to our knowledge, our research is the first attempt to examine the relations between IDLE and online course satisfaction, online self-regulated learning, and online learning self-efficacy among EFL learners. Consequently, we sought to examine whether IDLE has any significant effects on online course satisfaction via the mediation role of online self-regulated learning in the EFL setting.

2.6 The present study

While a significant body of literature has highlighted the positive associations between self-efficacy, informal digital learning, self-regulated learning, and course satisfaction (Lim et al., 2020; Aldhahi et al., 2022; Lee et al., 2023; Wu et al., 2023), the existing research landscape also encompasses contrasting or nuanced findings.

For instance, amidst the consensus on the beneficial impact of self-regulated learning on online course satisfaction, Landrum's (2020) study revealed contrasting results. Their findings did not identify self-regulation strategies as significant predictors of online learning satisfaction among students. This discrepancy underscores the imperative for deeper explorations into the nuanced aspects of self-regulated learning that might differentially contribute to learners' course satisfaction.

Moreover, despite considerable evidence supporting the positive impact of IDLE on language proficiency and cultural understanding, certain perspectives highlight potential limitations or contextual dependencies. Zhang and Liu (2022) underscored distinctions between form-focused and meaning-focused IDLE activities, suggesting that their effectiveness might be contingent upon learners' preferences or specific learning contexts.

Furthermore, within the domain of self-efficacy, while several studies accentuate its positive association with online learning outcomes, a subset of research presents more nuanced perspectives. For instance, Tang et al. (2022) established a significant positive correlation between online self-efficacy and students' success in online environments, whereas Zimmerman and Kulikowich (2016) observed variations in the impact of self-efficacy based on individual differences and task characteristics.

Acknowledging these divergent or nuanced findings underscores the multifaceted nature of the relationships among self-efficacy, informal digital learning, self-regulated learning, and course satisfaction. The complexity of these relationships seems contingent upon individual differences, contextual factors, and specific learning environments. Thus, comprehensive investigations considering these intricacies are crucial to obtain a deeper understanding of their interplay.

Against this backdrop, the purpose of this study was to empirically test a hypothesized model (see **Figure 1**) that examines the interrelationships among online learning self-efficacy, informal digital learning of English, online self-regulated learning, and their collective impact on online course satisfaction. The following hypotheses will guide this study:

2.6.1 H1: Online learning self-efficacy directly predicts online course satisfaction

Online learning self-efficacy (OLSE) is concerned with learners' beliefs in their ability to effectively perform online learning tasks and achieve desired outcomes (Bandura, 1982). OLSE is particularly crucial in shaping learners' motivation and performance in virtual environments (Zimmerman and Kulikowich, 2016; Tang et al., 2022). Previous research has demonstrated a positive association between OLSE and students' success and wellbeing in online learning, including online course satisfaction (Shen et al., 2013; Suryandani and Santosa, 2021; Aldhahi et al., 2022; Derakhshesh et al., 2022).

Students with higher OLSE are more likely to set challenging tasks and objectives and persevere in their efforts (Dong et al., 2022). As a result, they are less likely to avoid demanding tasks and view them as opportunities for growth and development (Cassidy, 2015; Supervia et al., 2022). In the context of EFL learners, higher self-efficacy beliefs have been associated with a willingness to engage in challenging language learning activities and a commitment to achieving language proficiency goals (Dong et al., 2022).

2.6.2 H2: Informal digital learning of English directly predicts online course satisfaction

IDLE involves learners' self-initiated activities using computer-based technologies and web-based resources in an informal context (Lai, 2019; Lee, 2021; Soyoof et al., 2023). Previous research has demonstrated the positive impact of IDLE on language learning outcomes, cultural awareness, and communication skills (Lee and Sylvén, 2021; Lee et al., 2022; Liu et al., 2023). Given the positive associations of IDLE with language learning outcomes and engagement, it is reasonable to hypothesize that learners who engage in IDLE activities may experience higher satisfaction with their online courses.

2.6.3 H3: Online self-regulated learning directly predicts online course satisfaction

Online self-regulated learning (SRL) is rooted in psychology and has been extended into the educational context, particularly in the domain of self-regulated learning (Jansen et al., 2019; Wolters and Brady, 2020; Theobald, 2021). In virtual classrooms, learners need to exercise self-control, self-monitoring, and self-evaluation to effectively engage in the learning process (Öztürk and Çakıroğlu, 2021; Lei and Lin, 2022; Saint et al., 2022). The positive influence of self-regulated learning on online course satisfaction has been
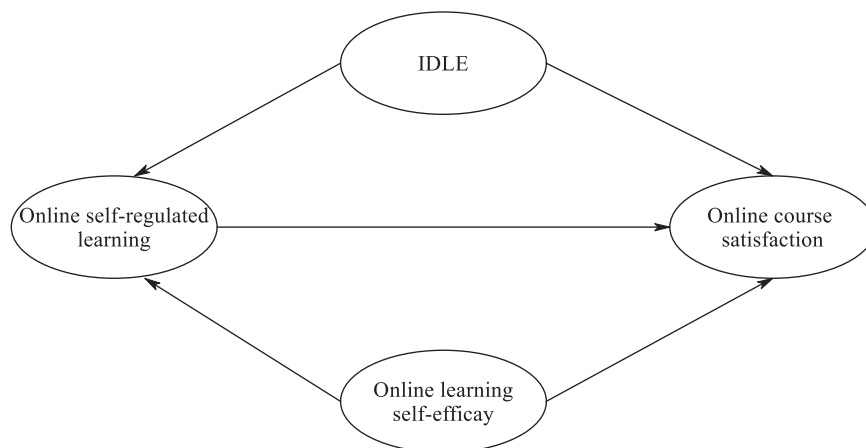
FIGURE 1
The hypothesized model.

observed in various educational contexts (Ejubovic and Puška, 2019; Wu et al., 2023).

2.6.4 H4: Online self-regulated learning acts as a mediating factor in the relationship between OLSE and IDLE, and course satisfaction

Online self-regulation serves as a mediating variable between OLSE, IDLE, and course satisfaction. When learners have higher OLSE and engage in IDLE activities, they are more likely to take a proactive approach to their learning, set challenging goals, and persist in the face of challenges. These behaviors are characteristic of effective self-regulated learners (Zimmerman, 1989). Moreover, previous research has shown that learners with higher OLSE and engagement in IDLE activities are more likely to exhibit stronger self-regulatory skills (Lai, 2019; Liu et al., 2023). Thus, it can be argued that online SRL acts as the mechanism through which OLSE and IDLE influence learners' satisfaction with online courses.

3 Materials and methods

3.1 Participants

The study included a diverse sample of 563 intermediate Chinese EFL college students, carefully chosen from multiple national universities across China. Among the participants, there were 247 males ($N = 247$) and 316 females ($N = 316$). The age of the participants ranged from 19 to 26 years, with a mean age of 20.88 years (SD = 2.06). All participants were undergraduate students pursuing various disciplines. Regarding their English proficiency level, the participants were categorized as intermediate. This ensured a homogenous language proficiency level for the study, which allowed for more precise investigations into the relationships between variables. The selection of participants was carried out using a rigorous stratified random sampling technique. Employing this approach, the researchers aimed to achieve comprehensive representation from multiple national universities across diverse geographic regions in China. Stratification based

on the geographical distribution of universities was crucial to account for potential regional variations in learning approaches, preferences, and experiences. Consequently, this sampling method minimized biases and enhanced the generalizability of the study's findings to the broader population of Chinese EFL learners.

The stratified random sampling procedure involved dividing the population of Chinese EFL college students from various national universities into distinct strata based on geographic regions. From each stratum, participants were randomly selected to participate in the research. This approach ensured that every participant had an equal chance of being included in the study, thereby ensuring the validity and representativeness of the findings.

3.2 Instruments

Four assessment instruments were employed in this study to measure various aspects of participants' experiences with online learning (see **Supplementary Appendix**). The first instrument, the Online Course Satisfaction Scale (OCSS), developed by Wei and Chou (2020), consists of 7 items assessing learners' satisfaction with an online course. These items gauge perceptions regarding course design and teachers' performance. Participants indicated their responses on a 5-point Likert scale, ranging from strongly disagree (1) to strongly agree (5).

For assessing informal digital English learning, the IDLE Scale was adapted from Lee and Drajati's (2019) original version. The scale includes four subscales: form-focused activities (FF), game-based activities (GB), receptive IDLE activities (RE), and productive IDLE activities (PR). Participants reported their engagement frequency in these activities on a 5-point Likert-type scale, varying from never (1) to very often—many times per day (5). A sample item included, "I use Google to check grammar and vocabulary."

To measure online learner self-regulation, the Online Self-Regulated Learning Questionnaire (OSLQ) was employed, based on Barnard et al.'s (2010) work. The OSLQ consists of 24 items and assesses constructs such as goal setting, time management, help seeking, task strategies, and self-evaluation. Participants responded to statements using a five-point Likert-type scale,

TABLE 1   The results of measurement model.

Constructs	Indicators	Standardized loading	t-value
OCSS	OCSS1	0.75	6.43
	OCSS2	0.81	7.29
	OCSS4	0.72	6.12
	OCSS5	0.76	6.57
	OCSS6	0.70	5.95
	OCSS7	0.79	7.03
IDLE	IDLE1	0.64	4.89
	IDLE2	0.72	5.95
	IDLE3	0.68	5.22
	IDLE4	0.71	5.75
	IDLE5	0.69	5.45
	IDLE6	0.75	6.32
	IDLE7	0.70	5.62
	IDLE8	0.67	5.13
	IDLE9	0.71	5.79
	IDLE11	0.66	4.98
	IDLE12	0.70	5.60
	IDLE13	0.73	6.04
OSLQ	OSLQ1	0.80	8.12
	OSLQ2	0.79	7.98
	OSLQ4	0.76	7.32
	OSLQ5	0.81	8.03
	OSLQ6	0.78	7.69
	OSLQ7	0.84	8.95
	OSLQ8	0.79	7.92
	OSLQ9	0.82	8.54
	OSLQ10	0.77	7.41
	OSLQ11	0.80	8.21
	OSLQ12	0.75	7.18
	OSLQ13	0.83	8.71
	OSLQ14	0.78	7.63
	OSLQ15	0.81	8.07
	OSLQ17	0.80	8.28
	OSLQ18	0.82	8.63
	OSLQ19	0.75	7.15
	OSLQ20	0.81	8.01
	OSLQ21	0.76	7.28
	OSLQ22	0.83	8.69
	OSLQ23	0.79	7.85
	OSLQ24	0.82	8.58
OLSS	OLSS1	0.70	6.02
	OLSS2	0.68	5.48
	OLSS3	0.71	6.16

(Continued)

TABLE 1   (Continued)

Constructs	Indicators	Standardized loading	t-value
	OLSS4	0.75	6.71
	OLSS5	0.73	6.32
	OLSS6	0.69	5.76
	OLSS8	0.72	6.19
	OLSS9	0.70	5.93
	OLSS10	0.68	5.43
	OLSS11	0.75	6.63
	OLSS12	0.73	6.28
	OLSS13	0.71	6.07
	OLSS14	0.74	6.38
	OLSS15	0.70	5.85
	OLSS16	0.68	5.39
	OLSS17	0.75	6.58
	OLSS18	0.73	6.25
	OLSS20	0.74	6.40
	OLSS21	0.70	5.90
	OLSS22	0.68	5.54
	OLSS23	0.75	6.65
	OLSS24	0.73	6.27
	OLSS25	0.71	6.04

OCSS, online course satisfaction; OSLQ, online self-regulated learning; OLSS, online learning self-efficacy.

ranging from strongly disagree to strongly agree. The OSLQ comprises six subscales, each reflecting different aspects of self-regulation. Reliability coefficients for the subscales were reported to be 0.80 and above (Barnard et al., 2009).

Lastly, the online learning self-efficacy scale (OLSS) by Tsai et al. (2020) was utilized, featuring 25 items assessing five distinct aspects of self-efficacy related to online learning. These aspects include participants' confidence in successfully completing an online course, engaging socially with peers, navigating Course Management System (CMS) tools, interacting with online instructors, and collaborating with classmates for academic purposes. Participants rated their agreement with each statement using a 5-point Likert scale, ranging from strongly disagree (1) to strongly agree (5).

3.3 Procedure

Participants were recruited through official communication channels of their respective national universities. Before data collection, the researchers sought ethical approval from the institutional review board at their affiliated university. Eligible participants received a formal invitation, which detailed the purpose, significance, and confidentiality of the study. Emphasis was placed on voluntary participation and informed consent, with the assurance that their involvement would not impact their academic standing or any other aspect of their university life. Data

collection occurred through an online survey questionnaire hosted on a secure and confidential platform. The survey period extended over 5 weeks, allowing participants sufficient time to complete the questionnaire at their convenience. Regular reminders were sent to encourage participation and maximize response rates. Informed consent was obtained from all participants, ensuring they were fully aware of the confidentiality and anonymity of their responses. To protect participants' privacy, personal identifiers were removed from the dataset. The research strictly adhered to the guidelines and principles outlined in the Declaration of Helsinki, ensuring the rights and welfare of the participants were safeguarded throughout the study. For data collection, a secure and confidential online survey platform was selected. This platform was chosen for its user-friendly interface, accessibility, and robust data management capabilities. Adhering to industry standards, the survey platform implemented data encryption and protection measures to ensure the confidentiality and integrity of participants' responses.

3.4 Data analysis

To analyze the data, descriptive statistics and correlations were computed using SPSS 28.0. Subsequently, confirmatory factor analysis (CFA) was performed through AMOS 26.0, a software application commonly utilized for Covariance-Based Structural Equation Modeling (CB-SEM) (Sarstedt et al., 2022).

We employed AMOS for effective assessment of construct validity and model fit, followed by SEM to explore hypothesized relationships between latent constructs. The choice of AMOS in our analysis stems from its advantages in the context of CB-SEM, providing a robust platform for evaluating complex relationships in our research framework. Notably, CB-SEM, distinct from partial least squares (PLS)-SEM, focuses on explaining covariances between observed variables and latent constructs, making it suitable for theory-driven models like ours (Hair et al., 2006; Dash and Paul, 2021). This aligns with our objective of elucidating interrelationships between self-efficacy, informal digital learning, self-regulated learning, and course satisfaction constructs.

To evaluate the adequacy of the model, widely recognized fit measures were employed, in accordance with the guidelines suggested by Hu and Bentler (1999). These fit measures encompassed the ratio of χ2 goodness-of-fit to degrees of freedom (df), the goodness of fit index (GFI), the comparative fit index (CFI), the Root-Mean-Square Error of Approximation (RMSEA), and the Standardized Root-Mean-Square Residual (SRMR). A χ2/df ratio below 3, accompanied by a *p*-value exceeding 0.05, indicated a good fit. Additionally, GFI and CFI values equal to or greater than 0.90 were considered indicative of a good fit, while RMSEA values below 0.08 and SRMR values below 0.10 were considered acceptable fit. To assess the significance of the indirect effects, bootstrapping analyses with 5,000 resamples were conducted, following Hayes' (2013) approach.

4 Results

Prior to conducting substantive analyses, we conducted thorough data quality checks to ensure the integrity and reliability

of our dataset. These checks encompassed a series of procedures aimed at identifying and addressing potential issues related to data completeness, accuracy, and consistency.

First, we examined missing data patterns across all variables included in the study. The analysis revealed that less than 5% of the data were missing, which were predominantly due to occasional non-responses by participants. We employed a principled approach to handle missing data, utilizing listwise deletion for cases with missing values, given the relatively low rate of non-response (Kline, 2023). Sensitivity analyses were conducted to evaluate the impact of missing data on our findings, confirming that the final sample remained representative of the original cohort.

Next, we assessed the distributional properties of the variables, examining for outliers, skewness, and kurtosis. Outliers were defined as values falling more than 1.5 times the interquartile range above the third quartile or below the first quartile (Kline, 2023). In our dataset, we observed a minimal number of outliers (less than 2% of cases), which were retained in the analysis as they did not exert undue influence on the results.

Additionally, we conducted a normality test on each variable item, following the approach outlined (Hair et al., 2006). The results, displayed in **Table 4**, reveal that the skewness values fall within the range of −0.19 to −0.08, indicating a relative proximity to 0 and suggesting a symmetric distribution of the data. Moreover, the kurtosis values range from 0.14 to 0.21, which are in close approximation to the expected value of 3, signifying that the data exhibits relatively normal tail behavior. Based on these observations, it is reasonable to conclude that our dataset demonstrates an approximate normal distribution, given the proximity of skewness and kurtosis values to 0 and 3, respectively.

Furthermore, we examined the consistency and reliability of the measures through Cronbach's alpha coefficients. All constructs demonstrated high internal consistency, with values exceeding the recommended threshold of 0.7 (Hair et al., 2006), affirming the reliability of our measurement model. Lastly, we conducted preliminary assessments to identify potential sources of non-response bias. We compared demographic characteristics, such as age, gender, educational background, and prior online learning experience, and key study variables between respondents and non-respondents, finding no substantial differences that would impede the generalizability of our findings. Additionally, we examined the response rates and patterns across different survey items and found no substantial variation that could indicate significant non-response bias.

Then, CFA was conducted using AMOS 26.0 to establish the validity and robustness of the measurement model. The analysis aimed to evaluate the proposed relationships between the latent constructs and their observed indicators and to assess the overall fit of the model to the data. The initial measurement model included indicators for each of the four latent constructs, with multiple observed indicators derived from their respective scales. However, the initial model did not meet the criteria for a satisfactory fit, as indicated by the following fit indices: $\chi^2(259) = 573.621$, CFI = 0.901, Tucker-Lewis Index (TLI) = 0.898, RMSEA = 0.086, SRMR = 0.074.

To enhance the construct validity and improve model fit, an iterative refinement process was undertaken. This process involved meticulous examination of modification indices and adherence to

TABLE 2 Convergent and divergent validity.

Variables	AVE	CR	1	2	3	4
1. Online self-efficacy	0.56	0.91	*0.74*			
2. IDLE	0.62	0.85	0.33**	*0.78*		
3. Online self-regulation	0.73	0.87	0.40***	0.41***	*0.85*	
4. Online satisfaction	0.67	0.81	0.44***	0.27*	0.51***	*0.81*

Online satisfaction, online course satisfaction; AVE, average variance extracted; CR, composite reliability. Italic font numbers are square roots of the AVE; off diagonals are correlation confidents; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

TABLE 3 HTMT values.

	1	2	3	4
1. Online self-efficacy	–			
2. IDLE	0.33	–		
3. Online self-regulation	0.40	0.41	–	
4. Online satisfaction	0.44	0.27	0.51	–

established guidelines to identify potential sources of model misfit. Consequently, two items from the online learning self-efficacy scale (OLSS7 and OLSS19) were identified as having low factor loadings, indicating inadequate representation of the underlying construct. As a result, these items were removed from the model to strengthen the construct's validity and conceptual coherence.

Similarly, during the refinement process, two items from the online self-regulated learning scale (OSLQ3 and OSLQ16) and one item each from the IDLE (IDLE10) and online course satisfaction (OCSS3) scales exhibited weak factor loadings. In response, these items were excluded from the model to enhance the clarity and unidimensionality of their respective constructs. Following these modifications, the revised measurement model underwent another round of CFA analysis, demonstrating a substantial improvement in fit indices. The fit indices for the revised model were as follows: $\chi^2(281) = 524.389$, CFI = 0.953, TLI = 0.958, RMSEA = 0.039, and SRMR = 0.046.

Comparing the fit indices between the initial and revised models unequivocally reveals that the revised model exhibited significantly better goodness-of-fit statistics. Based on the considerable improvement in fit indices and the enhanced construct validity achieved through the iterative refinement process, the revised model was deemed suitable for subsequent data analyses. Table 1 indicates the standardized loadings and *t*-values of the observed indicators on their respective latent constructs. The factor loadings of all indicators were significant ($p < 0.001$).

Then we examined convergent and divergent validity of the latent variables in the measurement model (see Table 2). Convergent validity indicates the extent to which indicators of a latent construct are related, while divergent validity assesses the distinctiveness of different constructs (Fornell and Larcker, 1981).

The average variance extracted (AVE) and composite reliability (CR) were calculated to assess convergent validity. The square roots of the AVE are presented in italic font in Table 2, and the correlation coefficients are displayed off the diagonals. The

correlations between different constructs provide evidence of divergent validity. The correlation coefficients are within an acceptable range, suggesting that the latent constructs are distinct from each other.

In addition to the Fornell–Larcker test, it is advisable to incorporate the evaluation of HTMT (heterotrait–monotrait ratio of correlations) values to conduct a thorough assessment of discriminant validity (Hair et al., 2006; Wijaya et al., 2022). When employing the HTMT criterion, a construct is considered to have strong discriminant validity if the HTMT value remains below the 0.9 threshold. This approach serves to confirm the discriminant validity in our study, as evident from the results presented in Table 3.

Overall, these results affirm the measurement model's convergent and divergent validity, indicating that the indicators of each latent construct are related, while the constructs themselves are distinguishable. The analysis supports the suitability of the measurement model for subsequent analyses and confirms the validity of the conceptual framework used to explore the relationships between the constructs in EFL learners.

Table 4 presents the descriptive statistics and reliability indices for the measured variables in the study. Participants reported a mean score of 3.81 (SD = 0.73) on the online learning self-efficacy scale, indicating a moderate level of self-efficacy in online learning. The internal consistency of the scale was found to be satisfactory (Cronbach's α = 0.81), indicating good reliability. For IDLE, participants had a mean score of 3.09 (SD = 0.95), suggesting a moderate level of engagement in informal digital learning activities for English. The internal consistency of the IDLE scale was high (Cronbach's α = 0.87), indicating strong reliability.

Regarding online self-regulation, the students reported a mean score of 3.66 (SD = 0.77), indicating a moderate level of self-regulated learning in the online context. The internal consistency of the scale was excellent (Cronbach's α = 0.92), indicating high reliability. In terms of online course satisfaction, participants had a mean score of 3.88 (SD = 0.69), suggesting a relatively high level of satisfaction with online learning activities. The internal consistency of the scale was found to be satisfactory (Cronbach's α = 0.84), indicating good reliability.

Then, the correlations among the constructs were calculated (see Table 5). As seen in Table 5, online self-efficacy is positively and significantly correlated with online self-regulation ($r = 0.40$, $p < 0.001$) and online course satisfaction ($r = 0.44$, $p < 0.001$). This indicates that higher levels of self-efficacy in online learning are associated with higher levels of online self-regulated learning and greater satisfaction with the online course. IDLE exhibits a positive and significant correlation with online self-regulation ($r = 0.41$, $p < 0.001$) and online course satisfaction ($r = 0.27$, $p < 0.05$).

TABLE 4 Descriptive statistics and reliability.

Variables	Mean	SD	Skewness	Kurtosis	Cronbach's alpha
1. Online self-efficacy	3.81	0.73	−0.19	0.21	0.81
2. IDLE	3.09	0.95	−0.11	0.20	0.87
3. Online self-regulation	3.66	0.77	−0.08	0.19	0.92
4. Online satisfaction	3.88	0.69	−0.06	0.14	0.84

TABLE 5 Correlations among the constructs.

Variables	1	2	3	4
1. Online self-efficacy	1.00			
2. IDLE	0.33**	1.00		
3. Online self-regulation	0.40***	0.41***	1.00	
4. Online satisfaction	0.44***	0.27*	0.51***	1.00

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

This suggests that a higher level of informal digital learning engagement in English is related to increased levels of online self-regulated learning and a higher degree of satisfaction with the online course. Online self-regulation is significantly and positively correlated with online course satisfaction ($r = 0.51$, $p < 0.001$). This indicates that a greater degree of self-regulated learning in the online context is associated with increased satisfaction with the online course. These findings highlight the interrelated nature of the constructs under investigation. Specifically, the positive correlations between online self-efficacy, IDLE, online self-regulation, and online course satisfaction support the hypothesized relationships posited in the conceptual framework. This suggests that learners who perceive themselves as efficacious in online learning and engage in informal digital learning activities in English are more likely to demonstrate self-regulated learning behaviors in the online course and experience higher levels of satisfaction with the learning experience.

The significant correlations observed between the latent variables lend support to the validity of the measurement model and provide evidence for the relationships postulated in the research hypotheses. The findings contribute to a better understanding of the complex interplay between online learning self-efficacy, IDLE, online self-regulated learning, and online course satisfaction in the context of EFL learners' online courses.

Structural equation modeling (SEM) was utilized to investigate the hypothesized relationships between the latent constructs. The model fit indices indicated a satisfactory fit to the data: $\chi^2(362) = 672.452$, $p = 0.000$, CFI = 0.952, TLI = 0.948, and RMSEA = 0.041, 95% CI = 0.036–0.047. These indices collectively suggest that the proposed structural model adequately represents the relationships between the latent constructs. **Figure 2** depicts the path diagram illustrating the hypothesized relationships between the constructs. All path coefficients were found to be statistically significant. The specific path coefficients are as follows: IDLE positively and significantly predicted online course satisfaction ($β = 0.34$). Learners who engage more in informal digital learning

activities for English tend to exhibit higher levels of satisfaction in online learning.

Online self-efficacy positively and significantly influenced online course satisfaction ($β = 0.41$). Learners' beliefs in their ability to perform well in online learning are associated with higher levels of satisfaction with the online course. Online self-regulated learning had a positive and significant effect on online course satisfaction ($β = 0.49$). Learners who demonstrate better self-regulated learning behaviors in the online context are more satisfied with online learning activities in the course. IDLE positively and significantly predicted online self-regulated learning ($β = 0.38$). Learners who engage more in informal digital learning activities for English tend to demonstrate higher levels of self-regulated learning in the online environment. Online learning self-efficacy had a positive and significant influence on online self-regulated learning ($β = 0.42$). Learners with higher levels of self-efficacy in online learning are more likely to exhibit better self-regulated learning behaviors in the online context.

These results provide empirical support for the hypothesized relationships between the constructs and highlight the significant role of IDLE, online self-efficacy, and online self-regulated learning in predicting learners' satisfaction with the online course. The findings contribute to a deeper understanding of the factors influencing learners' satisfaction in online learning environments and hold practical implications for EFL educators and instructional designers seeking to enhance the quality of online courses.

To evaluate the significance of the indirect effects, bootstrapping analyses were conducted using 5,000 resamples (Hayes, 2013). **Table 6** summarizes the results of the bootstrapping analyses, presenting the direct, indirect, and total effects in the mediation analysis. The direct effects between the predictor variables and the outcome variable were statistically significant. Specifically, online self-efficacy ($β = 0.41$, $p < 0.001$), IDLE ($β = 0.34$, $p < 0.001$), and online self-regulation ($β = 0.49$, $p < 0.001$) had positive and significant relationships with online course satisfaction. The mediation analysis further revealed significant indirect effects. Online self-efficacy showed a significant indirect effect on online course satisfaction through online self-regulated learning ($β = 0.20$, 95% CI [0.14, 0.26], $p < 0.001$). Similarly, IDLE had a significant indirect effect on online course satisfaction through online self-regulated learning ($β = 0.18$, 95% CI [0.12, 0.24], $p < 0.001$).

Furthermore, the total effects, which combine the direct and indirect effects, were also statistically significant. Online self-efficacy had a total effect of 0.61 (95% CI [0.55, 0.67], $p < 0.001$) on online course satisfaction, while IDLE had a total effect of 0.52 (95% CI [0.46, 0.58], $p < 0.001$).

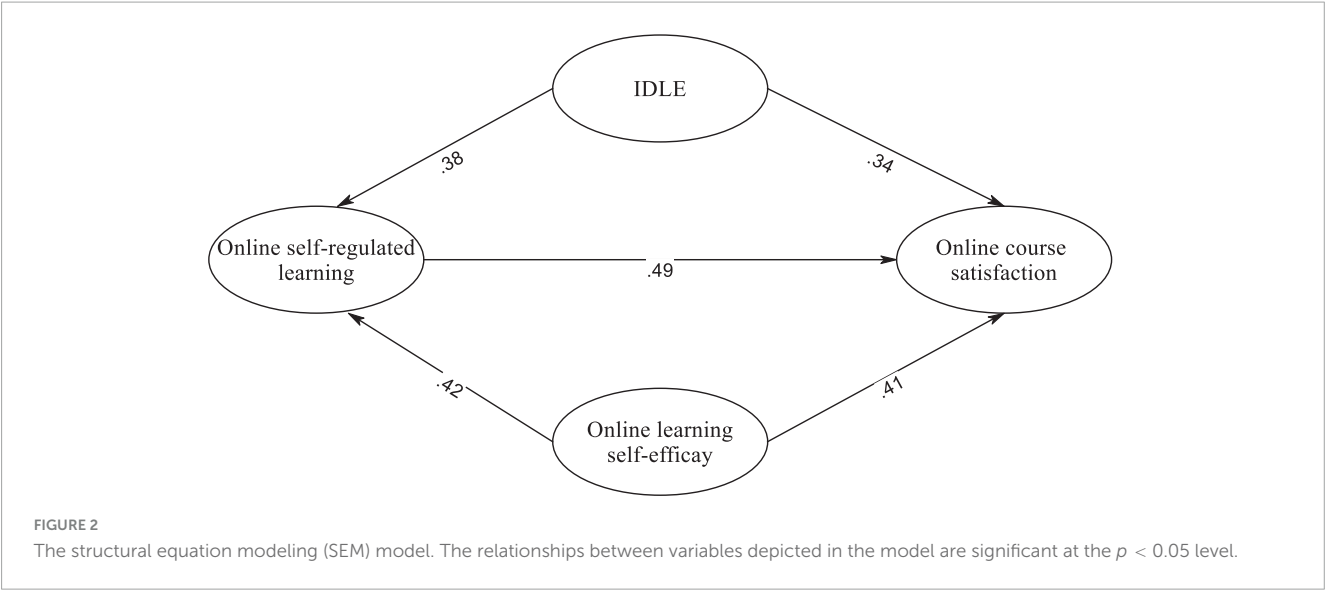Furthermore, we employed Harman's single-factor test, a well-recognized technique for assessing common method bias

FIGURE 2
The structural equation modeling (SEM) model. The relationships between variables depicted in the model are significant at the $p < 0.05$ level.

TABLE 6 The results of SEM.

Path	Coefficient (β)	Bootstrapped 95% CI	p-value	Effect
Online self-efficacy → online satisfaction	0.41	[0.35, 0.48]	< 0.001	Direct
IDLE → online satisfaction	0.34	[0.27, 0.41]	< 0.001	Direct
Online self-regulation → online satisfaction	0.49	[0.42, 0.56]	< 0.001	Direct
Online self-efficacy → OSRL → online satisfaction	0.20	[0.14, 0.26]	< 0.001	Indirect
IDLE → OSRL → online satisfaction	0.18	[0.12, 0.24]	< 0.001	Indirect
Online self-efficacy → online satisfaction	0.61	[0.55, 0.67]	< 0.001	Total
IDLE → online satisfaction	0.52	[0.46, 0.58]	< 0.001	Total

OSRL, online self-regulated learning; online satisfaction, online course satisfaction.

(Podsakoff et al., 2003). In this test, we conducted an exploratory factor analysis (EFA) on all items included in the study, with all items loaded onto a single factor. If common method bias was a significant concern, a single factor would account for a substantial proportion of the variance. However, the results of the EFA showed that the single factor accounted for 27.36% of the variance, suggesting that common method bias is unlikely to be a major issue in our dataset.

In addition, we performed a common latent factor analysis, following the procedure suggested by Kock and Lynn (2012). This analysis examines whether a common latent factor accounts for the covariance among the variables. The results indicated that a common latent factor did not significantly account for the variance in the observed variables, further supporting the absence of substantial common method bias.

Overall, the mediation analysis provided valuable insights into the underlying mechanisms through which online learning self-efficacy and IDLE influence online course satisfaction. The results indicate that both online self-efficacy and IDLE have direct effects on online course satisfaction, and these effects are partially mediated by online self-regulated learning. These findings contribute to a better understanding of the factors influencing online learning satisfaction in EFL learners and highlight the significance of self-efficacy and informal digital

learning experiences in fostering online course satisfaction and self-regulated learning behaviors in the online learning context. The findings have practical implications for educators and course designers seeking to optimize the online learning experience and enhance learner engagement and satisfaction in EFL contexts.

5 Discussion

In this research we sought to investigate a predictive role of online learning self-efficacy, self-regulated learning, and IDLE on online course satisfaction among EFL learner in the context of China. By doing this, the current study can offer a diverse range of both theoretical and pedagogical implications for EFL learning and teaching.

First, the findings indicated that online learning self-efficacy had a positive and direct impact on online course satisfaction. Our findings align with prior research that has indicated a positive relationship between online learning self-efficacy and learners' satisfaction with online courses (e.g., Shen et al., 2013; Suryandani and Santosa, 2021; Aldhahi et al., 2022; Derakhshesh et al., 2022). This finding supports Bandura's social cognitive theory, which posits that individuals' beliefs in their own ability to perform tasks and achieve desired outcomes play a crucial role in shaping

their motivation and performance (Bandura, 1982, 1997). The positive influence of online learning self-efficacy on online course satisfaction can be explained by the role of self-efficacy in reducing learners' anxiety and stress in virtual learning environments (Passiatore et al., 2019; Ferreira et al., 2020). In online settings, students may face unique challenges and uncertainties, and higher levels of self-efficacy can act as a buffer against the negative effects of anxiety and stress. As a result, learners with higher self-efficacy beliefs are more likely to feel confident and comfortable engaging in online activities, which in turn enhances their satisfaction and enjoyment in the learning process (Bolliger and Halupa, 2012; Kuo et al., 2014).

Furthermore, students with higher online learning self-efficacy are more inclined to set challenging tasks and goals, persevere in their efforts, and view difficulties as opportunities for growth (Dong et al., 2022). These self-directed and proactive behaviors are characteristic of effective self-regulated learners, which, in turn, can contribute to higher satisfaction with online courses (Zimmerman, 1989). When learners believe in their ability to succeed, they are more likely to take charge of their learning process, seek resources, and actively engage with the course materials, leading to a more fulfilling learning experience (Ji et al., 2022). As Landrum (2020) added, students' self-efficacy in virtual classes can be greatly conducive to students' course satisfaction of online learning.

Second, the findings of this study indicate that online self-regulated learning plays a significant and positive role in predicting online course satisfaction among EFL students. This finding is in line with previous research in the field of general education, which has also demonstrated the positive impact of self-regulated learning on learners' course satisfaction (Kuo et al., 2014; Ejubovic and Puška, 2019; Zalli et al., 2019; Lim et al., 2020; Wu et al., 2023). The ability of learners to regulate their learning process, plan their approach, and seek support from teachers and peers contributes to their resilience in the face of setbacks and challenges commonly encountered in online learning environments, leading to a higher level of satisfaction with the online course. Self-regulated learners demonstrate certain key characteristics that enhance their learning experience (Theobald, 2021) and, consequently, their satisfaction with online courses. Firstly, they are skilled in setting clear and achievable goals and objectives (Teng, 2021), which fosters their focus and motivation throughout the duration of the online course (Zimmerman, 1989, 1990). Having a sense of purpose and direction in their learning journey enables EFL students to approach online learning with determination and purpose, which in turn enhances their satisfaction with the progress they make and the achievements they attain (Kuo et al., 2014; Lim et al., 2020; Saint et al., 2022).

Moreover, self-regulated learners proactively employ various learning strategies to enhance their understanding and mastery of the course materials (Zimmerman, 2002). Engaging in activities such as seeking additional resources, organizing study materials, and actively participating in discussions enables them to take a more active role in their learning process. These proactive behaviors contribute to a more fulfilling and rewarding learning experience, resulting in higher levels of satisfaction with the online course. Additionally, online self-regulated learners take responsibility for managing their time and learning pace (Zimmerman and Kulikowich, 2016; Zalli et al., 2019; Xu et al., 2022). Their ability to function effectively with less dependence on external structures and guidance allows them to exercise autonomy in their learning

journey. This autonomy and flexibility in online learning empower learners as they feel more in control of their learning process, leading to a heightened sense of satisfaction.

Third, it was revealed that IDLE directly predicted online course satisfaction. This result underscores the critical role that self-initiated digital language learning activities play in shaping learners' satisfaction within online course environments. Numerous studies have recognized the efficacy of IDLE in fostering language learning outcomes, cultural understanding, and communication skills (Lee and Sylvén, 2021; Lee et al., 2022; Liu et al., 2023). These findings have consistently highlighted the positive impact of IDLE on learners' engagement and language acquisition processes.

The positive association between IDLE and online course satisfaction can be comprehensively understood through several interrelated factors supported by previous literature. Firstly, IDLE often allows learners to take charge of their language learning process in an informal context (Lai, 2019; Lee, 2021). This autonomous engagement empowers learners, providing them with a sense of control over their learning process, which positively impacts their satisfaction within online courses (Lee, 2019, 2021; Lee and Drajati, 2019). Secondly, the diverse range of resources available in IDLE, such as digital platforms, social media, blogs, and various computer-based technologies, offers learners opportunities to engage with language learning in a more naturalistic and unstructured manner (Lee et al., 2023; Soyoof et al., 2023). Such unstructured, self-directed learning experiences have been associated with increased motivation and interest among learners, which can contribute significantly to their satisfaction with the overall learning experience (Liu et al., 2023).

Furthermore, IDLE often bridges the gap between formal learning contexts and real-world applications, allowing learners to practice language skills in authentic situations (Lee and Sylvén, 2021). This practical, real-life application can enhance learners' confidence in using the language, thereby positively impacting their satisfaction with online language courses. The direct prediction of online course satisfaction by IDLE also underscores the evolving nature of language learning in digital environments (Lee and Drajati, 2019; Lee, 2021). In the contemporary educational landscape, where technology is ubiquitous, IDLE represents a flexible and accessible avenue for learners to supplement their formal learning experiences (Lee et al., 2022). This flexibility and accessibility contribute to learners' overall satisfaction, as they are able to tailor their language learning experiences to suit their preferences and needs.

Finally, the present research has shed light on the mediating role of online self-regulated learning in the relationship between IDLE, online learning self-efficacy, and online course satisfaction among EFL learners. The findings suggest that higher levels of online learning self-efficacy are associated with a greater propensity for engaging in self-regulated learning behaviors. Learners who possess strong self-efficacy beliefs in their ability to manage their learning effectively are empowered to set challenging goals, apply effort, and persist in their learning activities (Zimmerman, 2000; Jansen et al., 2019; Dinh and Nguyen, 2022). As a result, these self-efficacious learners are more likely to adopt effective learning strategies and utilize available resources optimally, leading to enhanced engagement and meaningful interactions with course materials, instructors, and peers. This heightened engagement,

in turn, contributes to their overall satisfaction with the online learning experience (Kuo et al., 2014).

Similarly, IDLE activities, being learner-centered, provide EFL learners with a sense of autonomy and responsibility over their own learning process (Lee, 2021). Consequently, it is reasonable to assume that engagement in IDLE may lead to greater online course satisfaction among EFL students. Indeed, studies have shown that autonomy can positively impact learners' sense of satisfaction in online environments (Mohammadi Zenouzagh et al., 2023). Furthermore, IDLE activities can foster the development of learners' social skills through increased interaction opportunities with others on interactive platforms, potentially boosting online course satisfaction. Moreover, learners who actively engage in IDLE experiences are likely to apply self-regulated learning strategies to effectively manage their informal learning activities. Such engagement in IDLE encourages learners to take responsibility for their learning, set goals, and monitor their progress, all of which align with the principles of self-regulated learning (Zhang and Liu, 2022; Soyoof et al., 2023). As learners apply self-regulated learning strategies in their IDLE experiences, they develop a strong foundation of self-regulatory skills that can extend to their formal online courses (Lee, 2021).

Furthermore, the role of online self-regulated learning in shaping learners' responses to challenges and setbacks during their online language learning journey cannot be overlooked. Self-regulated learners are more likely to perceive obstacles as surmountable and persist in the face of difficulties, resulting in a more positive emotional experience during online learning (Bandura, 1988). This positive emotional experience, in turn, contributes to their overall satisfaction with the course.

Additionally, the findings of this study align closely with the principles of SDT introduced by Deci and Ryan (1985). SDT posits that individuals harbor inherent psychological needs for autonomy, competence, and relatedness. When these fundamental needs find fulfillment, individuals are more inclined to willingly engage in activities and experience heightened satisfaction (Ryan and Deci, 2000). The SEM analysis conducted in this study unveiled direct and positive effects of both online learning self-efficacy and IDLE on online course satisfaction. This discovery implies that when learners possess belief in their capacity to excel in online learning and actively partake in informal digital language learning experiences that extend beyond the formal curriculum, they are more likely to find contentment with their online courses (Hsu et al., 2019).

From the vantage point of SDT, this heightened confidence among learners amplifies their sense of autonomy and competence, which, as per self-determination theory, stand as pivotal psychological needs (Deci and Ryan, 2015). This cultivated sense of autonomy and competence acts as a catalyst for elevated satisfaction with online courses. Furthermore, the involvement in informal digital language learning experiences beyond the prescribed curriculum serves as a means to gratify learners' need for autonomy and competence (Chen and Jang, 2010). Actively seeking supplementary opportunities for language learning beyond the obligatory coursework demonstrates autonomy, as learners take the reins of their own educational journey. This active engagement not only bestows them with a profound sense of autonomy but also propels the development of their language skills, ultimately heightening their perceived competence (Huang et al., 2019).

Moreover, the study divulged that online self-regulated learning operates as a partial mediator between online learning self-efficacy/IDLE and online course satisfaction. This observation mirrors the tenets of self-determination theory. Self-regulated learning showcases learners' ability to autonomously oversee and direct their learning process. Through the establishment of objectives and continuous monitoring of progress, learners wield control over their education, thereby satiating their need for autonomy (Deci and Ryan, 2000). Additionally, the deployment of effective learning strategies not only underscores learners' competence in navigating online course materials effectively (Roca and Gagné, 2008; Chiu, 2022), but also fortifies their sense of autonomy and competence, thus further heightening their satisfaction levels with online courses.

Although our study highlights the positive relationships between online learning self-efficacy, self-regulated learning, IDLE, and course satisfaction, it is crucial to consider alternative factors that might influence these results. One such factor could be the diverse nature of learners' prior experiences and skills in online learning environments. Individual differences, such as technological proficiency, prior exposure to online courses, or varying degrees of digital literacy, could significantly impact how learners engage with self-regulated strategies, IDLE activities, and ultimately perceive their efficacy in online learning contexts (Zimmerman and Kulikowich, 2016; Tang et al., 2022).

Furthermore, contextual aspects, including the design and structure of online courses, teaching methodologies, and the availability of resources and support, may also contribute to the observed relationships. For instance, certain course designs might favor or hinder the application of self-regulated learning strategies, impacting learners' satisfaction with the learning process (Landrum, 2020). Additionally, variations in the quality and type of IDLE activities or the accessibility of digital resources might influence learners' experiences and subsequently, their satisfaction levels with the online course (Lee et al., 2022; Liu et al., 2023).

Another plausible explanation could involve the dynamic nature of motivation and engagement in online learning. While our study emphasizes the positive associations between self-regulation, IDLE, and course satisfaction, other motivational factors, such as intrinsic motivation, interest in the subject matter, or social interaction within the virtual classroom, might interact with or moderate these relationships (Deci and Ryan, 1985; Lee and Sylvén, 2021; Liu et al., 2023). By acknowledging these potential alternative explanations, we recognize the complexity of the relationships investigated in our study. Future research could delve deeper into these factors to unravel their specific contributions and interactions, providing a more comprehensive understanding of the dynamics influencing online course satisfaction among EFL learners.

6 Conclusion

The present research investigated the intricate interrelationships between online learning self-efficacy, IDLE, online self-regulated learning, and online course satisfaction among EFL learners in the milieu of online language learning. The outcomes of this study reveal that online learning self-efficacy and

IDLE exert direct and positive effects on online course satisfaction among EFL learners. Learners' confidence in their online learning competencies and their engagement in informal digital language learning experiences play pivotal roles in shaping their satisfaction with online courses. Additionally, online self-regulation has been identified as a significant mediator, highlighting the importance of learners' self-regulatory behaviors in influencing the impact of self-efficacy and IDLE on course satisfaction.

Overall, these findings support the principles of self-determination theory (Deci and Ryan, 1985) by highlighting the importance of fulfilling psychological needs for autonomy and competence in promoting satisfaction with online courses. When learners believe in their abilities (online learning self-efficacy), engage in IDLE, and actively regulate their own learning process (self-regulated learning), they are more likely to experience higher levels of satisfaction with online courses.

7 Implications

The findings of this study make valuable contributions to our theoretical understanding of online learning, shedding light on the pivotal roles played by online learning self-efficacy, IDLE, and online self-regulation in shaping learners' satisfaction with online courses. This study underscores the need to consider not just the content and design of courses, but also the psychological and self-regulatory factors of learners, all of which are critical in promoting a successful online learning experience. This deeper insight enriches the evolving field of online education research and strengthens the theoretical foundations for comprehending learner satisfaction in virtual learning environments.

On a practical note, the implications of this research carry significant weight for educators, institutions, EFL stakeholders, and policymakers. The findings underscore the pivotal role of online learning self-efficacy in influencing course satisfaction. Educators and institutions should prioritize efforts to nurture learners' online learning self-efficacy. This can be achieved by providing learners with opportunities for success and by offering constructive feedback. By instilling a sense of efficacy in learners, educators can motivate them to tackle challenges, persist in their learning activities, and ultimately elevate their satisfaction with online courses.

It is important to note that increased satisfaction has been associated with more time spent online, improved grades, and overall academic success in L2 learning and teaching (Amoush and Mizher, 2023). Therefore, EFL stakeholders and policymakers should pay greater attention to this construct, striving to encourage friendly interactions among students that can enhance online learning self-efficacy and satisfaction within virtual environments. This study also underscores the significance of creating user-friendly online learning platforms that provide students with comfort, enjoyment, and reduced anxiety and stress. Such platforms are more likely to result in student satisfaction and, consequently, better L2 acquisition (Fathi and Mohammaddokht, 2021). Institutions should consider these implications when designing online language courses, as interactive and engaging learning environments foster learners' self-efficacy and encourage IDLE experiences, making a significant contribution to satisfaction and success in online courses.

Furthermore, the findings from this research underscore the significance of IDLE for EFL learners. Educators should actively encourage learners to engage in informal language learning activities through various digital platforms and resources. By promoting IDLE, educators can enhance learners' language proficiency, autonomy, and sense of belonging in the online language learning community. Given the mediating role of online self-regulated learning, educators should concentrate on teaching self-regulatory skills and strategies to learners. Providing explicit instruction and guidance on goal-setting, time management, and self-monitoring can empower learners to take control of their learning process and optimize their learning outcomes. This approach enhances their satisfaction with online courses and supports their academic success. In addition, policymakers and educational stakeholders can draw upon the insights from this research to develop evidence-based interventions and instructional strategies that support EFL learners in the online language learning context. By investing in research-backed initiatives, institutions can enhance the quality and effectiveness of online language instruction, ultimately leading to higher learner satisfaction and achievement.

8 Limitations

However, it is important to acknowledge several limitations in this study. Firstly, the findings presented here are derived from a specific sample of intermediate-level Chinese EFL college students attending various national universities. Although this specificity offers a focused perspective, it raises concerns about the extent to which the results can be generalized to other populations and language learning contexts. Given potential variations in learner profiles, needs, and experiences, it is advisable to exercise caution when extending these findings to diverse learner groups.

Secondly, the cross-sectional nature of this study, while providing a snapshot of relationships between the examined constructs, inherently constrains the establishment of causal links between variables. To gain a more profound understanding of temporal dynamics and causal pathways, future research could benefit from adopting longitudinal research designs. Such an approach could unveil the nuanced cause-and-effect relationships among online learning self-efficacy, IDLE, online self-regulated learning, and course satisfaction. Additionally, another notable limitation lies in the study's reliance on self-reported data, which introduces the potential for response bias. Although self-reports provide valuable insights into learners' perceptions, attitudes, and behaviors, they may not always perfectly align with objective measures. To enhance the validity and reliability of findings, future research might consider complementing self-report data with objective assessments and observations.

Building on the insights gained from this study, promising future research directions emerge. Researchers can contemplate intervention studies designed to enhance learners' self-efficacy, foster more effective IDLE experiences, and promote online self-regulated learning skills. Longitudinal investigations could monitor the long-term impact of such interventions on learners' satisfaction, academic achievement, and persistence in the online language learning environment. Furthermore, conducting comparative

studies across various cultural and language contexts offers valuable insights into the generalizability and applicability of the theoretical relationships explored in this research. These future directions contribute to the expansion of our knowledge and the ongoing improvement of online language learning experiences for diverse learner populations.

Data availability statement

The original contributions presented in this study are included in this article/Supplementary materials, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the School of Foreign Languages and Literatures, Chongqing Normal University, Daxuecheng, Chongqing 401331, China. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing–original draft. AX: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing–review & editing.

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.2023.1276266/full#supplementary-material

References

Abdelrady, A. H., and Akram, H. (2022). An empirical study of ClassPoint tool application in enhancing EFL students' online learning satisfaction. *Systems* 10:15. doi: 10.3390/systems10050154

Abdous, M. H. (2019). Influence of satisfaction and preparedness on online students' feelings of anxiety. *Internet High. Educ.* 41, 34–44. doi: 10.1016/j.iheduc.2019.01.001

Abdullah, A. (2022). Basic psychological need satisfaction and continued language learning during a pandemic: A structural equation modelling approach. *J. Psychol. Lang. Learn.* 4:e414322. doi: 10.52598/jpll/4/1/1

Abraham, M., Arficho, Z., Habtemariam, T., and Demissie, A. (2022). Effects of information communication technology-assisted teaching training on English language teachers' pedagogical knowledge and English language proficiency. *Cogent Educ.* 9:2028336. doi: 10.1080/2331186X.2022.2028336

Adilbayeva, U., Mussanova, G. A., Mombekova, N. B., and Suttibayev, N. A. (2022). Digital communication technology for teaching a foreign*ı* language and culture through reading. *Int. J. Soc. Cult. Lang.* 10, 21–30.

Al-Abyadh, M. H. A., and Abdel Azeem, H. A. H. (2022). Academic achievement: Influences of university students' self-management and perceived self-efficacy. *J. Intell.* 10:55. doi: 10.3390/jintelligence10030055

Aldhahi, M. I., Alqahtani, A. S., Baattaiah, B. A., and Al-Mohammed, H. I. (2022). Exploring the relationship between students' learning satisfaction and self-efficacy during the emergency transition to remote learning amid the coronavirus pandemic: A cross-sectional study. *Educ. Inform. Technol.* 27, 1323–1340. doi: 10.1007/s10639-021-10644-7

Al-Rahmi, W. M., Al-Adwan, A. S., Al-Maatouk, Q., Othman, M. S., Alsaud, A. R., Almogren, A. S., et al. (2023). Integrating communication and task–technology fit theories: The adoption of digital media in learning. *Sustainability* 15:8144. doi: 10.3390/su15108144

Amoush, K. H., and Mizher, R. A. (2023). Interaction as a predicator for EFL undergraduate university students' satisfaction with online English language courses. *Theory Pract. Lang. Stud.* 13, 927–937. doi: 10.17507/tpls.1304.14

Arrosagaray, M., González-Peiteado, M., Pino-Juste, M., and Rodríguez-López, B. (2019). A comparative study of Spanish adult students' attitudes to ICT in classroom, blended and distance language learning modes. *Comput. Educ.* 134, 31–40. doi: 10.1016/j.compedu.2019.01.016

Asratie, M. G., Wale, B. D., and Aylet, Y. T. (2023). Effects of using educational technology tools to enhance EFL students' speaking performance. *Educ. Inform. Technol.* [Online ahead of print]. doi: 10.1007/s10639-022-11562-y

Bahrani, T., and Sim, T. S. (2012). Informal language learning setting: Technology or social interaction? *Turk. Online J. Educ. Technol.* 11, 142–149. doi: 10.7456/10203100/002

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

Bandura, A. (1988). "Self-regulation of motivation and action through goal systems," in *Cognitive Perspectives on Emotion and Motivation*, eds V. Hamilton, G. H. Bower, and N. H. Frijda (Dordrecht: Kluwer Academic), 37–61. doi: 10.1007/978-94-009-2792-6_2

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: W.H. Freeman.

Barnard, L., Lan, W., To, Y., Paton, V., and Lai, S. (2009). Measuring self-regulation in online and blended learning environments. *Internet High. Educ.* 12, 1–16. doi: 10.1016/j.iheduc.2008.10.005

Barnard, L., Paton, V. O., and Lan, W. Y. (2010). Profiles in self-regulated learning in the online learning environment. *Int. Rev. Res. Open Distrib. Learn.* 11, 61–80. doi: 10.19173/irrodl.v11i1.769

Bervell, B., Umar, I. N., and Kamilin, M. H. (2020). Towards a model for online learning satisfaction (MOLS): Re-considering non-linear relationships among personal innovativeness and modes of online interaction. *Open Learn.* 35, 236–259. doi: 10.1080/02680513.2019.1662776

Bolaji, H. O., and Jimoh, H. A. (2023). Usability and utilization of ICT among educational administrators in secondary students in public school. *Ind. J. Educ. Res. Technol.* 3, 97–104.

Bolliger, D. U., and Halupa, C. (2012). Student perceptions of satisfaction and anxiety in an online doctoral program. *Dist. Educ.* 33, 81–98. doi: 10.1080/01587919.2012.667961

Cassidy, S. (2015). Resilience building in students: The role of academic self-efficacy. *Front. Psychol.* 6:1781. doi: 10.3389/fpsyg.2015.01781

Chen, J., Lin, C. H., Chen, G., and Fu, H. (2023). Individual differences in self-regulated learning profiles of Chinese EFL readers: A sequential explanatory mixed-methods study. *Stud. Second Lang. Acquisit.* 45, 955–978. doi: 10.1017/S0272263122000584

Chen, K. C., and Jang, S. J. (2010). Motivation in online learning: Testing a model of self-determination theory. *Comput. Hum. Behav.* 26, 741–752. doi: 10.1016/j.chb.2010.01.011

Chen, P. H., and Adesope, O. (2016). The effects of need satisfaction on EFL online learner satisfaction. *Distance Educ.* 37, 89–106. doi: 10.1080/01587919.2016.1155962

Chiu, T. K. (2022). Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic. *J. Res. Technol. Educ.* 54, S14–S30. doi: 10.1080/15391523.2021.1891998

Daneji, A. A., Ayub, A. F. M., and Khambari, M. N. M. (2019). The effects of perceived usefulness, confirmation and satisfaction on continuance intention in using massive open online course (MOOC). *Knowl. Manage. E-Learn.* 11, 201–214. doi: 10.34105/j.kmel.2019.11.010

Dash, G., and Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technol. Forecast. Soc. Change* 173:121092. doi: 10.1016/j.techfore.2021.121092

de Jong, P. G., Pickering, J. D., Hendriks, R. A., Swinnerton, B. J., Goshtasbpour, F., and Reinders, M. E. (2020). Twelve tips for integrating massive open online course content into classroom teaching. *Med. Teacher* 42, 393–397. doi: 10.1080/0142159X.2019.1571569

Deci, E. L., and Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *J. Res. Pers.* 19, 109–134. doi: 10.1016/0092-6566(85)90023-6

Deci, E. L., and Ryan, R. M. (2000). Self-determination theory and the facilitation of intrinsic motivation. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066X.55.1.68

Deci, E. L., and Ryan, R. M. (2015). "Self-determination theory," in *International Encyclopedia of the Social & Behavioral Sciences*, 2nd Edn, ed. J. D. Wright (Amsterdam: Elsevier), 486–491. doi: 10.1016/B978-0-08-097086-8.26036-4

Derakhshan, A., and Fathi, J. (2023). Grit and foreign language enjoyment as predictors of EFL learners' online engagement: The mediating role of online learning self-efficacy. *Asia Pac. Educ. Res.* 1–11. doi: 10.1007/s40299-023-00745-x

Derakhshesh, A., Fathi, J., Hosseini, H. M., and Mehraein, S. (2022). An investigation of the structural model of online course satisfaction, online learning self-efficacy, and online learning climate in the EFL context. *Comput. Assist. Lang. Learn. Electron. J.* 23, 261–281.

Dinh, T. C., and Nguyen, P. B. N. (2022). Impact of internet self-efficacy and self-regulated learning on satisfaction and academic achievement in online learning: A case study in Vietnam. *Int. J. Emerg. Technol. Learn.* 17:269. doi: 10.3991/ijet.v17i16.33819

Dizon, G., and Thanyawatpokin, B. (2021). Emergency remote language learning: Student perspectives of L2 learning during the COVID-19 pandemic. *JALT Call J.* 17, 349–370. doi: 10.29140/jaltcall.v17n3.431

Dong, L., Jamal Mohammed, S., Ahmed Abdel-Al Ibrahim, K., and Rezai, A. (2022). Fostering EFL learners' motivation, anxiety, and self-efficacy through computer-assisted language learning-and mobile-assisted language learning-based instructions. *Front. Psychol.* 13:899557. doi: 10.3389/fpsyg.2022.899557

Ejubovic, A., and Puška, A. (2019). Impact of self-regulated learning on academic performance and satisfaction of students in the online environment. *Knowl. Manage. E-Learn.* 11, 345–363. doi: 10.34105/j.kmel.2019.11.018

Elia, G., Solazzo, G., Lorenzo, G., and Passiante, G. (2019). Assessing learners' satisfaction in collaborative online courses through a big data approach. *Comput. Hum. Behav.* 92, 589–599. doi: 10.1016/j.chb.2018.04.033

Elliott, K. M., and Shin, D. (2002). Student satisfaction: An alternative approach to assessing this important concept. *J. High. Educ. Policy Manage.* 24, 197–209. doi: 10.1080/1360080022000013518

Elshami, W., Taha, M. H., Abuzaid, M., Saravanan, C., Al Kawas, S., and Abdalla, M. E. (2021). Satisfaction with online learning in the new normal: Perspective of students and faculty at medical and health sciences colleges. *Med. Educ. Online* 26:1920090. doi: 10.1080/10872981.2021.1920090

Fathi, J., and Ebadi, S. (2020). Exploring EFL pre-service teachers' adoption of technology in a CALL program: Obstacles, motivators, and maintenance. *Educ. Inform. Technol.* 25, 3897–3917. doi: 10.1007/s10639-020-10146-y

Fathi, J., and Mohammaddokht, F. (2021). Foreign language enjoyment and anxiety as the correlates of the ideal L2 self in the English as a foreign language context. *Front. Psychol.* 12:790648. doi: 10.3389/fpsyg.2021.790648

Fathi, J., Mohammaddokht, F., and Afzali, M. (2023). Exploring Iranian EFL teachers' attitudes toward the use of learning management systems in English classes. Íkala. *Rev. Lenguaje y Cult.* 28, 30–48. doi: 10.17533/udea.ikala.v28n1a02

Fernández-Gutiérrez, M., Gimenez, G., and Calero, J. (2020). Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish Autonomous Communities. *Comput. Educ.* 157:103969. doi: 10.1016/j.compedu.2020.103969

Ferreira, ÉD. M. R., Pinto, R. Z., Arantes, P. M. M., Vieira, ÉL. M., Teixeira, A. L., Ferreira, F. R., et al. (2020). Stress, anxiety, self-efficacy, and the meanings that physical therapy students attribute to their experience with an objective structured clinical examination. *BMC Med. Educ.* 20:296. doi: 10.1186/s12909-020-02202-5

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

Fujii, S., Mitsugi, M., Nakamura, K., Ono, Y., Yamagami, T., Takeuchi, N., et al. (2022). A Comparative study of learner satisfaction in synchronous and asynchronous online courses among Japanese EFL learners. *J. Pan Pac. Assoc. Appl. Linguist.* 26, 97–120. doi: 10.25256/PAAL.26.2.6

Guo, W., Bai, B., and Song, H. (2021). Influences of process-based instruction on students' use of self-regulated learning strategies in EFL writing. *System* 101:102578. doi: 10.1016/j.system.2021.102578

Gyamfi, G., and Sukseemuang, P. (2018). EFL learners' satisfaction with the online learning program, Tell Me More. *Turkish Online J. Distance Educ.* 19, 183–202. doi: 10.17718/tojde.382798

Hair, J., Black, B., Babin, B., Anderson, R. E., and Tatham, R. L. (2006). *Multivariate Data Analysis*, 6th Edn. New York, NY: Prentice Hall.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Hew, K. F., Hu, X., Qiao, C., and Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Comput. Educ.* 145:103724. doi: 10.1016/j.compedu.2019.103724

Hsu, H. C. K., Wang, C. V., and Levesque-Bristol, C. (2019). Reexamining the impact of self-determination theory on learning outcomes in the online learning environment. *Educ. Inform. Technol.* 24, 2159–2174. doi: 10.1007/s10639-019-09863-w

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equat. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Huang, Y. C., Backman, S. J., Backman, K. F., McGuire, F. A., and Moore, D. (2019). An investigation of motivation and experience in virtual learning environments: A self-determination theory. *Educ. Inform. Technol.* 24, 591–611. doi: 10.1007/s10639-018-9784-5

Jansen, R. S., Van Leeuwen, A., Janssen, J., Jak, S., and Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educ. Res. Rev.* 28:100292. doi: 10.1016/j.edurev.2019.100292

Ji, H., Park, S., and Shin, H. W. (2022). Investigating the link between engagement, readiness, and satisfaction in a synchronous online second language learning environment. *System* 105:102720. doi: 10.1016/j.system.2022.102720

Jiang, H., Islam, A. A., Gu, X., and Spector, J. M. (2021). Online learning satisfaction in higher education during the COVID-19 pandemic: A regional comparison between Eastern and Western Chinese universities. *Educ. Inform. Technol.* 26, 6747–6769. doi: 10.1007/s10639-021-10519-x

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., and Wu, N. (2023). Effects of automatic speech recognition technology on EFL learners' willingness to communicate and interactional features. *Educ. Technol. Soc.* 26, 37–52.

Karbakhsh, R., and Ahmadi Safa, M. (2020). Basic psychological needs satisfaction, goal orientation, willingness to communicate, self-efficacy, and learning strategy use as predictors of second language achievement: A structural equation modeling approach. *J. Psycholing. Res.* 49, 803–822. doi: 10.1007/s10936-020-09714-7

Kim, S., and Kim, D. J. (2021). Structural relationship of key factors for student satisfaction and achievement in asynchronous online learning. *Sustainability* 13:6734. doi: 10.3390/su13126734

Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford publications.

Kock, N., and Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *J. Assoc. Inform. Syst.* 13:40. doi: 10.17705/1jais.00302

Kuo, Y. C., Walker, A. E., Schroder, K. E., and Belland, B. R. (2014). Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *Internet High. Educ.* 20, 35–50. doi: 10.1016/j.iheduc.2013.10.001

Lai, C. (2019). Technology and learner autonomy: An argument in favor of the nexus of formal and informal language learning. *Annu. Rev. Appl. Ling.* 39, 52–58. doi: 10.1017/S0267190519000035

Landrum, B. (2020). Examining students' confidence to learn online, self-regulation skills and perceptions of satisfaction and usefulness of online classes. *Online Learn.* 24, 128–146. doi: 10.24059/olj.v24i3.2066

Lee, J. S. (2019). Informal digital learning of English and second language vocabulary outcomes: Can quantity conquer quality? *Br. J. Educ. Technol.* 50, 767–778. doi: 10.1111/bjet.12599

Lee, J. S. (2021). *Informal Digital Learning of English: Research to Practice*. London: Routledge, doi: 10.4324/9781003043454

Lee, J. S., and Drajati, N. A. (2019). Affective variables and informal digital learning of English: Keys to willingness to communicate in a second language. *Aust. J. Educ. Technol.* 35, 168–182. doi: 10.14742/ajet.5177

Lee, J. S., Kiaer, J., and Jeong, S. (2023). The role of informal digital learning of Korean in KFL students' willingness to communicate. *J. Multiling. Multicult. Dev.* 1–17. doi: 10.1080/01434632.2023.2216671

Lee, J. S., and Sylvén, L. K. (2021). The role of Informal Digital Learning of English in Korean and Swedish EFL learners' communication behaviors. *Br. J. Educ. Technol.* 52, 1279–1296. doi: 10.1111/bjet.13082

Lee, J. S., Yeung, N. M., and Osburn, M. B. (2022). Foreign Language Enjoyment as a mediator between Informal Digital Learning of English and willingness to communicate: A sample of Hong Kong EFL secondary students. *J. Multiling. Multicult. Dev.* 1–19. doi: 10.1080/01434632.2021.2020802

Lei, J., and Lin, T. (2022). Emergency online learning: The effects of interactional, motivational, self-regulatory, and situational factors on learning outcomes and continuation intentions. *Int. Rev. Res. Open Distrib. Learn.* 23, 43–60. doi: 10.19173/irrodl.v23i3.6078

Lim, C. L., Ab Jalil, H., Ma'rof, A. M., and Saad, W. Z. (2020). Self-regulated learning as a mediator in the relationship between peer learning and online learning satisfaction: A study of a private university in Malaysia. *Malays. J. Learn. Instruct.* 17, 51–75. doi: 10.32890/mjli2020.17.1.3

Liu, G., Ma, C., Bao, J., and Liu, Z. (2023). Toward a model of informal digital learning of English and intercultural competence: A large-scale structural equation modeling approach. *Comput. Assist. Lang. Learn.* 1–25. doi: 10.1080/09588221.2022.2164778

Liu, G. Z., Rahimi, M., and Fathi, J. (2022). Flipping writing metacognitive strategies and writing skills in an English as a foreign language collaborative writing context: A mixed-methods study. *J. Comput. Assist. Learn.* 38, 1730–1751. doi: 10.1111/jcal.12707

Martin, F., and Bolliger, D. U. (2022). Developing an online learner satisfaction framework in higher education through a systematic review of research. *Int. J. Educ. Technol. High. Educ.* 19, 1–21. doi: 10.1186/s41239-022-00355-5

Miller, Z. F., and Godfroid, A. (2020). Emotions in incidental language learning: An individual differences approach. *Stud. Sec. Lang. Acquis.* 42, 115–141. doi: 10.1017/S027226311900041X

Mohammadi Zenouzagh, Z., Admiraal, W., and Saab, N. (2023). Learner autonomy, learner engagement and learner satisfaction in text-based and multimodal computer mediated writing environments. *Educ. Inform. Technol.* [Online ahead of print]. doi: 10.1007/s10639-023-11615-w

Moore, J. C. (2005). *The Sloan Consortium Quality Framework and the Five Pillars*. Boston, MA: The Sloan Consortium. doi: 10.4018/978-1-59140-555-9.ch245

Nashaat, N., Abd El Aziz, R., and Abdel Azeem, M. (2021). The mediating role of student satisfaction in the relationship between determinants of online student satisfaction and student commitment. *E-Learn. High. Educ.* 2021:404947. doi: 10.5171/2021.404947

Ngah, A. H., Kamalrulzaman, N. I., Mohamad, M. F. H., Rashid, R. A., Harun, N. O., Ariffin, N. A., et al. (2022). The sequential mediation model of students' willingness to continue online learning during the COVID-19 pandemic. *Res. Pract. Technol. Enhanc. Learn.* 17, 1–17.4. doi: 10.1186/s41039-022-00188-w

Nikou, S., and Maslov, I. (2023). Finnish university students' satisfaction with e-learning outcomes during the COVID-19 pandemic. *Int. J. Educ. Manage.* 37, 1–21. doi: 10.1108/IJEM-04-2022-0166

Öztürk, M., and Çakıroğlu, Ü. (2021). Flipped learning design in EFL classrooms: Implementing self-regulated learning strategies to develop language skills. *Smart Learn. Environ.* 8:2. doi: 10.1186/s40561-021-00146-x

Pangarso, A., and Setyorini, R. (2023). The drivers of e-learning satisfaction during the early COVID-19 pandemic: Empirical evidence from an Indonesian private university. *Cogent Educ.* 10:2149226. doi: 10.1080/2331186X.2022.2149226

Paposa, K. K., and Paposa, S. S. (2023). From brick to click classrooms: A paradigm shift during the pandemic—Identifying factors influencing service quality and learners' satisfaction in click classrooms. *Manage. Labor Stud.* 48, 182–196. doi: 10.1177/0258042X211066234

Passiatore, Y., Pirchio, S., Oliva, C., Panno, A., and Carrus, G. (2019). Self-efficacy and anxiety in learning English as a Foreign language: Singing in class helps speaking performance. *J. Educ. Cult. Psychol. Stud. (ECPS J.)* 20, 121–138. doi: 10.7358/ecps-2019-020-passi

Pham, L., Limbu, Y. B., Bui, T. K., Nguyen, H. T., and Pham, H. T. (2019). Does e-learning service quality influence e-learning student satisfaction and loyalty? Evidence from Vietnam. *Int. J. Educ. Technol. High. Educ.* 16, 1–26. doi: 10.1186/s41239-019-0136-3

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. doi: 10.1037/0021-9010.88.5.879

Posner, M. I., and Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Dev. Psychopathol.* 12, 427–441. doi: 10.1017/S0954579400003096

Quang, N. N., Linh, P. N., and Hien, N. T. T. (2022). Tasks, self-efficacy, and L2 motivational self system in an online emergency EFL speaking class: A mixed-methods study. *Jalt Call J.* 18, 1–33. doi: 10.29140/jaltcall.v18n1.518

Qutob, M. M. (2018). The relationship between EFL learners' satisfaction within the classroom environment and their speaking skills. *English Lang. Teach.* 11, 116–124. doi: 10.5539/elt.v11n7p116

Rahimi, M., and Fathi, J. (2022). Employing e-tandem language learning method to enhance speaking skills and willingness to communicate: The case of EFL learners. *Comput. Assist. Lang. Learn.* 1–37. doi: 10.1080/09588221.2022.2064512

Ranadewa, D. U. N., Gregory, T. Y., Boralugoda, D. N., Silva, J. A. H. T., and Jayasuriya, N. A. (2021). Learners' satisfaction and commitment towards online learning during COVID-19: A concept paper. *Vision* 27:09722629211056705. doi: 10.1177/09722629211056705

Rashidi, N., and Moghadam, M. (2014). The effect of teachers' beliefs and sense of self-efficacy on Iranian EFL learners' satisfaction and academic achievement. *Tesl-Ej* 18:23.

Roca, J. C., and Gagné, M. (2008). Understanding e-learning continuance intention in the workplace: A self-determination theory perspective. *Comput. Hum. Behav.* 24, 1585–1604. doi: 10.1016/j.chb.2007.06.001

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivation: Classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020

Sagarra, N., and Zapata, G. C. (2008). Blending classroom instruction with online homework: A study of student perceptions of computer-assisted L2 learning. *ReCALL* 20, 208–224. doi: 10.1017/S0958344008000621

Saint, J., Fan, Y., Gašević, D., and Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Comput. Educ.* 3:100060. doi: 10.1016/j.caeai.2022.100060

Sarstedt, M., Ringle, C. M., and Hair, J. F. (2022). "Partial least squares structural equation modeling," in *Handbook of Market Research*, eds C. Homburg, M. Klarmann, and A. Vomberg (Cham: Springer), 587–632. doi: 10.1007/978-3-319-57413-4_15

Schunk, D. H. (1995). "Self-efficacy and education and instruction," in *Self-Efficacy, Adaptation, and Adjustment: Theory, Research, and Application*, ed. J. E. Maddux (New York, NY: Plenum Press), 281–303. doi: 10.1007/978-1-4419-6868-5_10

Schunk, D. H., and Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Read. Writing Q.* 23, 7–25. doi: 10.1080/10573560600837578

She, L., Ma, L., Jan, A., Sharif Nia, H., and Rahmatpour, P. (2021). Online learning satisfaction during COVID-19 pandemic among Chinese university students: The serial mediation model. *Front. Psychol.* 12:743936. doi: 10.3389/fpsyg.2021.743936

Shen, B., Wang, Y., Yang, Y., and Yu, X. (2023). Relationships between Chinese university EFL learners' academic emotions and self-regulated learning strategies: A structural equation model. *Lang. Teach. Res.* 13621688221144832. doi: 10.1177/13621688221144832

Shen, D., Cho, M. H., Tsai, C. L., and Marra, R. (2013). Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. *Internet High. Educ.* 19, 10–17. doi: 10.1016/j.iheduc.2013.04.001

Shih, H. F., Chen, S. H. E., Chen, S. C., and Wey, S. C. (2013). The relationship among tertiary level EFL students' personality, online learning motivation and online learning satisfaction. *Proc. Soc. Behav. Sci.* 103, 1152–1160. doi: 10.1016/j.sbspro.2013.10.442

Soyoof, A., Reynolds, B. L., Vazquez-Calvo, B., and McLay, K. (2023). Informal digital learning of English (IDLE): A scoping review of what has been done and a look towards what is to come. *Comput. Assist. Lang. Learn.* 36, 608–640. doi: 10.1080/09588221.2021.1936562

Supervia, U. P., Bordás, S. C., and Robres, Q. A. (2022). The mediating role of self-efficacy in the relationship between resilience and academic performance in adolescence. *Learn. Motivat.* 78:101814. doi: 10.1016/j.lmot.2022.101814

Suryandani, P. D., and Santosa, M. H. (2021). North Bali students' online learning self-efficacy, engagement, and satisfaction. *Vision* 10, 47–52. doi: 10.21580/vjv10i28921

Tang, Y., Tseng, H., and Tang, X. (2022). The impact of information-seeking self-efficacy and online learning self-efficacy on students' performance proficiency. *J. Acad. Librariansh.* 48:102584. doi: 10.1016/j.acalib.2022.102584

Teng, L. S. (2021). Individual differences in self-regulated learning: Exploring the nexus of motivational beliefs, self-efficacy, and SRL strategies in EFL writing. *Lang. Teach. Res.* 13621688211006881. doi: 10.1177/13621688211006881

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemp. Educ. Psychol.* 66:101976. doi: 10.1016/j.cedpsych.2021.101976

Torrado, M., and Blanca, M. J. (2022). Assessing satisfaction with online courses: Spanish version of the learner satisfaction survey. *Front. Psychol.* 13:875929. doi: 10.3389/fpsyg.2022.875929

Tsai, C.-L., Cho, M.-H., Marra, R., and Shen, D. (2020). The self-efficacy questionnaire for online learning (SeQoL). *Distance Educ.* 41, 472–489.

Tsai, S. C. (2019). Implementing interactive courseware into EFL business writing: Computational assessment and learning satisfaction. *Interactive Learn. Environ.* 27, 46–61. doi: 10.1080/10494820.2018.1451896

Wei, H. C., and Chou, C. (2020). Online learning performance and satisfaction: Do perceptions and readiness matter? *Distance Educ.* 41, 48–69. doi: 10.1080/01587919.2020.1724768

Wiers-Jenssen, J., Stensaker, B. R., and Gr⊘gaard, J. B. (2002). Student satisfaction: Towards an empirical deconstruction of the concept. *Q. High. Educ.* 8, 183–195. doi: 10.1080/1353832022000004377

Wijaya, T. T., Cao, Y., Bernard, M., Rahmadi, I. F., Lavicza, Z., and Surjono, H. D. (2022). Factors influencing microgame adoption among secondary school mathematics teachers supported by structural equation modelling-based research. *Front. Psychol.* 13:952549. doi: 10.3389/fpsyg.2022.952549

Wolters, C. A., and Brady, A. C. (2020). College students' time management: A self-regulated learning perspective. *Educ. Psychol. Rev.* 33, 1319–1351. doi: 10.1007/s10648-020-09519-z

Wu, W., Hu, R., Tan, R., and Liu, H. (2022). Exploring factors of middle school teachers' satisfaction with online training for sustainable professional development under the impact of COVID-19. *Sustainability* 14:13244. doi: 10.3390/su142013244

Wu, Y., Xu, X., Xue, J., and Hu, P. (2023). A cross-group comparison study of the effect of interaction on satisfaction in online learning: The parallel mediating role of academic emotions and self-regulated learning. *Comput. Educ.* 199:104776. doi: 10.1016/j.compedu.2023.104776

Xu, Q., Wu, J., and Peng, H. (2022). Chinese EFL university students' self-efficacy for online self-regulated learning: Dynamic features and influencing factors. *Front. Psychol.* 13:912970. doi: 10.3389/fpsyg.2022.912970

Yang, C., and Xu, D. (2023). Predicting student and instructor e-readiness and promoting e-learning success in online EFL class during the COVID-19 pandemic: A case from China. *PLoS One* 18:e0284334. doi: 10.1371/journal.pone.0284334

Zalli, M. M. M., Nordin, H., and Hashim, R. A. (2019). The role of self-regulated learning strategies on learners' satisfaction in massive open online course (MOOC): Evidence from Malaysia MOOC. *Int. J. Innovat. Technol. Explor. Eng.* 8, 2286–2290. doi: 10.35940/ijitee.J1138.0881019

Zhang, Y., and Liu, G. (2022). Revisiting informal digital learning of English (IDLE): A structural equation modeling approach in a university EFL context. *Comput. Assist. Lang. Learn.* 1–33. doi: 10.1080/09588221.2022.2055081

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *J. Educ. Psychol.* 81, 329–339. doi: 10.1037/0022-0663.81.3.329

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educ. Psychol.* 25, 3–17. doi: 10.1207/s15326985ep2501_2

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 (London: Academic Press). 3–39. doi: 10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Pract.* 41, 64–70. doi: 10.1207/s15430421tip4102_2

Zimmerman, W. A., and Kulikowich, J. M. (2016). Online learning self-efficacy in students with and without online learning experience. *Am. J. Distance Educ.* 30, 180–191. doi: 10.1080/08923647.2016.1193801

Zou, C., Li, P., and Jin, L. (2022). Integrating smartphones in EFL classrooms: Students' satisfaction and perceived learning performance. *Educ. Inform. Technol.* 27, 12667–12688. doi: 10.1007/s10639-022-11103-7

Learning analytics for enhanced professional capital development: a systematic review

Javier de La Hoz-Ruiz[1], Mohammad Khalil[2]*,
Jesús Domingo Segovia[1] and Qinyi Liu[2]

[1]Department of Didactics and School Organization, University of Granada, Granada, Spain, [2]Center for the Science of Learning & Technology, Faculty of Psychology, University of Bergen, Bergen, Norway

Background/Motivation: This article presents a systematic review aimed at examining the utilization of learning analytics (LA) to enhance teachers' professional capital.

Aim: The study focuses on three primary research questions: (1) exploring the characteristics and approaches of LA in professional capital, (2) investigating suggestions from LA for assessing and improving professional capital, and (3) examining variables studied in enhancing the most intricate dimension of professional capital using LA.

Methodology: To address the research objectives, a systematic review was conducted focusing on the key concepts "learning analytics" and "professional capital." Following the procedures outlined encompassed in four stages: identification, screening, inclusion, and adequacy. The PRISMA 2009 protocol guided the systematic review process.

Principal findings: The findings of the study underscore the efficacy of LA as a catalyst for improving professional capital, particularly through collaborative learning and the utilization of tools like forums and online learning platforms. Social capital emerges as a pivotal component in integrating diverse types of professional capital, fostering opportunities for knowledge creation and social networking.

Conclusion/Significance: In conclusion, the study highlights the paramount significance of addressing teachers' professional capital development through collaborative approaches and leveraging technology, particularly in primary education. The article concludes by emphasizing the imperative for more research and knowledge dissemination in this field, aiming to ensure equity in learning and address the challenges posed by the COVID−19 pandemic.

KEYWORDS

professional capital, learning analytics, social capital, professional learning community, communities of practice

1 Introduction

Achieving goals 4 and 17 of the 2030 Agenda on Sustainable Development Goals (SDGs) is currently a major challenge for education systems and requires the collaborative efforts of teachers, families, communities and students. Goal 4 aims to ensure inclusive and equitable quality education, promoting learning opportunities for all. On the other hand, Goal 17 focuses on strengthening the implementation of SDGs through global partnerships. In this

context, addressing educational challenges (Goal 4) involves the collaboration of teachers, families, communities, and students. This collaboration is essential to ensure equitable access to quality education and contribute to sustainable development. Key indicators include the educational completion rate, access to preschool, the proportion of trained teachers, foreign investments, and international cooperation. The interconnection of these goals highlights the need for coordinated efforts to make a meaningful impact on education and sustainable development.

In 2021, UNESCO highlighted the pressing need for new schools and governance structures with the capacity to energize and unify a community of individuals with the shared purpose of seeking knowledge while being committed to enhancing the quality and equity of education. In this regard, the action of school leaders and the advancement of these schools and communities as extended communities of professional practice are fundamental (Bolívar and Domingo, 2023). The concept of Third Generation Professional Learning Communities is central to this discussion. These communities — in addition to their core goal of improving the learning of all and among all — aim to foster internal capacities for improvement through interactive and committed professionalism (Hargreaves and O'Connor, 2018). This approach also places a significant emphasis on accumulating more professional capital (Hargreaves and Fullan, 2012, 2013, 2020) generated through the extension of the community of practice into to the school environment, local community, and across professional and inter-institutional networks. As a result, there is pressing needs for fluid networks of interrelation, communication, and support for learning for all and for all, with a shared and networked form of leadership, unified by a broad perspective of "leadership from the middle" (Rincón, 2019).

These new scenarios outlined by UNESCO in 2021, coupled with the complexity of the current post-pandemic era and the transition toward a "new normal" (Hargreaves, 2020; Darling-Hammond, 2022), highlight the need for educational institutions to adapt. This adaptation is particularly crucial in challenging and vulnerable contexts (e.g., communities with limited resources, schools with access challenges). In turn, there is a growing need for more innovative and engaged schools, which involves embracing new forms of governance that foster their development as communities of professional practice. Such communities' initiate projects and create cultures and environments with a shared commitment to educational improvement. Thus, the educational improvement largely depends on the ability of school leaders to connect everyone (teachers, families, and the local community) to form Communities of Professional Practice (Kimble and Hildreth, 2008). Professional capital plays a fundamental role in the building these internal capacities for improvement (Hargreaves and Fullan, 2012). This concept is one of the central issues of the present work, together with the sense of sustainability and quality of education.

Education systems are at a crossroads in their efforts to facilitate professional learning for teachers that supports both institutional improvement and ensures quality and equity of learning and sustainable development. Therefore, it is of great interest to the academic field to identify the measures that are being used to extract information about professional capital as well as to investigate the improvement processes. However, few studies have focused on informing the design and implementation of intentional frameworks to enhance teachers' professional capital through the growth of social networks (Yoon et al., 2018). To fill this gap, it is worth highlighting the potential of the emerging field of *learning analytics (hereinafter LA)*.

LA aims to develop tools to raise awareness of the presence of learning activities and processes, i.e., to make such processes available for analysis. LA can be applied to collect and analyze information about teaching activities and can also help to improve professional capital. A review of previous studies on professional capital reveals the existence of sufficient knowledge regarding the utilization of learning analytics for developing skills related to this capital (Tong and Razniak, 2017; Jan et al., 2019; Silva et al., 2020; Demir, 2021; Minga-Vallejo et al., 2021; Yassine et al., 2022). However, to the best of our knowledge, there are no systematic reviews focusing on the application of LA to improving professional capital. While previous works by Tong and Razniak (2017) and Demir (2021) explored aspects of professional capital, they focused on collaborative leadership and social capital, respectively, without specifically considering LA as a means of improvement. Similarly, Jan et al. (2019), Silva et al. (2020), Yassine et al. (2022), and Minga-Vallejo et al. (2021) covered to a large extent Social Network Analysis (SNA) and online learning environments but did not concentrate on professional capital improvement through LA, to name but a few. Despite providing valuable insights, these studies leave a gap in understanding how LA can be specifically employed to enhance professional capital. Recognizing this gap is crucial for guiding future research in this area.

Gaining insight into the most recent studies in this field holds great significance for researchers seeking to uncover potential avenues for future exploration in learning analytics, ultimately contributing to the advancement of professional capital. To that end, we follow a systematic literature review to answer the following research questions:

> *RQ1*: What are the characteristics and approaches of existing learning analytics in professional capital?
>
> *RQ2*: What does learning analytics suggest should be evaluated and analyzed to improve professional capital?
>
> *RQ3*: What is the most untangled dimension of learning analytics in professional capital, and how learning analytics is improving this dimension?

The review study is structured as follows: we begin by establishing a pertinent background, followed by a narrative that reports on our systematic review of the literature. We then presents the results of our synthesis and discuss significant insights and findings. Lastly, we discuss limitations and draw conclusions.

2 Professional capital: human, social, and decisional capital

Professional capital refers to the skills and knowledge that a person possesses and that allow them to perform their work effectively, as well as the relationships and networks they build within the educational environment. Professional capital — the most important factor in social production and activity — is a concept related to the

value of individuals or groups and can be used to enhance long-term growth (Giddens, 1999).

In the educational field, this term refers to the combination of knowledge, skills, and experience that an education professional possesses and that are valuable to their work performance in education. Day et al. (2006) define professional capital as theoretical and practical knowledge acquired through initial and continuing training, as well as professional experience in the educational field, all of which enable the teacher to develop effective pedagogical and didactic skills and competencies. On the other hand, Hargreaves et al. (2002) define it as knowledge, skills and competences acquired through critical reflection and experiential learning, which enable the education professional to make informed decisions and develop effective strategies to improve teaching and learning. Organization for Economic Co-operation and Development (2019) considers this term as the combination of knowledge, skills, competencies, and values that an education professional possesses, and that allows them to exercise their work ethically and effectively in the educational field.

It is worth noting that the world's highest-performing education and economic systems are adopting the strategy of fostering professional capital. Countries and communities investing in professional capital are therefore making a long-term investment in developing human capital.

The key to this concept is systemic development and the integration of three types of capital – human, social, and decisional – into the teaching profession. Professional capital is concerned with collective responsibility (rather than individual autonomy), rigorous training, continuous learning, going beyond the evidence, and being open to the needs and priorities of students and society (Hargreaves and Fullan, 2012).

From this standpoint, developing good teachers for all students requires teachers to be highly committed, well prepared, engaged in continuous training, adequately paid, and involved in good teamwork to maximize their own progress and make effective judgments by using all their ability and experience (Hargreaves and O'Connor, 2018).

From an economic perspective, the process of teacher professional development is a worthwhile, long-term investment that will add to the net value of professional capital. By investing in innovative, professional, and high-quality teachers, their professional capital can be increased and circulated, and the teachers will be expected to introduce significant innovations into their teaching practice (Liu et al., 2020). To implement innovative teaching in a challenging educational environment, teachers must have high levels of professional capital and make appropriate investments in professional practice to improve performance.

According to Tong and Razniak (2017), the development of effective professional capital requires collaborative leadership, professional development, and adult learning. As the collaborative culture gains momentum within the school environment, greater collaboration promotes inclusion, trust, risk-taking, and fosters connectivity among staff. According to DuFour (2003), capacity building through collaborative teamwork is important for cultivating a positive learning environment for all. One of the important factors that administrators should consider is being aware of the challenges within their school community. In other words, teacher engagement can begin to generate professional capital (Hargreaves and Fullan, 2012), i.e., teaching wisdom, collaborative ability, and mastery of the

content they develop, which can extend from the cloud to the classroom (Hu et al., 2018).

To view the entire process of school education from the perspective of professional capital, it is important to highlight the three dimensions indicated by Hargreaves and Fullan (2012): (a) human, (b) decision-making and (c) social capital.

2.1 Human capital

Human capital within the field of education refers to the set of skills, knowledge, experiences, and competencies that people possess and that allow them to exercise their educational work effectively. Importantly, human capital in education refers not only to the skills and knowledge of education professionals, but also to the skills and competencies of students that are the result of the education and training they receive (Coleman, 1988).

Some of the most notable definitions of human capital within the educational area include that of Schultz (1961), who defines this construct as the set of knowledge, skills, and values that individuals possess and that have been acquired through education, training and experience, which allow them to perform effectively in the educational field. For Becker (2010), human capital is the set of competencies and skills that enable individuals to perform effectively in the workplace, including the ability to adapt to changing situations and learn continuously. In the case of education professionals, this refers to their ability to enhance student learning.

2.2 Decision-making capital

The term "decisional capital" refers to the power that a person or entity has to make important and strategic decisions in an organization, company, or institution. Some important definitions related to this concept are presented by authors such as Simon (1977), and Grant (1996), by which they highlight the influence of a person, group, or entity when it comes to making important decisions that impact the functioning of an organization. In business terms, Nonaka and Takeuchi (1995) conceive this power as the ability of a person or entity to make decisions and carry out actions that significantly determine the success or failure of a company or project in the short and long term.

The significance of decisional capital within the broader framework of professional capital has been addressed in the literature. For example, McKenzie et al. (2011) assert that improving the governance of the decision-making process will yield progressive benefits through the meticulous planning of developmental initiatives. In addition, Visone (2018) shows how leaders who trust teachers, value their contributions, and provide opportunities for decision-making and leadership also supported the development of decisional capital and social capital, a concept that will be the focus of the next section.

2.3 Social capital

In education, social capital refers to the benefits that individuals may derive from their connections and relationships with others

(Coleman, 1988) by allowing them access to assets such as information, advice, experience, materials, and confidence that can facilitate positive changes in teachers' beliefs and practices (e.g., Daly et al., 2010; Penuel et al., 2013). As Leana (2011) states, high social capital generates an increase in human capital. Thus, for example, if efforts are focused on increasing individual talent, much more work will be required to build social capital. In contrast, individuals gain confidence, learn, and receive feedback by being surrounded by the right types of people and having the appropriate relationships and interactions in their environment.

A growing body of work has sought to identify the mechanisms that facilitate social relations among teachers, pointing to various characteristics of educational infrastructure that support social capital as a source of development in schools, such as grade level assignment and formal leadership positions (Spillane et al., 2015). However, it is important to recognize that normative dimensions play a pivotal role in fostering social relationships. In schools where teachers adhere to shared norms such as trust and collective responsibility, in schools where teachers adhere to shared norms such as trust and collective responsibility, there is a greater likelihood of improvement (Bryk and Schneider, 2002).

2.4 Learning analytics and professional capital

Learning analytics is defined as the measurement, collection, analysis and presentation of data on students and their contexts, in order to understand and optimize learning and the environments in which it occurs (Long and Siemens, 2011). LA has considerable value because it can be used as a means to extract methodologies and more effective processes and tools in data measurement, collection, analysis, and reporting of professional capital (Khalil and Ebner, 2016). According to Khalil and Ebner (2016), the methods of LA can be categorized as follows: (a) data mining techniques; (b) statistics and mathematics; (c) text mining, semantics, and linguistic analysis; (d) display; (e) social network analysis; (f) qualitative analysis; and (g) gamification.

There are several discussions including institutional reports such as D2L's "The State of Learning Analytics in 2020" (2020) demonstrating the importance of the field in education. This report brings insight on the increasing usage of data to personalize education and improve the student experience. The field is also strongly related to Artificial Intelligence (AI) where LA depends on methods of AI to understand learning; for example, machine learning and related data-driven approaches. For that reason, it should be noted that LA is a combination of different disciplines such as computer science, statistics, psychology, and education.

In general terms, learning analytics allows us to identify opportunities for improvement in the teaching and learning process. This can help professionals identify the skills and knowledge they need to improve their performance and therefore their professional capital. In addition, LA can help generate professional capital by improving the teaching and learning process, enhancing the retention and completion of training programs, and identifying trends and patterns in the workforce.

Having explained the key variables, the present study proposes how these should be worked on by identifying the lines of research and their interconnections based on the information contained in the databases via carrying out a systematic literature review to understand the structure and knowledge gaps of the scientific domain.

3 Methodology

In order to respond to the research objectives, a systematic review was conducted, focusing on the two key concepts to be analyzed: "learning analytics" and "professional capital." We adopted the procedures of Kitchenham and Charters (2007) and Pollock and Berge (2018) when carrying out the systematic review, which comprises four stages: identification, screening, inclusion, and adequacy. The PRISMA 2009 protocol was also adopted as a guide for producing this systematic review (Liberati et al., 2009).

3.1 Identification and screening

3.1.1 Identification: data source

The following databases and their rational were utilized in this study:

(1) Web of Science (WOS) is one of the most reputable collections of journal articles, indexing both the Social Science Citation Index (SSCI) and Science Citation Index (SCI).[1]
(2) Scopus is a database of great international relevance. Like the SCI, it not only collects bibliographic information, but also analyzes the behavior of the citations received by journals, which allows generating a large number of bibliometric and citation indicators, such as the h index.
(3) ACM is the database chosen for the topic of this study, being "the most complete full-text database in the world for articles and bibliographic literature on computing and information technology." The annual proceedings of the Learning Analytics and Knowledge Conference (LAK) are published in the ACM Digital Library.

3.1.2 Search and screening strategy

The following key terms were integrated in the systematic review and used in the search formula: ALL FIELDS/ (ALL = "learning analytic*" OR ALL = "academic analytic*" OR ALL = "teaching analytic*") AND (ALL = "social capital" OR ALL = "human capital" OR ALL = "decisional capital" OR ALL = "professional capital") in the three databases, after duplicate citations, 657 articles were extracted at this initial stage. The words used in the search equation are the general terms related to the focus of the study (learning analytics, professional capital) accompanied by their sister terms within the area (academic analytics, teaching analytics, social capital, human capital, decisional capital). We decided to include "learning analytics," "academic analytics," and "teaching analytics" to expand our search umbrella since various authors have debated conceptual differences between the three terms, but common practice often employs them interchangeably.
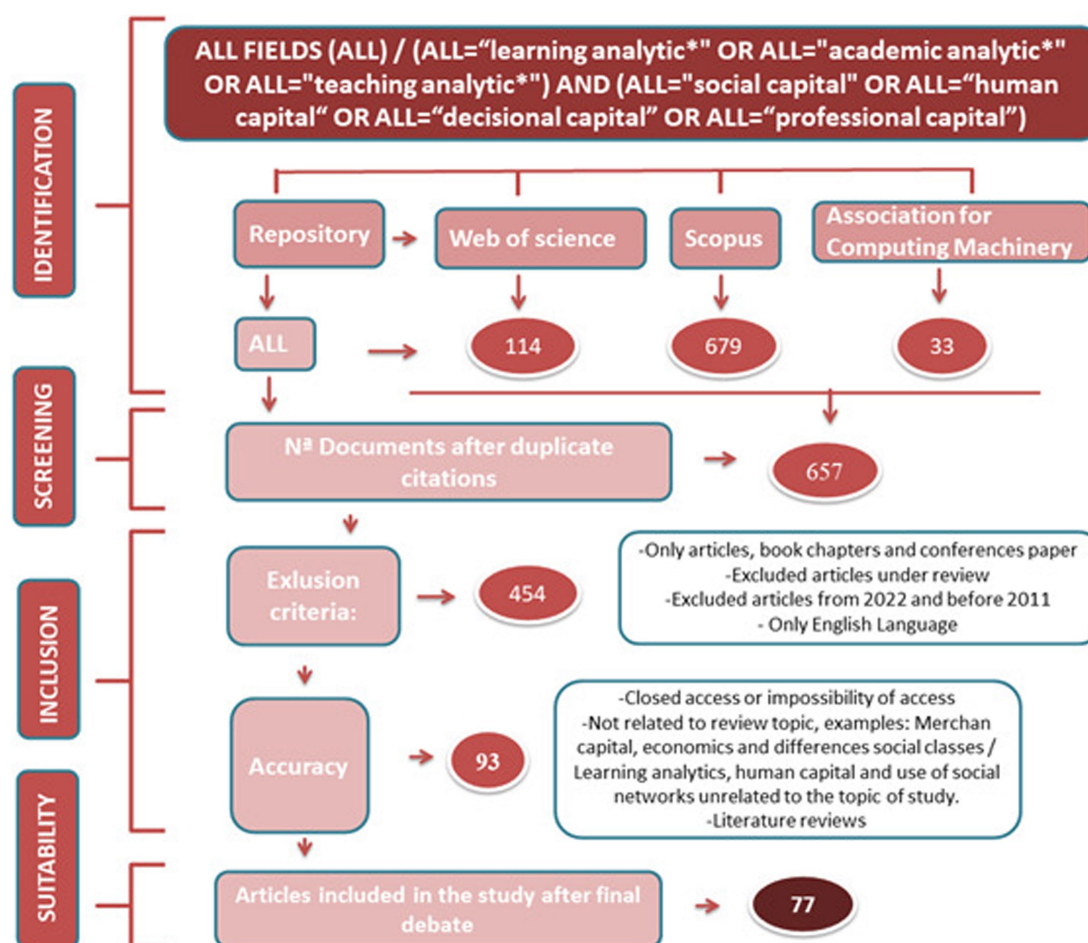
---

1   https://webofscience.com

FIGURE 1
Search and selection process of studies to be reviewed.

By incorporating the three, we aim to encompass a broader spectrum of research and relevant resources. The search was conducted on February 23, 2022, after which the articles were subjected to the process of reading, screening, and analysis. Figure 1 shows the search and selection process for the studies reviewed.

3.2 Inclusion and eligibility criteria

After specifying the above search terms, the search was further narrowed by applying the following criteria as described in Table 1, resulting in a final corpus of 77 articles.

We also adopted a quality assessment as referred by Schön et al. (2017). Table 2 displays the checklist used to assess the quality of the included studies. All primary studies (77 documents) were evaluated on the basis of quality indicators.

The first item (QA1) evaluates the purpose of each study. This question was answered positively in 82% of the studies. The second point (QA2) measures whether the study presents a detailed description of the approach, and the answer to this question was positive in 77% of the studies. The third item (QA3) asks about a method of validating the result, with only 21% of studies using adequate validation methods. The fourth point (QA4) evaluates

whether the studies are based on opinions or points of view. Only 31% of studies responded positively. Finally, the fifth item (QA5) looks at the number of citations received by studies, and the answers demonstrated that 53% of the studies had more than five citations in other studies.

3.3 Limitations

Following the quality guidelines for systematic reviews (Moher et al., 2009; Alexander, 2020), we established the inclusion and exclusion criteria, while recognizing the complexities and implications of these. We acknowledge that the review uses only three databases. This systematic review could have benefited from other databases as well as other impact indicators. We also acknowledge that the language represents another bias in the databases, as these repositories predominantly consider English-speaking articles.

Additionally, the study's reliance on current platforms might overlook emerging technologies that could influence these domains. The exclusive emphasis on primary education raises questions about the applicability of findings across different educational levels. Cultural influences on the implementation and effectiveness of learning analytics strategies may not be fully captured. Future research

TABLE 1 Inclusion and exclusion criteria of the systematic review.

Criteria	Inclusion	Exclusion
Topic and focus of study	Use of Learning analytics to improve professional capital in education.	Mercantile capital, economy, and different social classes. Learning analytics, human capital, and use of social networks not related to the subject of study.
Language	English	Other language
Publication period	January 2011–December 2021	Articles excluded 2022 onward and those before 2011 (the field of learning analytics emerged in 2011)
Type of publication	Articles, book chapters, and conference communications	Books, posters, workshop documents, editorials, and reports.
Publication status	Peer-reviewed articles	Non-peer-reviewed and in-press articles
Other	Accessible	Inaccessible, and literature reviews

should consider addressing these limitations to offer a more nuanced and widely applicable understanding of the dynamics between learning analytics and professional capita.

4 Findings and discussion

This systematic review aimed to address the following research questions:

4.1 What are the characteristics and approaches of existing learning analytics in professional capital? (RQ1)

Figure 2 reveals a significant increase in scientific production over the last ten years. In addition, these findings highlight that in the year in which the SDGs emerged (2015), and the explosion of the COVID-19 pandemic (2019), the publications peaked, with 17 articles in 2015 and 19 in 2019, although it should be noted that 2012 (the year following the appearance of LA) saw a third peak of scientific production, with 14 articles.

This Figure 2 also shows the scientific production with respect to four emerging themes in recent years. These are sustainable goals, informal learning, collaborative learning, and COVID-19, all of which have seen a considerable increase in published output from 2015 to the present, particularly Sustainable Development Goals. However, from 2011 to the present, the highest number of publications were related to informal learning (19), followed by SDGs (17), collaborative learning (9), and finally COVID-19 (3).

LA within professional capital can be seen as an indicator of improved quality and effectiveness of learning (Hu et al., 2018).

TABLE 2 Quality criteria used to assess the adequacy of the study.

Item	Assessment criteria	Score	Description
QA1	Were the objectives of the research clearly stated?	−1	The objectives were not described.
		0	The objectives were partially but unclearly described
		1	Yes, the objectives were well described and clear
QA2	Does the article include a detailed description of the proposed solution or approach?	−1	No, details were missing
		0	Partially, if you wish to use the approach or solution, you must read the references
		1	Yes, the approach can be used based on the presented details
QA3	Is the proposed solution or approach valid?	−1	No
		0	It was partially validated in a laboratory, or only portions of the proposal were validated
		1	Yes, by a case study
QA4	Does the article present an opinion or viewpoint?	−1	Yes
		0	Partially because the corresponding work was explained, and the work was set into a specific context
		1	No, the paper was based on research
QA5	Has the study been cited in other scientific publications?	−1	No, no one cited the study
		0	Partially. Between one and five scientific papers cited the study
		1	Yes, more than five scientific papers cited the study.

However, there are other issues that can also help us understand this observation and that must be taken into account. For example, de Laat and Schreurs (2013) show that informal learning has a strong relationship with the increase in professional capital. However, as powerful as informal learning can be, a challenge arises when attempting to use it for professional development. Informal learning activities are mostly implicit, spontaneous, and invisible to others, and as such, this problem presents an interesting challenge for the LA field,
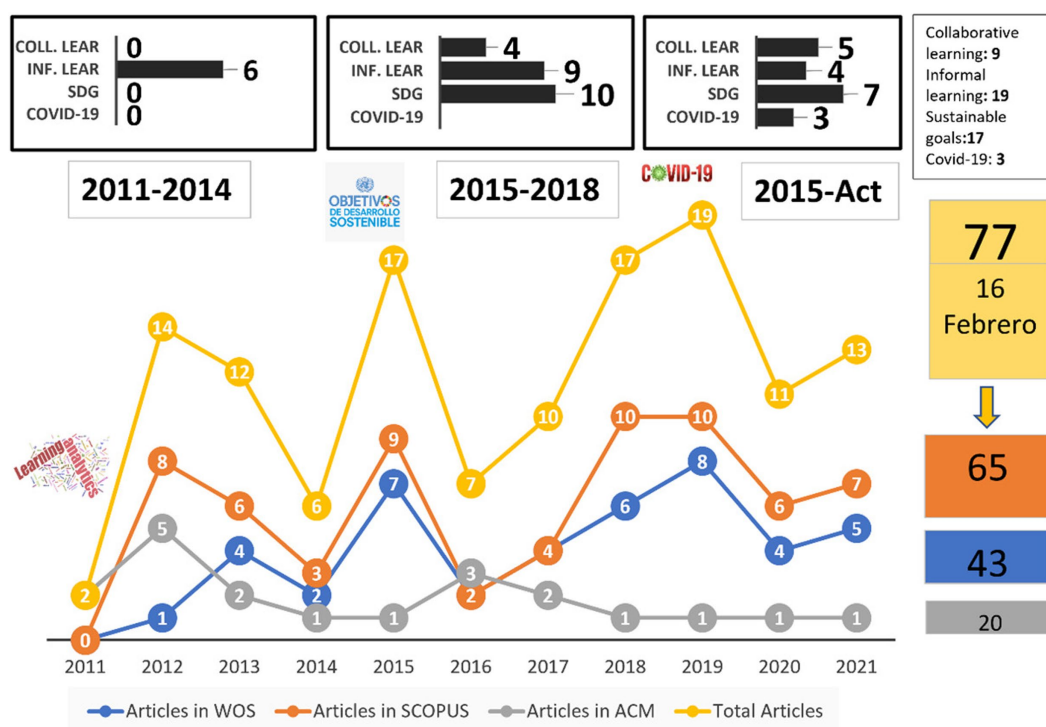
FIGURE 2
Number of accumulated publications on learning analytics in the professional capital of teachers according to source, emerging themes, and key time points.

that is, finding ways to capture and analyze traces of informal (social) learning in everyday life and networks (Cross et al., 2004).

Similarly, collaborative learning between professionals emerges, which can begin to generate professional capital (Hargreaves and Fullan, 2012), that is, teaching wisdom, collaboration, and mastery of the content they develop, which can extend from the cloud to the classroom. While much of learning analytics can focus on the classroom and/or educational institution, many transformations are also taking place in learning on and through the web to achieve collaborative output, and such changes are therefore equally open to inquiry from learning, network, and analytical perspectives, as a considerable amount of information can be extracted from social networks (Haythornthwaite, 2011).

Stoll et al. (2010) has demonstrated the importance of focusing on collective knowledge and growth that permeates community life. Through continuous inquiry and reflection, members are encouraged to seek new knowledge by continually examining their practices and engaging in thoughtful dialog, applying new ideas, solving problems, and finding solutions.

The concept of collaborative learning leads us to the term *network learning* (LeCun et al., 2015). This refers to the collective advancement of knowledge and the development of shared identities that come together in the community aspect of social learning, based on the well-known concept of communities of professional practice — a theme of central relevance in the collaborative learning literature in recent years (Lave, 1988). Therefore, the collaborative learning concept tends to be closely linked to the improvement of professional capital through the use of LA.

As mentioned, the results show an increase in scientific production in this area of research during the last ten years, identifying the key moments and significant events that help us understand the evolution of this production. For example, the introduction of the SDGs in 2015 coincided with a surge in published output in this field, which can be explained by the fact that the educational approach adopted to achieve the targets of Objective 4 (quality in education) involves creating collaborative capacities among teachers through the construction of professional capital.

Thus, developing these skills in preparation for a career requires a sustained and progressive growth of professional habits. The community of professional practice represents an alternative, informal way to achieve this goal (Cook et al., 2012), a term that is repeated again in the literature because it fosters a new way of learning as students observe and emulate mentors, while engaging in a cycle of "learning to be" in order to master a particular discipline (Khousa and Atif, 2018).

Subsequently, another relevant event in the trajectory of this field is the Covid−19 pandemic. This crisis has had a profound impact on education. In the context of the pandemic, professional capital has become even more important, as there have been significant changes in the way people work, along with the skills and knowledge needed to adapt to these changes. Pedroso et al. (2021) found that those school leaders who had greater professional capital were more effective in their response to the pandemic.

In Figure 3 uses the Sankey diagram is used to identify how scientific production is distributed by type of document, database, quartile, and nationality of the authors.

FIGURE 3
Number of publications by type of document, quartile, and author nationality.
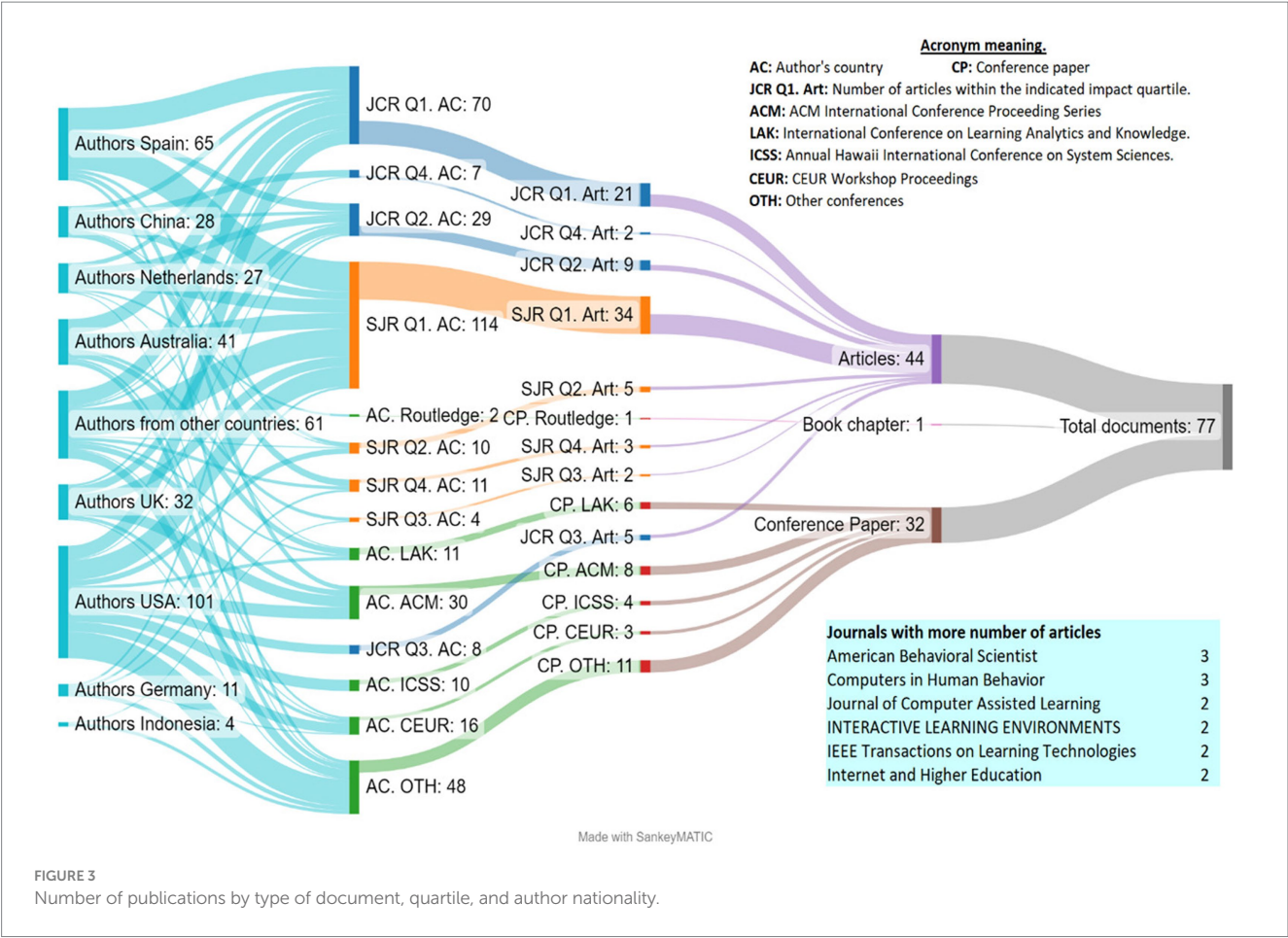
First, moving from left to right in Figure 4, it is clear that most of the documents are journal articles ($n = 44$), followed by conference communications ($n = 32$), and only one book chapter. Of the 44 articles, most were published in Quartile 1 of the SJR ($n = 34$) and Quartile 1 of the JCR ($n = 21$), and we can see that most of the remaining articles have an impact factor. Continuing with the diagram, in terms of the number of authors per quartile based on their author's affiliated, and in accordance with the previous findings, most of the authors appear in Quartile 1 of the JRS ($n = 114$ authors) and Quartile 1 of the JCR ($n = 70$ authors). The most predominant nationalities are the USA ($n = 101$ Authors), Spain ($n = 65$), UK ($n = 32$), China ($n = 28$), and the Netherlands ($n = 27$). Finally, the journal with the highest number of articles on the topic is American Behavioral Scientist and the *Journal of Professional Capital and Community*.

Second, regarding the 32 conference communications, the ACM Conference on Web Search and Data Mining (8 documents) and The International Learning Analytics and Knowledge Conference (6 documents) are noteworthy, where American authors account for most of the published conference proceedings, followed by the Germans and the British.
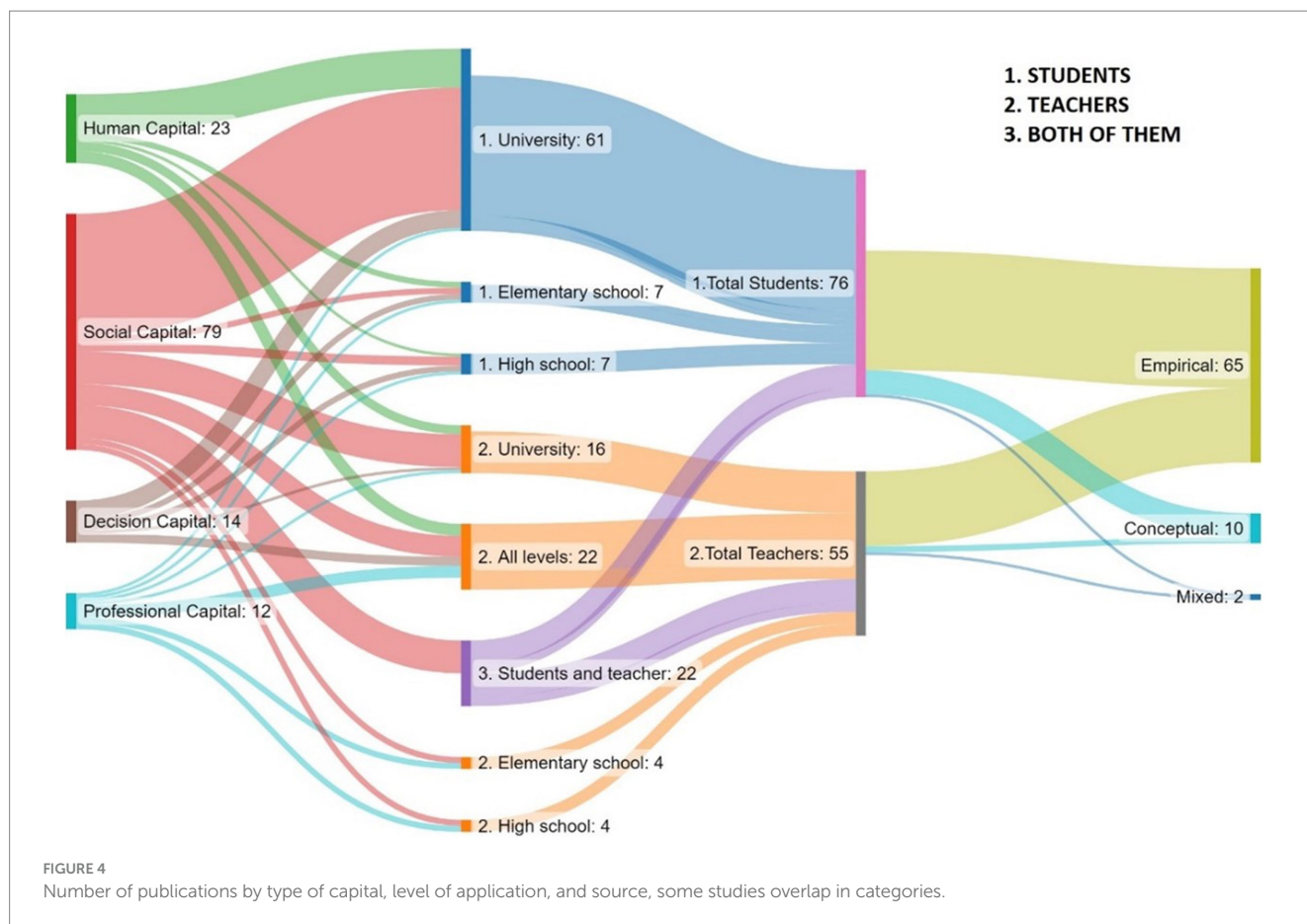
Third, only one book chapter was included (with American authorship from the Routledge publishing house).

LA is not only associated with the improvement of professional capital, but also in other areas, occupying the most relevant positions in high-impact databases. As later confirmed by Paraschiv et al. (2016), the speed at which new scientific articles are published has increased drastically, as well as the process of monitoring the most recent high-impact publications. LA is even present in the formulation of new educational policy reforms in countries such as China (Yu et al., 2016). In the United States, education and government are redefining their partnerships and working together to create competency-based, industry-driven education at the local, state, and national levels through LA methodologies (Baumann et al., 2014), which can give us an insight into the rise of specific and focused conferences on LA in the United States.

In line with the data shown in Figure 3, studies such as those of Monés et al. (2020), and Muñoz-Merino et al. (2022) agree in showing the increasing importance of LA in the Spanish context, in addition to its effectiveness in the use and improvement of the quality of education provided by teachers (Michos et al., 2020; Llopis-Albert and Rubio, 2021). Consequently, this is one of the focal points of Figure 5, a Sankey diagram which indicates that most of the articles are empirical ($n = 65$), followed by conceptual ($n = 10$), and mixed ($n = 2$).

Continuing in the same lines, the following categorization shows the focus of the study, i.e., whether the research focuses on students or teaching staff. On a total of 76 occasions, different groups of students are studied at varying levels, while on 55 occasions the teaching staff are the focus of the study. It should be noted that conceptual studies tend to focus on students, with a notable lack of conceptual studies focused on teachers.

FIGURE 4
Number of publications by type of capital, level of application, and source, some studies overlap in categories.

The bulk of scientific output centered on the student demographic is primarily directed toward university-level students, with a shortage of research on school students. However, teaching staff studies are more balanced, addressing all levels of education. Moreover, there is a notable predominance of publications at the university level (albeit to a lesser extent than the student-focused articles), and, as noted above, there is again a lack of research related to professional capital and learning analytics at the primary education level. Finally, the articles that focus on both teachers and students address these populations together and at all levels.

The last block on the right shows the number of investigations on the various types of professional capital. At first glance, it is evident that social capital accounts for the majority of the studies, while there is a similar distribution of human, decisional, and professional (general) capital studies.

Consequently, it is evident that conceptual studies are aimed at improving social capital in students, suggesting a gap in terms of translating this work to teachers, while providing a foundation upon which productive learning can be established (Kovanovic et al., 2014).

Demir's (2021) highlights the need for more research on the relationship between these dimensions and the organizational structure of schools to promote the desired outcomes of teacher social capital. Based on this, empirical studies show the considerable importance of social capital within the broader spectrum of professional capital (Hargreaves and Fullan, 2012), leading to the adequate exchange of knowledge and practices. These dynamics contribute to the creation of challenging learning environments that
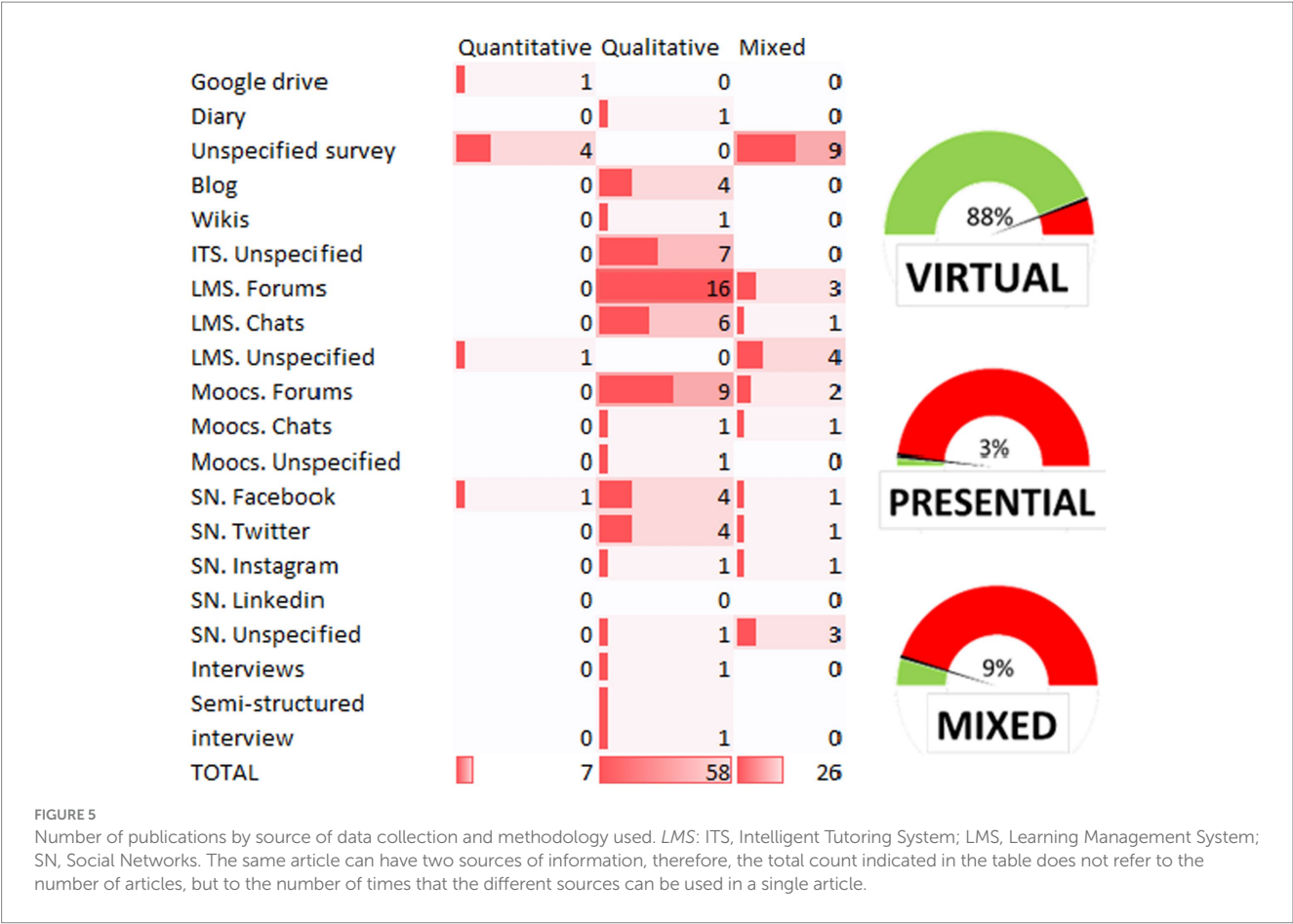
are conducive to increased creativity, thereby enabling the acquisition of fresh insights into student success.

As a result, there is a gap in terms of studies on social capital at elementary and primary levels. An argument can be made for including social capital as an explicit component of the ability of community schools to use data on student outcomes to increase student success. On this basis, several articles support Newmann et al.'s (1997) description of the organizational capacity of schools to meet performance expectations. For instance, Smylie and Evans (2006) highlight the role of Coleman's (1988) concept of social capital for policy implementation, while studies such as those of Yoon et al. (2018) found a relationship between the presence of forms of social capital as part of organizational capacity and the frequency and extent of data usage among teachers and administrators.

In light of research on organizational learning, it appears that social capital provides opportunities for the creation of new knowledge, such as possible solutions to persistent problems of student success, and research on organizational routines as mechanisms for change and preservation in organizations (Kerrigan, 2015).

Another argument is the ability of social capital to lay the foundations for the other two types of professional capital (human and decisional). Evidence for this possibility can be found in the study by Coleman (1988), who demonstrated that the effect of social capital is especially important in the creation of human capital, while social capital in both the family and the community plays a role in creation of human capital in the next generation. In this context, universities

FIGURE 5
Number of publications by source of data collection and methodology used. *LMS*: ITS, Intelligent Tutoring System; LMS, Learning Management System; SN, Social Networks. The same article can have two sources of information, therefore, the total count indicated in the table does not refer to the number of articles, but to the number of times that the different sources can be used in a single article.

can use learning analytics to help teachers understand and monitor pedagogical practices that are designed to build social relationships among students and actively engage them in their learning environment (Carceller et al., 2013).

LA methods can be employed to identify the potential impact of socialization efforts of active participation in "learning relationships" for professional development. In this regard, LA is aim at formulating tools to raise awareness of the presence of learning activities and processes. This enables the subsequent collection and analysis of such processes to ultimately improve activities, all of which are key elements tied to the next question to be addressed (Ferguson and Buckingham, 2012; Siemens, 2013).

4.2 What does learning analytics suggest should be evaluated and analyzed to improve professional capital? (RQ2)

To respond to this question, three perspectives are raised: what dynamics are most frequently used for the collection of these data? How and what are the techniques employed to extract useful information from these data? And, finally, what strategies are being employed to improve professional capital? In order to delve deeper into the first issue, Figure 5 is presented.

First, this figure visualizes the data collection methods in the form of a speedometer, showing that in 88% of the cases the data is extracted

virtually, 9% use mixed methods (physical and virtual), and only 3% of the studies collect data entirely in person.

The horizontal bar in the table indicates the methodology used, while the vertical bar refers to the type of source or tool used for collection. Regarding the type of methodology, qualitative methods predominate, appearing on 59 occasions in the documents analyzed, followed by mixed methodologies (26 times), and qualitative methods (7 times).

Moving to the vertical axis, we can first observe which of the qualitative sources were used most frequently, following the categorizations presented in Figure 1. It is noteworthy that both LMS platforms ($n = 16$) and MOOCs ($n = 9$) are used for data extraction, followed by LMS chats ($n = 6$). It is also worth highlighting the homogenous distribution of data gathered from social networks (Facebook $n = 4$/Twitter $n = 4$) and blogs ($n = 4$). This is followed by mixed methodologies, where unspecified questionnaires ($n = 9$) are used together with LMSs without specifically specifying the source ($n = 4$), as in the case of social networks ($n = 3$). Finally, in the quantitative column, the data are collected from unspecified questionnaires ($n = 4$).

It is important to note the significance of the data collection method, which is useful for defining individual, school, and systematic objectives for professional development. Furthermore, this process offers invaluable insights into teachers' learning needs, which are crucial for planning meaningful professional development initiatives (Guskey, 2002). However, a large part of the sample of articles also focuses on data gathered from students, so we must address these

findings from both perspectives, even though they are directly interrelated (Harris and Sass, 2011). At the same time, it is necessary to consider that after addressing the first question, the data will focus on the social capital dimension as being key to the improvement of professional capital (Day, 2013). Consequently, these data are highly relevant as they reveal patterns that shed light on the perceived sense of community (Dawson, 2008), participation and social connections (Fournier et al., 2011), disconnected students and teachers, as well as communication between them (Macfadyen and Dawson, 2010).

Based on these findings, and after contextualizing the information, this can be used to identify the potential impact of socialization efforts via active participation in forums, which our study suggests is the main tool for data extraction, a trend supported by numerous articles concerned on improving professional capital, such as those of Poole and van de Ven (1989), Järvelä et al. (2015), and Hammond (2019). The results of these studies highlight the importance of online forums for professional development, especially regarding collaboration and knowledge sharing. Forums can encourage reflection, discussion, and shared learning, which can help teachers improve their practice and raise the quality of teaching overall. In addition, these studies show how forums can be effectively designed and evaluated to support online teacher professional development (Rocco, 2010). Therefore, this information can guide teaching and learning practices by providing an insight into the role played by online discussion forums in both online and blended modes of delivery. However, a limitation of forums, as noted in the study by Kavanaugh and Patterson (2001), is that by relying on forum data, we focus only on the comparatively small proportion of enrolled students who actively participated in them. To summarize, participation in discussion forums provides access to professional capital, which can then be used to access resources and support student outcomes, while having a positive impact on professional development (Misanchuk and Anderson, 2002; Šmite et al., 2017), leading us to encourage the creation of professional capital for quality teaching and learning.

Although there is a notable use of qualitative methods in our analyzed corpus, there are still a few papers that use mixed methods, such as the studies by Peeters and Vaidya (2016) or Saunders (2014), where they show how mixed data collection models can provide a more complete and detailed view of professional capital. By combining quantitative and qualitative methods, objective and subjective data can be collected to provide a deeper understanding of teaching practice and the impact of professional training and development (Carson et al., 2020; Gidari and Kakana, 2022). In addition, these mixed models can help address the limitations of individual data collection methods and improve the validity and reliability of results (Fetters, 2016).

Figure 6 also shows a scarcity of primary usage of quantitative methods. This finding aligns with the research described by Zhang (2022) who pointed out that the data on professional capital should be collected through mixed methods that combine quantitative and qualitative techniques. Quantitative data collection can provide useful information, but it cannot capture the full complexity of teaching practice, as teaching and learning processes are multifaceted and contextual (Ross and Bruce, 2007). Therefore, it is also necessary to use qualitative methods to understand the experience and perspective of teachers in the classroom and in their professional environment.

In addition, the use of a single quantitative method can lead to a fragmented and incomplete view of professional capital, as it does not take into account the context in which learning and teaching occurs. Thus, there is the need for a more holistic approach that integrates various sources of information and perspectives. In general, these studies indicate that to understand professional capital properly, it is necessary to use mixed research techniques that combine quantitative and qualitative methods (Popham, 1999).

However, it is imperative to explore other methods of data collection, such as the direct observation of individuals, as employed in the study by Hu et al. (2018), which captured most of the typical learning experiences of students in the classroom. In addition, techniques that facilitate the management of data from collection to analysis are increasingly being used. These systems include MySQL (Bieke and Maarten, 2012), an open source relational database management system, which in practice can help improve efficiency in the management of information of teaching staff, which allows more time for teaching and less time for administration or monitoring student attendance and academic performance (Paredes and Chung, 2012).

Having identified the relevant aspects of LA in the collection of evidence on professional capital, the next analytical step is to attempt to transform the findings into ordered and meaningful knowledge that allows conclusions to be drawn, for which the results shown in Figure 6 were used.
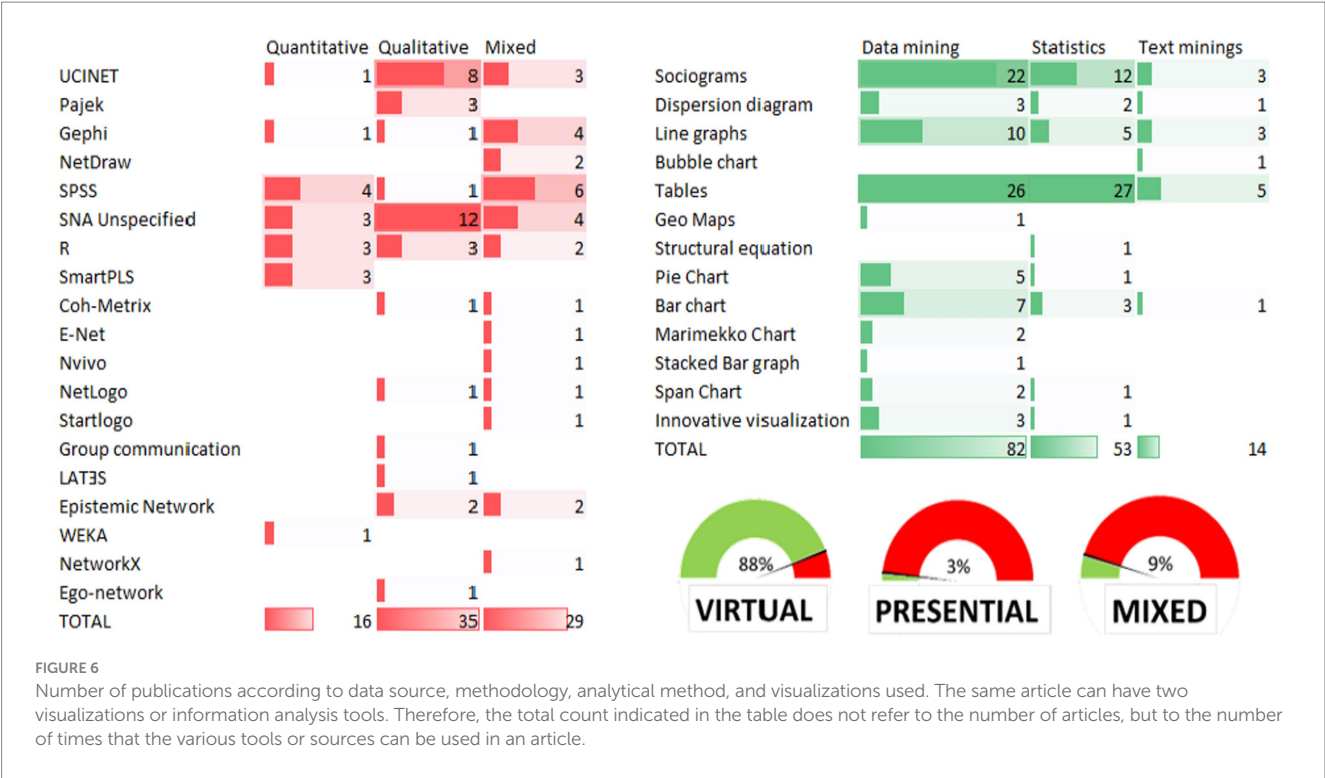
Generally speaking, social learning analysis provides a methodological lens for making sense of these massive data streams (Ahn, 2013). Detailed inspection of the results confirms that within professional capital research, the most used category of analysis within learning analytics is network analysis. This allows detecting patterns of behavior and learning communities (Hernández and Navarro, 2018), while identifying the most influential students and teachers within a group to design pedagogical strategies tailored to the specific needs of each of them.

First, a speedometer at the bottom-right part of the figure shows the origin of the type of data to be analyzed (virtual, face-to-face, or mixed). This figure shows that in 88% of the cases, data were extracted from a virtual sample, 9% were mixed (a mixture of physical and virtual methods), and only 3% of the studies collected data entirely in person.

Next, two tables are presented. The table with red bars compares the analytical tools used (vertical bar) in the studies with the type of methodology used (horizontal bar), while the table with green bars shows the visualization used (vertical bar) to show the data after the various analyses were performed (horizontal bar).

The red table indicates that the most frequently used methodology is qualitative ($n=35$), followed by mixed ($n=29$), and quantitative ($n=16$). Following this rationale, for purely qualitative methodologies, the most widely used instrument is UCINET ($n=8$), followed by R and Pajek ($n=3$ for both), while in 12 occasions the analysis tool was unspecified, which is a limitation of our study. For hybrid methodologies, there is a notable use of SPSS ($n=6$), followed by Gephi ($n=4$), and Ucinet ($n=3$). Finally, for the quantitative methodologies, homogeneous results appear, with the use of tools such as SPSS ($n=4$), R (x3), and SmartPLS ($n=3$).

This table also shows that l data mining is the most frequently used analytical technique ($n=82$), followed by statistical techniques ($n=53$) and, in a limited number of cases, text mining ($n=14$). In the same order, the use of tables as a visualization method is predominant within data mining ($n=26$), followed by sociograms ($n=22$), line

FIGURE 6
Number of publications according to data source, methodology, analytical method, and visualizations used. The same article can have two visualizations or information analysis tools. Therefore, the total count indicated in the table does not refer to the number of articles, but to the number of times that the various tools or sources can be used in an article.

graphs ($n = 10$), bar graphs, and pie charts. Of the studies using statistical procedures, tables are again the most frequently used ($n = 27$), followed by sociograms ($n = 12$) and line graphs ($n = 5$). Finally, a similar distribution was found for text mining studies, with the predominant use of tables ($n = 5$), followed by sociograms ($n = 3$), and line graphs ($n = 3$).

Therefore, as this study demonstrates, network analysis is the most used technique to analyze data related to professional capital because it allows us to understand the complexity of the relationships and connections between teachers and colleagues, students, parents, as well as other stakeholders within the educational system. This approach is evident in the research of Bryant et al. (2017), where social capital is used to facilitate the resolution of complex problems in a large and interdisciplinary team, or Chapman et al. (2016), who conducted a collaborative research and found that schools reported greater evidence of an impact on positive outcomes for disadvantaged students. In addition, the need to use these types of techniques has increased in the wake of the COVID-19 pandemic, as seen in studies such as those of Raaper and Brown (2020).

A key tool within this branch of LA is UCINET (Borgatti et al., 2002), a social network analysis software that is widely used in social and organizational research. This instrument is used to analyze and visualize social networks and provides tools to study the relationships between actors, the flow of information, and the characteristics of the network itself. Therefore, UCINET appears to be a reliable social media tool in the educational field for the analysis of professional capital (Wang et al., 2017). This approach could be particularly useful for those seeking to understand professional capital and how it relates to success at work. For example, this instrument helps educators to reflect on whether their individual social networks are sufficient to allow cooperative actors to be willing to share useful knowledge and/ or whether cooperative culture can maintain and further strengthen

enhanced creativity (Sam Liu, 2017). It is also possible to identify opportunities to expand the network and connect with influencers (Tahmasebi and Askaribezayeh, 2021).

However, in recent years, scientific articles have provided evidence to support the validity of using hybrid methodologies to conduct analyses in professional capital (Leong and Ibrahim, 2015). Nonetheless, UCINET reappears in the column of hybrid procedures, which in turn suggests that this instrument is one of the most reliable in this field. Indeed, Alwafi (2021) emphasizes the importance of using mixed methods to study professional capital on Twitter.

Nevertheless, SPSS and R have also become more present in mixed methods (Muenchen, 2011), although they feature most centrally in quantitative analysis and are accompanied by other programs when the qualitative component comes into play. However, in recent years, R has become more present in qualitative works, as shown in studies such as Oliveira e Sá and de Castro (2020), where R is used qualitatively to analyze leadership, collaboration, and reflexivity from a dialogic and innovative perspective with the aim of improving professional development. Jan et al. (2019) states that while network analysis is effective in detecting key participants, subgroups, and certain aspects of a community of practice, a specific measure of network analysis cannot be correlated with a particular presence in a research community. Therefore, network analysis should be complemented by a qualitative analytical technique.

It is also unsurprising that the combination of *data mining* and statistics frequently appears in our sample of studies, since they allow for analyzing large datasets and extracting significant patterns and relationships, which can be very useful in decision-making and the development of strategies for promoting professional capital. For example, Xing and Gao (2018) discuss the relationship between online discourse and engagement in Twitter's professional learning communities, and by analyzing the most relevant topics in

conversations, these authors gained an insight in the perception that users have about a person and opportunities for collaborations, purely through the use of data mining techniques and statistics.

The visualizations of these data assume a relevant role within LA (Vieira et al., 2018), and in this regard, sociograms emerge as one of the most used strategies for facilitating the presentation of data to interested parties (Tubaro et al., 2016) and have a long tradition in the field (Keithlucas, 1957). Because sociograms provide a clear representation of the relationships and connections between teachers in a network while establishing the roles and positions of teachers therein (Ferreira et al., 2018), they help identify opportunities for collaboration and professional development among teachers (Tatum et al., 2013). Moreover, they facilitate the identification of gaps in professional capital and possible solutions (Loomis, 1948), along with the monitoring and evaluation of professional capital over time (Atkinson, 1949). Network analysis also facilitates the identification of the most effective strategies for improving the quality of teaching and the evaluation of the impact of professional development policies and programs on students and teachers (Zhao, 2009). This concept is central to the content presented in Figure 7.

To conclude addressing this research question, Figure 7 shows the programs and types of platforms used for improving professional capital, as well as the evolution of the latter variable over time.
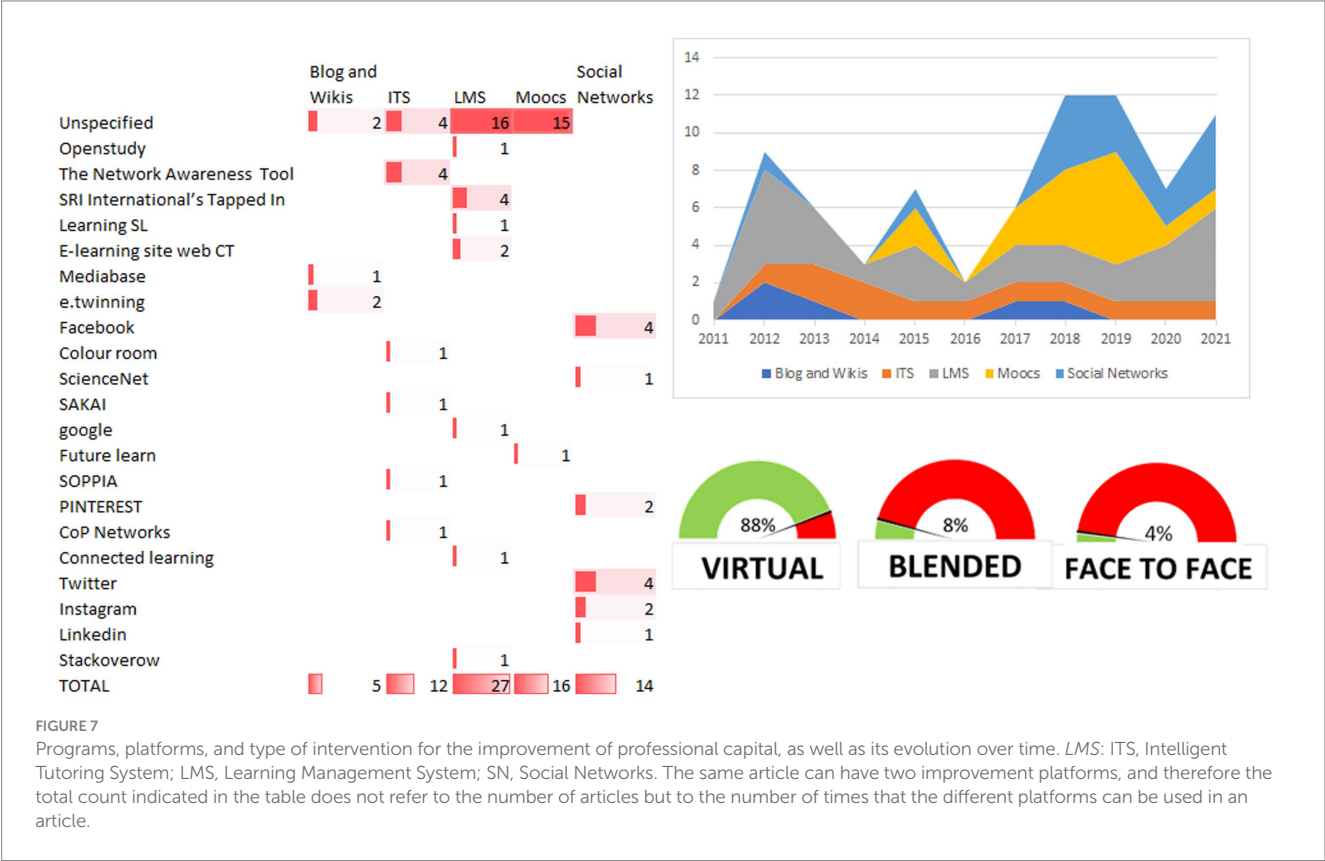
On the bottom right of the figure, a speedometer shows the intervention modality (virtual, blended, or face-to-face). In 88% of cases the intervention is implemented virtually, 8% used a blended intervention, while only 4% used a face-to-face modality.

Moving to the table with the red bars, the columns show the type of platform used for the intervention, while the s rows indicate the

specific name of the service used for the intervention. Regarding the type of platform, the use of LMS and MOOCS ($n = 27$ and $n = 16$, respectively) are notable, followed by social networks ($n = 14$), ITS ($n = 12$) and finally blogs and wikis ($n = 5$). The most prominent LMS of the rest is SRI International's Tapped In ($n = 4$). MOOCs have unspecified names, while the most used social networks are Facebook and Twitter (both n = 4). Within ITS, we can highlight The Network Awareness Tool ($n = 4$) and finally e.twinning (x2) within blogs and wikis.

On the temporal exploration of the included articles, from 2011 to 2013, there was a predominant use of LMS (without the presence of MOOCs and social networks), peaking in 2012 with a total of twelve appearances. The period of 2014–2017 saw the lowest volume of scientific production, peaking in 2015 with a total of 7 cases in which these services were used for improving professional capital. It should also be noted that the use of MOOCs appears, but to a much lesser extent. Finally, 2017 to 2021 represents the most significant period, where the field appears to be emerging, reaching a peak number of scientific publications to indicate the important role played by MOOCs and the use of social networks.

First, and according to the results, LMSs emerge as the most prevalent and consistently employed platforms. This preference is due to their robust features, including organized learning management, planning, evaluation tools, communication, and collaborative capabilities. These attributes make LMSs an ideal form of intervention in the quest to promote professional capital. These ideas are supported by studies such as that of de Laat and Schreurs (2013), where LMSs are used to implement an approach to professional development that is connected to the informal day-to-day networking activities in the

FIGURE 7
Programs, platforms, and type of intervention for the improvement of professional capital, as well as its evolution over time. *LMS*: ITS, Intelligent Tutoring System; LMS, Learning Management System; SN, Social Networks. The same article can have two improvement platforms, and therefore the total count indicated in the table does not refer to the number of articles but to the number of times that the different platforms can be used in an article.

workplace, while providing instructions to develop automated and scalable LA tools that facilitate the establishment of informal networks to better leverage their learning potential.

However, it should also be noted that LMSs have a considerable impact when used with social network integration. Some studies show that although students like to use Facebook as an LMS, many of them find as many advantages as disadvantages when compared to a traditional LMS such as Moodle (Al-Dhanhani et al., 2015). Facebook is clearly preferred by students to communicate instantly with their teachers and participate in discussions, but not for sharing materials and submitting assignments (Verdu et al., 2021).

Regarding the intervention platforms used to improve professional capital, the emergence of MOOCs has become evident, and in recent years these have formed a central axis for collaborative work between students and teachers, as shown by the considerable increase in scientific production in this specific area. This notion is supported by the literature review of Zhu et al. (2022), who analyzed 166 articles between 2011 and 2021 to summarize the trends and critical problems of integrating LA in MOOCs, revealing that this approach was more often used for research purposes than in practice (i.e., learning and teaching). In addition, approximately 60% of the articles adopted student registration data, which also indicates a trend toward the gap found in our study, where most of the articles have focused on students rather than teachers.

Consequently, some of the benefits brought to the field include the ease of keeping up with the latest advances in their field and improving their teaching practice, and the greater flexibility offered in terms of schedules and pace of learning, allowing educators to take online courses and training in their own time and place. Moreover, educators are given the opportunity to become familiar with the latest online teaching technologies and tools (Rincón-Flores et al., 2019) and have the chance to interact and collaborate with other educators around the world, which can provide new perspectives and enrich teaching practices (Paton et al., 2018).

In this regard, studies such as that of Solórzano-García and Navío-Marco (2019) show how Moocs generate knowledge through the increase of professional capital, which has an impact on the improvement of learning communities through the recognition of other members. In addition, it is worth highlighting the algorithms inserted within MOOCs, such as the Latent Dirichlet Allocation (LDA), which is used to recognize latent knowledge (Zarra et al., 2018). Therefore, it is clear that the integration of social networks, algorithms, or blogs within the same platform seems to be one of the most effective interventions, while the isolated application of a single service on the same platform generates difficulties and gaps in learning (Koedinger et al., 2015).

As for social media, the surge in its use over the same time period appears to be undeniable, and this emerging field reveals information about how teachers gain experience, the effectiveness of leadership structures, how councils, professional, and personal networks support teaching, and why and how school reforms spread across districts (Leana, 2011; de Laat and Schreurs, 2013; Spillane et al., 2015; Yoon et al., 2018). This critical mass of research on the potential of social capital and social networks to impact educational contexts has encouraged scholars and school leaders to move toward research and intervention design informed by the findings.

However, there are few studies of these interventions, particularly with regard to how they have been implemented, what mechanisms

promote change, and, most importantly, how the interventions have improved teaching and learning. We need to have clear models and strategies for implementing those interventions, and in the following studies, we show some specific examples of the use of social networks today.

The study by Subekti et al. (2019) assessed the compatibility of collaboration between teachers through interactions on their social networks using the Weight Decision Matrix (WDM) algorithm. Lu et al. (2020) used social networks to identify hidden patterns in the network, finding isolated networks and attempted to unite them; while Oktavia and Sujarwo (2020) used information gathered from the social media platform to generate recommendation systems from learning partners that can provide suggestions for educational institutions.

Follow-up studies of the groups examined should also be conducted to investigate how social capital dynamics evolve over time and how they contribute to the creation of group identity (Ranieri et al., 2012).

Given the results of this research, it is logical to focus on the improvement of social capital as a key to the creation of professional capital, since it provides opportunities for the generation of new knowledge, such as possible solutions to persistent problems of student and faculty success, and research on organizational routines as drivers of change and preservation in organizations. To achieve these goals, it is necessary to identify the most influential variables in social capital as a mechanism to design the most effective interventions adapted to each context and needs, as well as the learning analytics studies involved, all of which will be the focus of our next and final research question.
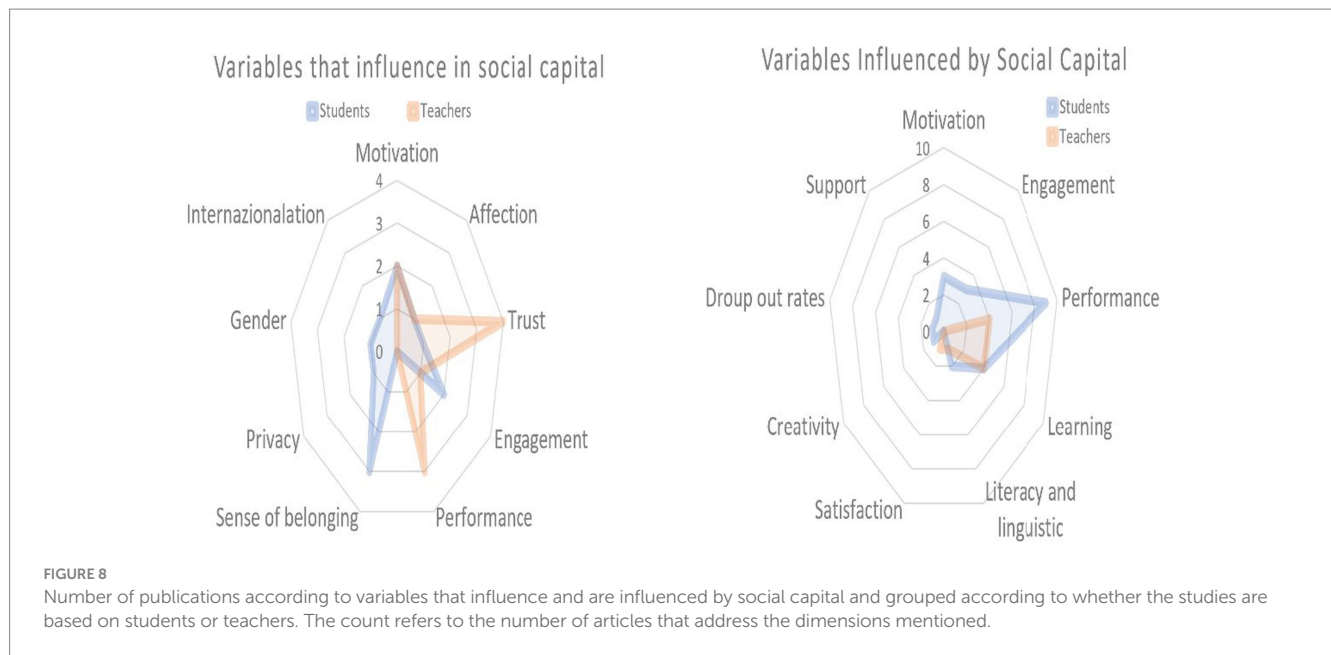
4.3 What is the most untangled dimension of learning analytics in professional capital, and how learning analytics is improving this dimension? (Q3)

Our findings from the first research questions shows that social capital is the largest area untangled by the field of learning analytics (see Figure 4). In this section, we will further explore the variables through which learning analytics is enhancing social capital within the professional capital context.

To answer this, it is necessary to analyze the articles that address social capital in a specific way, to observe which variables are of influence and are influenced. To facilitate this task, a spider diagram was created (see Figure 8 and accompanying notes to understand the numbers) to help clarify the most significant variables involved in social capital. The right-hand panel shows the variables that influence social capital, while the left panel shows the variables influenced by social capital, both divided according to whether these refer to students or teachers.

Beginning with the variables that influence social capital with respect to the student, the feeling of belonging to the community appears as the most influential (with a total of 3 articles), followed by commitment and motivation ($n = 2$ each). Concerning teachers, the trust dimension is the most predominant object of study with a total of 4 articles, followed by performance ($n = 3$), motivation ($n = 2$), and finally commitment, with a single article.

Number of publications according to variables that influence and are influenced by social capital and grouped according to whether the studies are based on students or teachers. The count refers to the number of articles that address the dimensions mentioned.

On the other hand, the variables influenced by social capital according to studies conducted in students include various aspects of performance with a total of 9 articles, followed by learning ($n = 4$), and motivation and commitment ($n = 3$ for each). Regarding the teaching staff, there is a lack of scientific production compared to that related to students, since only 4 works are concerned with learning and performance as dimensions influenced by social capital.

In this regard, this study supports works such as that of Moolenaar et al. (2012), in which social capital appears as a variable that can increase professional capital. However, we need to take into account the variables that support social capital in this process, such as the sense of belonging among students, because people who do not feel part of a group may feel isolated, disconnected, and less committed, while those who feel that they belong to a group are more likely to feel engaged and connected to other members of the group. In our work, this commitment is also shown to be an influential variable in social capital.

As explained, Ahn and Davis (2020) provide solid evidence to show how the sense of belonging and social capital are theoretically and empirically intertwined, and similarly, Cheung (2011) states that people who feel part of a group may also be more willing to collaborate in community projects, contribute ideas and resources, as well as help the other members of the group. Finally, results such as those reported by Glass and Gesing (2018) agree with our study regarding the importance of a sense of belonging and commitment as decisional variables in social capital.

This work on social capital in students has reported improvements in student performance, as supported by the data of our review, along with studies such as those of Salimi et al. (2022) demonstrating that social and personal integrative benefits play a mediating role in the relationship between online social capital and academic performance.

From the perspective of teachers, trust between colleagues emerges as a key aspect requiring attention, since this lays the foundations for increasing social capital. Indeed, when trust is high, people take a step forward voluntarily and work together in an efficient and optimal manner. As Allan and Persson (2020) state, they embrace a common purpose, take risks, think creatively, help each other, and communicate openly and sincerely.

Based on the works just described, there appears to be a difference between students and teachers in terms of the volume of work conducted on social capital, with a greater amount of work dedicated to students. Therefore, the need arises for the teachers themselves to collaboratively work in the classroom and as agents involved in the community.

It appears to us that trust, motivation, and social connections are intertwined and, consequently, require greater consideration. For example, Nahapiet and Ghoshal (1998) analyzed the social concept of 'intellectual capital 'by which they refer to the knowledge of a social community, such as an organization or professional practice groups. Thus, a key question is how can we create the motivation and confidence to sustain a spiral of construction of intellectual and social capital in networked practices?

This response must be based on the formation of Communities of Professional Practice with a shared purpose. In fact, numerous studies in our sample directly relate social capital to communities (Solórzano-García and Navío-Marco, 2019; Alwafi, 2021). It can be noted, then, that the concept of social capital has provided researchers and education practitioners with valuable insights into community building (Bryk and Schneider, 2002; Penuel et al., 2013). Communities of professional practice create social capital by providing opportunities for professional interaction and dialog to improve teaching practices. A leader can develop the collective capacity of a staff to achieve better results by ensuring that teams use collaborative time to engage in dialog and processes that positively impact student learning.

In short, the variables analyzed by LA that support social capital in this study will give us guidance on how to work effectively to improve professional capital, all of which involves creating communities that promote a shared identity within a network of people as well as the collective development of a particular domain or theme.

Methodologically, our study offers valuable insights by scrutinizing the diverse data collection strategies employed in learning analytics for professional capital improvement. The emphasis on forums and tools provides not only a comprehensive overview but also serves as a practical guide for educators seeking effective ways to harness the potential of data.

Interdisciplinary collaboration takes center stage in our discussion, highlighting its pivotal role in elevating the rigor of studies. We advocate for mixed methods analysis, emphasizing the importance of collaborative efforts to enrich the exploration of learning analytics and professional capital.

Furthermore, we recommend to explore MOOCs and learning analytics interventions, to incorporate automated community discovery techniques, and to adapt to the challenges posed by the COVID-19 pandemic, as points to serve as guideposts for scholars seeking to advance the field (learning analytics in professional capital) and address emerging issues. In essence, this study not only sheds light on the current state of learning analytics and professional capital but also lays the groundwork for a future research agenda that is both informed and innovative.

5 Conclusion

This study provides a systematic review of the intersection of learning analytics and professional capital. Three databases were searched and prompted the following conclusions:

5.1 On characteristics and approaches of learning analytics within professional capital

LA for Professional Capital Improvement is an effective line of work. LA helps to understand how to identify and establish indicators and lines of action, which can lead to collaboration with schools, their management teams, and the local community, as well as with education authorities, networks and professional associations. Our analysis revealed that collaborative learning not only takes place in formal contexts, but also occurs informally (Shepard et al., 2018), and even takes place in networks.

Increasing professional capital requires understanding how educational institutions are shaped, articulated, and productively optimized as a community for educational success. LA provides an overview of the stock of social and professional capital used by communities. In the context of the pandemic, professional capital has taken on even more important because of significant changes in the way people work and the skills and knowledge needed to adapt to these changes.

Social capital serves as the foundation for the other two types of professional capital (human and decisional capital). This form of capital plays a pivotal role in creating opportunities for the acquisition of new knowledge. While it is clear that conceptual studies have predominantly focused on improving the social capital of students, there is an evident gap in our understanding of social capital among teachers, particularly within primary education. Addressing this gap is essential for cultivating productive learning environments (Kovanovic et al., 2014).

The field analyzed occupies very predominant positions in high impact databases. This indicates a tendency to plan future research in different areas and themes. Most of the authors working in this line of research are Spanish.

5.2 Learning analytics uses the following data collection, analysis, and improvement strategies to increase professional capital

The forum is the main tool for extracting data on professional capital. The use of the UCINET tool appears to be significant and allows the identification of the training and professional development needs of teachers (Zhao, 2008), as well as the consequent design of strategies to improve the quality of education through data mining techniques, all with the help of visualizations such as sociograms.

The LMS appears to be the most predominant and consistently used platform over time, although it should be borne in mind that MOOCs and social networks have become a central axis for collaborative work between students and teachers in recent years. These environments offer flexibility and a variety of themes, fostering interactivity while boasting a large data storage capacity. The key methodology for this particular field seems to involve knowledge of techniques, tools, and platforms for social network analysis. However, few studies have focused on informing the design and implementation of intentional frameworks for creating social capital of teachers through the development of social networks (Yoon et al., 2018).

In general, the outcome of interventions through the use of platforms and web services is a useful means of making recommendations to teachers. These make it possible to suggest projects and contacts, and help build communities as well as select their members, thus maximizing social capital (Pham et al., 2012)

5.3 Social capital stands out as the most studied dimension within the intersection of learning analytics and professional capital

Students who feel they belong to a group are more likely to feel engaged and connected to other members of the group. This translates into improved performance and confidence among co-workers. This is therefore a key aspect to develop when aiming to increase social capital.

In general, this research has provided new insights into the current state of how a specific field — such as Learning Analytics — can help to understand how the processes of building professional capital in education are collected, analyzed, and improved. All these findings point to a number of guidelines and implications for the quality of professional capital training:

There is a need for teachers themselves (and other relevant stakeholders) to undertake collaborative classroom work. These efforts must be based on the formation of communities of practice united by a shared purpose, aligned with the goals set out in the 2030 agenda. The utilization of LA helps to restore equity in learning, which suffered significant setbacks during the pandemic period.

The analysis for the improvement of professional capital should be carried out using mixed methods. These analyses should stimulate

awareness, develop networking skills, and provide information on learning outcomes in learning networks. Professional development within an organization — specifically primary education — is severely affected by a lack of scientific output. This gap suggests another promising avenue of research, that is, to examine the current themes that are giving rise to professional collaborations.

Finally, it is suggested that future studies be undertaken on MOOC and LA interventions to improve learning and teaching practices. Active interdisciplinary collaboration increases the rigor of studies and the dissemination of knowledge. In addition, we stress the need to include automated community discovery techniques in e-learning environments to facilitate and enhance their use. We also emphasize the urgency of conducting further advanced research to uncover other hidden opportunities (Yassine et al., 2022).

In short, identifying the strategies involving learning analytics within the context of professional capital is presented as a powerful approach toward developing an understanding of knowledge about networks. This, in turn, has a positive impact on the overall quality of education. Consequently, this theme represents a novel focal point for delving into the underlying reasons and objectives behind the improvement of social capital.

Author contributions

JH-R: Writing – original draft, Writing – review & editing. MK: Writing – original draft, Writing – review & editing. JD: Writing – original draft, Writing – review & editing. QL: Writing – original draft, Writing – review & editing.

Funding

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.2024.1302658/ full#supplementary-material

References

Ahn, J. (2013). What can we Learn from Facebook activity? Using social learning analytics to observe new media literacy skills. Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium.

Ahn, M. Y., and Davis, H. H. (2020). Sense of belonging as an indicator of social capital. Int. J. Sociol. Soc. Policy 40, 627–642. doi: 10.1108/IJSSP-12-2019-0258

Al-Dhanhani, A., Mizouni, R., Otrok, H., and Al-Rubaie, A. (2015). Analysis of collaborative learning in social network sites used in education. Soc. Netw. Anal. Min. 5:303. doi: 10.1007/s13278-015-0303-z

Alexander, P. A. (2020). Methodological guidance paper: the art and science of quality systematic reviews. Rev. Educ. Res. 90, 6–23. doi: 10.3102/0034654319854352

Allan, J., and Persson, E. (2020). Social capital and trust for inclusion in school and society. Educ. Citizensh. Soc. Justice 15, 151–161. doi: 10.1177/1746197918801001

Alwafi, E. (2021). Tracing changes in teachers' professional learning network on twitter: comparison of teachers' social network structure and content of interaction before and during the COVID-19 pandemic. J. Comput. Assist. Learn. 37, 1653–1665. doi: 10.1111/jcal.12607

Atkinson, G. (1949). The Sociogram as an instrument in social-studies teaching and evaluation. Elem. Sch. J. 50, 74–85. doi: 10.1086/459108

Baumann, T., Harfst, S., Swanger, A., Saganski, G., Alwerfalli, D., and Cell, A. (2014). Developing competency-based, industry-driven manufacturing education in the USA: bringing together industry, government and education sectors. Procedia Soc. Behav. Sci. 119, 30–39. doi: 10.1016/j.sbspro.2014.03.006

Becker, G. (2010). Human capital: a theoretical and empirical analysis, with special reference to education.

Bieke, S., and Maarten, D. L. (2012). Network awareness tool—learning analytics in the workplace: detecting and analyzing informal workplace learning. Proceedings of the 2nd international conference on learning analytics and knowledge, New York, NY. 59–64.

Bolívar, A., and Domingo, J. (2023). Comunidades de práctica profesional y mejora de los aprendizajes. Análisis y estudios 74

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002). Ucinet for windows: Software for social network analysis. Harvard, MA: Analytic Technologies.

Bryant, L. H., Freeman, S. B., Daly, A., Liou, Y.-H., and Branon, S. (2017). Making sense: unleashing social capital in interdisciplinary teams. J. Profes. Capital Commun. 2, 118–133. doi: 10.1108/JPCC-01-2017-0001

Bryk, A. S., and Schneider, B. (2002). Trust in Schools: a core resource for improvement. Manhattan, NY: Russell Sage Foundation.

Carceller, C., Dawson, S., and Lockyer, L. (2013). Improving academic outcomes: does participating in online discussion forums payoff? Int. J. Technol. Enhanced Learn. 5, 117–132. doi: 10.1504/IJTEL.2013.059087

Carson, R. L., Kuhn, A. P., Moore, J. B., Castelli, D. M., Beighle, A., Hodgin, K. L., et al. (2020). Implementation evaluation of a professional development program for comprehensive school physical activity leaders. Prev. Med. Rep. 19:101109. doi: 10.1016/j. pmedr.2020.101109

Chapman, C., Chestnutt, H., Friel, N., Hall, S., and Lowden, K. (2016). Professional capital and collaborative inquiry networks for educational equity and improvement? J. Prof. Cap. Community 1, 178–197. doi: 10.1108/JPCC-03-2016-0007

Cheung, C. (2011). Children's sense of belonging and parental social capital derived from school. J. Genet. Psychol. 172, 199–208. doi: 10.1080/00221325.2010.520362

Coleman, J. S. (1988). Social Capital in the Creation of human capital. *Am. J. Sociol.* 94, S95–S120. doi: 10.1086/228943

Cook, J., Schmidt, A., Kunzmann, C., and Braun, S. (2012). The challenge of integrating motivational and affective aspects into the design of networks of practice. CEUR Workshop Proceedings, 957. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84924874062&partnerID=40&md5=1c0f3aa53642d91aa8b58bc4c1cbe62e

Cross, R. L., Parker, A., and Cross, R. (2004). *The hidden power of social networks: Understanding how work really gets done in organizations.*

Daly, A. J., Moolenaar, N. M., Bolivar, J. M., and Burke, P. (2010). Relationships in reform: the role of teachers' social networks. *J. Educ. Adm.* 48, 359–391. doi: 10.1108/09578231011041062

Darling-Hammond, L. (2022). Reimagining American education: possible futures: the policy changes we need to get there. *Phi Delta Kappan* 103, 54–57. doi: 10.1177/00317217221100012

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Dev. Psychopathol.* 20, 775–803. doi: 10.1017/S0954579408000370

Day, C. (2013). "The new lives of teachers" in *Back to the future: Legacies, continuities and changes in educational policy, practice, and research*. eds. M. A. Flores, A. A. Carvalho, F. I. Ferreira and M. T. Vilaça (Dordrecht: Sense Publishers), 57–74.

Day, C., Kington, A., Stobart, G., and Sammons, P. (2006). The personal and professional selves of teachers: stable and unstable identities. *Br. Educ. Res. J.* 32, 601–616. doi: 10.1080/01411920600775316

de Laat, M., and Schreurs, B. (2013). Visualizing informal professional development networks: building a case for learning analytics in the workplace. *Am. Behav. Sci.* 57, 1421–1438. doi: 10.1177/0002764213479364

Demir, E. K. (2021). The role of social capital for teacher professional learning and student achievement: a systematic literature review. *Educ. Res. Rev.* 33:100391. doi: 10.1016/j.edurev.2021.100391

DuFour, R. (2003). Building a professional learning community. *School Administrator* 60, 13–18.

Ferguson, R., and Buckingham, S. (2012). Social learning analytics: five approaches. Proceedings of the 2nd international conference on learning analytics and knowledge, New York, NY

Ferreira, A. L., Brasil, T. L., and Acioly-Régnier, N. M. (2018). O sociograma e os processos grupais: Uma experiência no campo educacional. *Comunicações* 25:137. doi: 10.15600/2238-121X/comunicacoes.v25n2p137-166

Fetters, M. D. (2016). "Haven't we always been doing mixed methods research?": lessons learned from the development of the horseless carriage. *J. Mixed Methods Res.* 10, 3–11. doi: 10.1177/1558689815620883

Fournier, T., Bruckert, E., Czernichow, S., Paulmyer, A., and Poulain, J. P. (2011). The THEMA study: a sociodemographic survey of hypercholesterolaemic individuals. *J. Hum. Nutr. Diet.* 24, 572–581. doi: 10.1111/j.1365-277X.2011.01168.x

Gidari, S., and Kakana, D. (2022). *Using a mixed-method to evaluate a kindergarten teachers' Professional development Programme and to investigate teachers' Professional Growth*

Giddens, A. (1999). *Consecuencias de la modernidad.*

Glass, C. R., and Gesing, P. (2018). The development of social capital through international students' involvement in campus organizations. *J. Int. Stud.* 8, 1274–1292. doi: 10.5281/zenodo.1254580

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strateg. Manag. J.* 17, 109–122. doi: 10.1002/smj.4250171110

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching* 8, 381–391. doi: 10.1080/135406002100000512

Hammond, M. (2019). A review of recent papers on online discussion in teaching and learning in higher education. *Online Learn.* 9:1782. doi: 10.24059/olj.v9i3.1782

Hargreaves, A. (2020). "The Day after: education and equity after the global pandemic" in *Flip the system US: How teachers can transform education and save democracy*. ed. M. Soskil (Boca Raton, FL: CRC Press), 64–73.

Hargreaves, A., and Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. Ashland: Blackstone Publishing

Hargreaves, A., and Fullan, M. (2013). The power of professional capital. *Learn. Profes.* 34:36.

Hargreaves, A., and Fullan, M. (2020). Professional capital after the pandemic: revisiting and revising classic understandings of teachers' work. *J. Prof. Capital Commun.* 5, 327–336. doi: 10.1108/JPCC-06-2020-0039

Hargreaves, D., Miell, D., and Macdonald, R. (2002). "What are musical identities, and why are they important" in *Musical Identities* (Oxford: Oxford University Press)

Hargreaves, A., and O'Connor, M. T. (2018). *Collaborative professionalism: When teaching together means learning for all. 1st Edn*. Thousand Oaks: Corwin.

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

Haythornthwaite, C. (2011). Learning networks, crowds and communities. Proceedings of the 1st international conference on learning analytics and knowledge-LAK'11, Banff, Alberta.

Hernández, E., and Navarro, M. J. (2018). La participación en redes escolares locales para promover la mejora educativa, un estudio de caso. *Profesorado, Revista de Currículum y Formación del Profesorado* 22, 71–90. doi: 10.30827/profesorado.v22i2.7715

Hu, S., Torphy, K. T., Opperman, A., Jansen, K., and Lo, Y.-J. (2018). What do teachers share within socialized knowledge communities: a case of Pinterest. *J. Prof. Cap. Community* 3, 97–122. doi: 10.1108/JPCC-11-2017-0025

Jan, S. K., Vlachopoulos, P., and Parsell, M. (2019). Social network analysis and online learning communities in higher education: a systematic literature review. *Online Learn.* 23:1. doi: 10.24059/olj.v23i1.1398

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., et al. (2015). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educ. Technol. Res. Dev.* 63, 125–142. doi: 10.1007/s11423-014-9358-1

Kavanaugh, A. L., and Patterson, S. J. (2001). The impact of community computer networks on social capital and community involvement. *Am. Behav. Sci.* 45, 496–509. doi: 10.1177/00027640121957312

Keithlucas, A. (1957). Using the Sociogram in teaching Houseparents. *Child Welfare* 36, 1–7.

Kerrigan, M. R. (2015). Social Capital in Data-Driven Community College Reform. *Community Coll. J. Res. Pract.* 39, 603–618. doi: 10.1080/10668926.2013.866061

Khalil, M., and Ebner, M. (2016). What is learning analytics about? A survey of different methods used in 2013–2015. *arXiv*. doi: 10.48550/arXiv.1606.02878

Khousa, E. A., and Atif, Y. (2018). Social network analysis to influence career development. *J. Ambient. Intell. Humaniz. Comput.* 9, 601–616. doi: 10.1007/s12652-017-0457-9

Kimble, C., and Hildreth, P. (2008). *Communities of practice-Vol. 2: Creating learning environments for educators*. Charlotte, NC: Information Age Publishing.

Kitchenham, B., and Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering. Technical report, EBSE technical report EBSE-2007-01.*

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., and Bier, N. L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. Proceedings of the second (2015) ACM conference on learning @ scale, New York, NY.

Kovanovic, V., Joksimovic, S., Gasevic, D., and Hatala, M. (2014). What is the source of social capital? The association between social network position and social presence in communities of inquiry. *CEUR Workshop Proc.* 1183, 21–28.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life.* Cambridge: Cambridge University Press.

Leana, C. R. (2011). The missing link in school reform. *Standford Innov. Rev.* 34, 3–4.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Leong, L. W., and Ibrahim, O. (2015). Role of information system (IS), social networking technology (SNT) and WEB 2.0 for improving learning outcomes: a case of Malaysian universities. *Procedia Soc. Behav. Sci.* 211, 111–118. doi: 10.1016/j.sbspro.2015.11.017

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 6:e1000100. doi: 10.1371/journal.pmed.1000100

Liu, X., Chang, J., and Zhang, L. (2020). Development and validation of a scale for teacher professional capital for ICT-enhanced teaching innovation. 2020 Ninth International Conference of Educational Innovation through Technology (EITT), Porto, Portugal.

Llopis-Albert, C., and Rubio, F. (2021). Application of learning analytics to improve higher education. *Multidiscip. J. Educ. Soc. Technol. Sci.* 8:2. doi: 10.4995/muse.2021.16287

Long, P., and Siemens, G. (2011). What is learning analytics. In Proceedings of the 1st international conference learning analytics and knowledge, LAK, Banff, Alberta.

Loomis, C. P. (1948). The Most frequently chosen Sociogram. *Sociometry* 11, 230–234. doi: 10.2307/2785112

Lu, Q., Lu, Q., Huang, J., Ge, Y., Wen, D., Chen, B., et al. (2020). EgoVis: a visual analysis system for social networks based on egocentric research. *Int. J. Cooperat. Informat. Syst.* 29, 1930003–1930002. doi: 10.1142/S0218843019300031

Macfadyen, L. P., and Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: a proof of concept. *Comput. Educ.* 54, 588–599. doi: 10.1016/j.compedu.2009.09.008

McKenzie, J., van Winkelen, C., and Grewal, S. (2011). Developing organisational decision-making capability: a knowledge manager's guide. *J. Knowl. Manag.* 15, 403–421. doi: 10.1108/13673271111137402

Michos, K., Lang, C., Hernández-Leo, D., and Price-Dennis, D. (2020). Involving teachers in learning analytics design: lessons learned from two case studies. Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. New York, NY.

Minga-Vallejo, R.-E., Ramírez-Montoya, M.-S., and Rodríguez-Conde, M.-J. (2021). Methods for the evaluation of social learning (2017-2021): systematic literature review. Ninth international conference on technological ecosystems for enhancing Multiculturality (TEEM'21), Barcelona, Spain.

Misanchuk, M., and Anderson, T. (2002). Building Community in an Online Learning Environment: Communication, cooperation and collaboration. Available at: https://www.semanticscholar.org/paper/Building-Community-in-an-Online-Learning-and-Misanchuk-Anderson/cd6eaa59b4bc4688352dcf00858cadbea1588577

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 339:b2535. doi: 10.1136/bmj.b2535

Monés, A. M., Damoulis, Y. D., Acquila-Natale, E., Álvarez, A., Rodríguez, M. C., Pérez, R. C., et al. (2020). Achievements and challenges in learning analytics in Spain: the view of SNOLA. *RIED* 23:2. doi: 10.5944/ried.23.2.26541

Moolenaar, N. M., Sleegers, P. J. C., and Daly, A. J. (2012). Teaming up: linking collaboration networks, collective efficacy, and student achievement. *Teach. Teach. Educ.* 28, 251–262. doi: 10.1016/j.tate.2011.10.001

Muenchen, R. A. (2011). *R for SAS and SPSS users second edition conclusion. En R for SAS and Spss users*, 2nd Edn (pp. 647–661). Berlin: Springer.

Muñoz-Merino, P. J., Moreno-Marcos, P. M., Rubio-Fernández, A., Tsai, Y.-S., Gašević, D., and Delgado Kloos, C. (2022). A systematic analysis of learning analytics using multi-source data in the context of Spain. *Behav. Inform. Technol.* 42, 643–657. doi: 10.1080/0144929X.2022.2039767

Nahapiet, J., and Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Acad. Manag. Rev.* 23, 242–266. doi: 10.2307/259373

Newmann, F., King, M. B., and Rigdon, M. (1997). Accountability and school performance: implications from restructuring schools. *Harv. Educ. Rev.* 67, 41–75. doi: 10.17763/haer.67.1.14141916116656q6

Nonaka, I., and Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford: Oxford University Press.

Oktavia, T., and Sujarwo, S. (2020). Exploration of recommender generator system to support social learning platform of higher education institution. *ICIC Exp. Lett.* 14, 489–496. doi: 10.24507/icicel.14.05.489

Oliveira e Sá, S., and de Castro, P. A. (2020). "Characteristics of the pedagogical supervisor in context of a constructive and reflective supervision" in *Computer supported qualitative research*. eds. A. P. Costa, L. P. Reis and A. Moreira (New York: Springer International Publishing), 274–287.

Organization for Economic Co-operation and Development. (2019). Education at a glance. Available at: https://www.oecd-ilibrary.org/education/education-at-a-glance-2019_f8d7880d-en

Paraschiv, I. C., Dascalu, M., McNamara, D. S., and Trausan-Matu, S. (2016). "Finding the needle in a haystack: who are the most central authors within a domain?" in *Adaptive and adaptable learning*. eds. K. Verbert, M. Sharples and T. Klobučar (New York: Springer International Publishing), 632–635.

Paredes, W. C., and Chung, K. S. K. (2012). Modelling learning & performance: a social networks perspective. Proceedings of the 2nd international conference on learning analytics and knowledge, Vancouver, British Columbia.

Paton, R. M., Fluck, A. E., and Scanlan, J. D. (2018). Engagement and retention in VET MOOCs and online courses: a systematic review of literature from 2013 to 2017. *Comput. Educ.* 125, 191–201. doi: 10.1016/j.compedu.2018.06.013

Pedroso, J. E. P. Jr., Siason, N. D., and Tangco-Siason, A. (2021). Principal's leadership practices during the COVID 19 pandemic: an exploratory study. *Int. J. Arts Human. Stud.* 1, 76–87. doi: 10.32996/ijahs.2021.1.1.12

Peeters, M. J., and Vaidya, V. A. (2016). A mixed-methods analysis in assessing students' professional development by applying an assessment for learning approach. *Am. J. Pharm. Educ.* 80:77. doi: 10.5688/ajpe80577

Penuel, W. R., Frank, K. A., Sun, M., Kim, C. M., and Singleton, C. A. (2013). The organization as a filter of institutional diffusion. *Teach. Coll. Rec.* 115, 1–33. doi: 10.1177/016146811311500105

Pham, M. C., Derntl, M., Cao, Y., and Klamma, R. (2012). Learning analytics for learning blogospheres. Proceedings of the 11th international conference on Advances in Web-Based Learning, Berlin, Heidelberg.

Pollock, A., and Berge, E. (2018). How to do a systematic review. *Int. J. Stroke* 13, 138–156. doi: 10.1177/1747493017743796

Poole, M. S., and van de Ven, A. H. (1989). Using paradox to build management and organization theories. *Acad. Manag. Rev.* 14, 562–578. doi: 10.2307/258559

Popham, W. J. (1999). *Modern educational measurement: Practical guidelines for educational leaders 3rd Edn*. London: Pearson.

Raaper, R., and Brown, C. (2020). The Covid-19 pandemic and the dissolution of the university campus: implications for student support practice. *J. Prof. Cap. Community* 5, 343–349. doi: 10.1108/JPCC-06-2020-0032

Ranieri, M., Manca, S., and Fini, A. (2012). Why (and how) do teachers engage in social networks? An exploratory study of professional use of Facebook and its implications for lifelong learning. *Br. J. Educ. Technol.* 43, 754–769. doi: 10.1111/j.1467-8535.2012.01356.x

Rincón, S. (2019). "Las redes escolares como entornos de aprendizaje para los líderes educativos" in *Cómo cultivar el liderazgo educativo. Trece miradas*. eds. J. Weinstein and G. Muñoz (Santiago: Universidad Diego Portales), 355–388.

Rincón-Flores, E. G., Montoya, M. S. R., and Mena, J. (2019). Engaging MOOC through gamification: systematic mapping review. Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality, New York, NY.

Rocco, S. (2010). Making reflection public: using interactive online discussion board to enhance student learning. *Reflective Pract.* 11, 307–317. doi: 10.1080/14623943.2010.487374

Ross, J., and Bruce, C. (2007). Professional development effects on teacher efficacy: results of randomized field trial. *J. Educ. Res.* 101, 50–60. doi: 10.3200/JOER.101.1.50-60

Salimi, G., Heidari, E., Mehrvarz, M., and Safavi, A. A. (2022). Impact of online social capital on academic performance: exploring the mediating role of online knowledge sharing. *Educ. Inf. Technol.* 27, 6599–6620. doi: 10.1007/s10639-021-10881-w

Sam Liu, C.-H. (2017). Remodelling progress in tourism and hospitality students' creativity through social capital and transformational leadership. *J. Hospital. Leisure Sport Tourism Educ.* 21, 69–82. doi: 10.1016/j.jhlste.2017.08.003

Saunders, R. (2014). Effectiveness of research-based teacher professional development. *Australian J. Teach. Educ.* 39, 5–6. doi: 10.14221/ajte.2014v39n4.10

Schön, E.-M., Thomaschewski, J., and Escalona, M. J. (2017). Agile requirements engineering: a systematic literature review. *Comput. Stand. Interfaces* 49, 79–91. doi: 10.1016/j.csi.2016.08.011

Schultz, T. W. (1961). Investment in Human Capital. *Am Econ Rev.* 51, 1–17.

Shepard, S., Boudet, H., Zanocco, C. M., Cramer, L. A., and Tilt, B. (2018). Community climate change beliefs, awareness, and actions in the wake of the September 2013 flooding in Boulder County, Colorado. *J. Environ. Stud. Sci.* 8, 312–325. doi: 10.1007/s13412-018-0479-4

Siemens, G. (2013). Learning analytics: the emergence of a discipline. *Am. Behav. Sci.* 57, 1380–1400. doi: 10.1177/0002764213498851

Silva, L., Mendes, A. J., and Gomes, A. (2020). Computer-supported collaborative learning in programming education: a systematic literature review. 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal.

Simon, H. A. (1977). *The new science of management decision*.

Šmite, D., Moe, N. B., Šāblis, A., and Wohlin, C. (2017). Software teams and their knowledge networks in large-scale software development. *Inf. Softw. Technol.* 86, 71–86. doi: 10.1016/j.infsof.2017.01.003

Smylie, M. A., and Evans, A. E. (2006). "Social capital and the problem of implementation" in *New directions in education policy: Confronting complexity* (Albany, NY: State University of New York Press), 187–208.

Solórzano-García, M., and Navío-Marco, J. (2019). Developing social entrepreneurs through distance education: the value of commitment and interactivity with the learning community. *Int. J. Mobile Learn. Organ.* 13, 30–50. doi: 10.1504/IJMLO.2019.096466

Spillane, J. P., Hopkins, M., and Sweet, T. M. (2015). Intra-and interschool interactions about instruction: exploring the conditions for social capital development. *Am. J. Educ.* 122, 71–110. doi: 10.1086/683292

Stoll, B. J., Hansen, N. I., Bell, E. F., Shankaran, S., Laptook, A. R., Walsh, M. C., et al. (2010). Neonatal outcomes of extremely preterm infants from the NICHD neonatal research network. *Pediatrics* 126, 443–456. doi: 10.1542/peds.2009-2959

Subekti, A., Ferdiana, R., and Santosa, P. I. (2019). Social media mapping for business communication. 2019 international conference on information and communications technology, ICOIACT 2019, Yogyakarta, Indonesia.

Tahmasebi, A., and Askaribezayeh, F. (2021). Microfinance and social capital formation-a social network analysis approach. *Socio Econ. Plan. Sci.* 76:100978. doi: 10.1016/j.seps.2020.100978

Tatum, P. E., Bell, C., and Hosokawa, M. (2013). Use of a Sociogram to teach teamwork. *J. Am. Geriatr. Soc.* 61, S153–S154.

Tong, W., and Razniak, A. (2017). Building professional capital within a 21st century learning framework. *J. Prof. Cap. Community* 2, 36–49. doi: 10.1108/JPCC-06-2016-0018

Tubaro, P., Ryan, L., and D'angelo, A. (2016). The visual Sociogram in qualitative and mixed-methods research. *Sociol. Res. Online* 21, 180–197. doi: 10.5153/sro.3864

Verdu, M. J., De Castro, J.-P., Regueras, L. M., and Corell, A. (2021). MSocial: practical integration of social learning analytics into Moodle. *IEEE Access* 9, 23705–23716. doi: 10.1109/ACCESS.2021.3056914

Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: a systematic literature review and research agenda. *Comput. Educ.* 122, 119–135. doi: 10.1016/j.compedu.2018.03.018

Visone, J. D. (2018). Developing social and decisional capital in US National Blue Ribbon Schools. *Improv. Sch.* 21, 158–172. doi: 10.1177/1365480218755171

Wang, Z., Chen, X., and Yu, M. (2017). Interactive effect of leader–member tie and network centrality on leadership effectiveness. *Soc. Behav. Personal. Int. J.* 45, 1197–1210. doi: 10.2224/sbp.6351

Xing, W., and Gao, F. (2018). Exploring the relationship between online discourse and commitment in twitter professional learning communities. *Comput. Educ.* 126, 388–398. doi: 10.1016/j.compedu.2018.08.010

Yassine, S., Kadry, S., and Sicilia, M.-A. (2022). Detecting communities using social network analysis in online learning environments: systematic literature review. *WIREs Data Mining Knowl. Discov.* 12:e1431. doi: 10.1002/widm.1431

Yoon, S. A., Anderson, C. W., Baker-Doyle, K., de Laat, M., de los Santos, E., Frank, K. A., et al. (2018). Networked by design: interventions for teachers to develop social capital. *ICLS* 2, 1235–1242.

Yu, X., Guan, J., and Leng, J. (2016). Using learning analytics to support personalized learning and quality education: a case study of China's "everyone connected" project. En Z. Gong, D. K. W. Chiu and D. Zou (Eds.), *Current developments in web based learning* (pp. 196–201). New York: Springer International Publishing.

Zarra, T., Chiheb, R., Faizi, R., and El Afia, A. (2018). MOOCs' recommendation based on forum latent Dirichlet allocation. Proceedings of the 2nd international conference on smart digital environment, Rabat, Morocco.

Zhang, J. (2022). *Teachers' professional learning communities in China: a mixed-method study on Shanghai primary schools*. London: Routledge.

Zhao, J. (2008). "A sociogram analysis on group interaction in an online discussion forum" in *Advances in web based learning—Icwl 2008, proceedings*. eds. F. Li, J. Zhao, T. K. Shih, R. Lau, Q. Li and D. McLeod, vol. *5145* (Berlin: Springer-Verlag), 377–389.

Zhao, J. (2009). Group interaction in a web 2.0 based learning environment: a sociogram analysis. *Int. J. Contin. Eng. Educ. Life Long Learn.* 19, 191–205. doi: 10.1504/IJCEELL.2009.025027

Zhu, M., Sari, A. R., and Lee, M. M. (2022). Trends and issues in MOOC learning analytics empirical research: a systematic literature review (2011–2021). *Educ. Inf. Technol.* 27, 10135–10160. doi: 10.1007/s10639-022-11031-6

Supporting social interactions to improve MOOC participants' learning outcomes: a literature review

Hong Sen Loh[1], Gabrielle Martins van Jaarsveld[2]*,
Canan Mesutoglu[1] and Martine Baars[1]

[1]Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, Netherlands, [2]Centre for Education and Learning, Delft University of Technology, Delft, Netherlands

Introduction: The need for more scalable, accessible and affordable education, coupled with technological advancements in information sharing technology and collaborative platforms has led to the growth of MOOCs (massive open online courses). The growth of MOOCs has resulted in learning becoming increasingly distributed, affordable and flexible compared to traditional classroom education, allowing individuals in disadvantaged groups to access high quality educational materials. However, new challenges emerge, most notably with MOOCs' low completion rates. Among the reasons for a low completion rate, lack of interaction with peers is cited as a major reason, yet, interventions to promote social interaction has received relatively less attention. From a constructivist perspective, social interactions among peers are essential in helping individuals learn. This systematic literature review aimed to understand social interaction interventions in MOOC settings. This includes constructs measured as learning outcomes and how they were measured.

Methods: Literature articles were sourced from multiple databases and filtered for inclusion using the PRISMA process and its four phases. Twenty articles were included in the final review.

Results: Results of the review showed that social interactions fall into three categories: discussion forums, learning groups and one-to-one interactions with peers. Learning outcomes investigated were grouped into four categories: knowledge, social engagement, learning engagement and learner experience.

Conclusion: Across the studies, there were clear positive effects of social interaction on learning outcomes. Intervention strategies include sending prompts for individuals to use discussion boards and also grouping learners by homogeneity based on their learning engagement. However, more experimental studies are necessary to bolster the evidence of a causal impact of social interactions. Also, further research should be done to understand potential mediating factors that contribute to the success of implementing social interactive elements in MOOCs.

KEYWORDS

MOOC, social interaction, social learning, collaborative learning, learning outcomes

1 Introduction

The evolution of information communication technology has changed the landscape of many fields, education being one of the many. Individuals now have access to an increasing quantity of information from a vast array of sources. More online learning platforms have popped up and virtual classrooms are becoming more commonly used. Cloud storage has made resource sharing easier and as the internet and information communication technologies continue to grow, higher education institutes have also been adapting to this trend of digitalisation. The digitalisation of higher education has also appeared in the form of Massive Open Online Courses (MOOCs), an increasingly popular means of learning.

MOOCs are online courses created by universities, organisations or industry experts for potential learners interested in a topic area (Liyanagunawardena et al., 2013). Across university sites and platforms like edX, Udemy, and Coursera, thousands of MOOCs are currently available for free. From subjects like arts and humanities to computing, from nutrition to business management, participants can sign up for MOOCs from any part of the world, often for free. Classes come in the form of video lectures, documents and assignments. Since only a digital device is required, learning is also made highly convenient for participants.

A combination of factors has driven this popularisation of MOOCs, including the advancements in file sharing and video conferencing technology (Al-Samarraie, 2019) and education institutions' desire to reach a wider audience (Boggs et al., 2021). Additionally, a declining number of educators worldwide (University of Technology Sydney, 2023) implies a need to have reusable materials and instructional methods that can cater to students' learning needs without necessarily increasing the number of educators. MOOCs appear to provide the answer to this issue.

In higher education, MOOCs are generally followed in two ways. First, participants rely entirely on the MOOCs for learning content, however, there are still peer discussions about the content and assignments of the MOOCs. Students also receive support from educators. MOOCs are also used in a blended learning approach where students get their preliminary knowledge from MOOCs before doing seminars and taking assessments that are not part of the MOOC. A variation of the blended learning approach sees MOOC content used only as supporting materials to what has already been provided to students (Holotescu et al., 2014; Manli, 2014). For small university programmes where there are limited courses and educators, Andersen et al. (2019) propose using MOOCs to supplement students' education, noting that educators should still have scheduled sessions for students to clarify doubts. Assessing students' knowledge can be done via presentations at certain checkpoints of a MOOC and a final project for students to work on based on the MOOC materials.

MOOCs present multiple benefits for educators and participants. For educators, MOOCs are a scalable way to teach since learning content can be reused and their time spent does not increase as learners attending the MOOC increases (Pérez-Sanagustín et al., 2017). The extra time and effort can be channelled into improving learning materials. And since MOOCs take place in an online environment, data from participants' learning behaviours and progress can be obtained for learning analytics to optimise learning (Urrutia et al., 2017; Onah et al., 2018). For MOOC participants, having the flexibility to complete course deliverables based on their schedule is highly valued (Shapiro et al., 2017). With traditional courses in schools, students are locked into a study after enrolment but with MOOCs, students can dive into a domain and explore course materials without having the pressure of committing to finish it (Shapiro et al., 2017).

Despite the multiple benefits associated with learning in MOOC setting, historically high dropout rates have been a cautioning point of MOOCs being unable to live up to traditional learning environments. Initial studies on MOOC participant behaviour and completion rates have the median completion rate for courses to be around 10–13% after accounting for participants who became inactive shortly after signing up (Onah et al., 2014; Jordan, 2015; Reich and Ruipérez-Valiente, 2019). Reich and Ruipérez-Valiente (2019) commented on the likely possibility of MOOC platforms catering towards more affluent individuals in higher education, courses becoming less affordable and less accessible for most individuals if participant achievement rates remain low. Participant attrition has been linked to language-related barriers (Gomez-Zermeno and De La Garza, 2016), participant motivation (Wang and Baker, 2015; Xiong et al., 2015), and unmatching expectations (Eriksson et al., 2016), to name a few factors. Remedies have been suggested such as adding subtitles in English or a native language, profiling learners and having different retention strategies for participants with differing motivations (Xiong et al., 2015). Another cause of attrition was a lack of interaction and participants feeling a sense of isolation (Xiong et al., 2015; El Said, 2016). While discussion forums exist on MOOC platforms, understanding how to promote higher-quality online social interaction might play a role in improved completion rates.

A key difference between learning via a MOOC and learning in a traditional setting is the social environment learners are exposed to. In traditional learning environments, social interactions between students have been shown to foster beneficial learning outcomes such as more understanding of content, higher accuracy when quizzed and greater confidence in their knowledge (Tullis and Goldstone, 2020). In online learning contexts, interaction has been shown to result in benefits such as higher engagement with course content (Sunar et al., 2017; De Felice et al., 2021). However, multiple studies point to learners having difficulty with social interaction and a lack of it when learning online (Baber, 2021; Wut and Xu, 2021; Azmat and Ahmad, 2022; Ivanec, 2022). This finding supports the conclusion by Aldowah et al. (2019) that MOOCs lacking social support, social presence and peer interaction results in high participant dropout rates. Researchers agree that participants' feelings of isolation must be addressed and promoting peer engagement can play a role in increasing participants' learning success with the MOOC course (Wang et al., 2018; Aldowah et al., 2019; Williams et al., 2019). Therefore, this review aims to address this issue by presenting a timely synthesis of research done on ways social interactions between peers have been implemented in MOOCs.

1.1 Literature review

1.1.1 MOOCs

MOOCs have experienced substantial growth since 2011 (Onah et al., 2014). A relatively recent investigation by Reich and Ruipérez-Valiente (2019) concluded that while total MOOC enrolments have increased compared to the early years, completion rates have remained

low. Research has looked at attrition factors from multiple different perspectives. For course factors, longer courses, older courses and courses using peer grading (as opposed to auto-grading) were associated with lower completion rates (Jordan, 2015). Self-paced MOOCs—as opposed to MOOCs with fixed dates for assignment submissions positively predicted higher student satisfaction, arguably because it gives students more autonomy in their learning (Hew et al., 2020). For factors related to learner contexts and characteristics, participants with higher levels of education were found to have higher self-regulated learning sub-processes and higher self-efficacy, both of which are associated with MOOC completion (Hood et al., 2015). Additionally, the likelihood of completion was higher for students with prior experience with MOOCs, and participants with a higher self-reported commitment such as participants who specified the number of hours they intended to spend (Greene et al., 2015). Courses that emphasised active learning with assessments that went beyond knowledge recall and that allowed students to apply what they had learnt were associated with higher satisfaction (Hew et al., 2020). Joo et al. (2018) highlighted students' perception of a MOOC's usefulness and its perceived ease of use to have a positive influence on satisfaction with the course which was associated with continuance with the MOOC. On a related note, Jung and Lee (2018) focused on learning engagement, finding it to be a mediator between some of the previously mentioned variables—self-efficacy, perceived usefulness and ease of use—and learning persistence. In summary, previous research has put forward strong explanations to explain attrition rates, attributing them to course factors (e.g. course length, grading system, flexibility), learner characteristics and perceptions of course usefulness.

Research and practice have invested greatly to understand MOOC attrition rates from multiple perspectives but an area of investigation that has received less attention is the impact of social factors. There are inconclusive findings as to whether social interaction play a huge role in learning in MOOC contexts. For example, Hew (2014) found the quality of peer interaction to be highly rated as important by participants from three top-rated MOOCs. However, a contrasting finding was made by Gameel (2017), concluding that participants' course satisfaction did not increase when provided opportunities for discussion on forums as they found it to be chaotic, often going off-top and overwhelming. This lack of clarity calls for a need to synergise research on the impacts of social interaction on MOOC learning outcomes and how to promote it.

To understand the effects of social interaction on MOOCs, it is necessary to understand the types of social interaction that happen in MOOCs. Interactions usually take place on discussion forums that are often built into the MOOC platform (Mayende et al., 2017). However, Veletsianos et al. (2015) showed that social interactions can happen beyond the confines of discussion forums, with learners sharing their experiences with family or discussing course content with individuals who might be friends or other learners enrolled in the same MOOC. There also seem to be different degrees of connection with some learners preferring to post messages on discussion boards and others directly reaching out to other learners through forums or social media to form closer connections (Veletsianos et al., 2015). It is thus important to recognise that different forms of social interactions can exist when promoting them in MOOC settings. Investigating these differences and their impact on participants' learning outcomes is imperative should educators want MOOCs to stay relevant in education.

1.1.2 Social interaction in online learning environments

Multiple prominent learning theories can be used as starting points to understand how social interaction potentially impacts learning. Social Learning Theory (Bandura, 1978) proposes that individuals learn behaviours through modelling behaviours of others, such as peers or teachers. In an example context of problem-solving, an individual learns via observation and imitation of another's methods before later applying the method to a similar problem.

Constructivism emphasises learning to be an active process and knowledge to be constructed by learners (Piaget, 1964; Narayan et al., 2013). According to Piaget (1964), the construction of knowledge happens through assimilation—fitting new experiences such that they are consistent with existing mental schemas, and accommodation—revising existing mental schemas to be consistent with new experiences and information. The second form of constructivism, social constructivism, proposes a similar idea, but individuals learn in a social setting by sharing their knowledge and assimilates and accommodates new information according to what was shared in that group.

This idea of learning from others has similarities with Siemens' theory of Connectivism (Siemens, 2005) where learning occurs when networks are formed between information sources. Knowledge is the result of connecting ideas and concepts from an array of sources which could be expanding on pre-existing ideas. Thus, having a diverse number of information sources and being able to tap into them is critical in learning and acquiring knowledge.

The three theories highlight key complementary areas that are necessary for learning. From Connectivism, the importance of connecting with different individuals is highlighted. The different individuals are information sources where we can observe and imitate others, as posed by Social Learning Theory. Information gained through observation and interaction with others leads to active reflection of new knowledge and revision of pre-existing knowledge, as proposed by Constructivism. Salomon and Perkins (1998) and Vygotsky (1978) emphasised the importance of social interaction and collaboration in an individual's learning process as learning is social in nature. Through interaction with others, students are exposed to a variety of thinking processes which promote learning. Knowledge is more effectively and efficiently developed in a social context as the construction of one's knowledge can be built on the understanding of peers (Brown et al., 1989) and problems or concepts that are too difficult for the individual become solvable with the assistance and guidance of others (Vygotsky, 1978).

In the context of MOOCs, courses should ideally incorporate elements of the above outlined in the theories: connecting participants, discussions to promote sharing of different opinions and approaches towards learning topics and support structures for participants to learn from and guide one another.

With regard to learning outcomes, multiple different indicators of MOOC learner success have been outlined. While course completion or dropout rates seem to be the metric of assessment for interventions (Jordan, 2015; Xiong et al., 2015), other researchers have chosen to focus on other metrics such as final grades (Deng et al., 2019) or learning engagement as measured by a rating score or motivation (Ramesh et al., 2014) or engagement as measured by completion of course activities (Coffrin et al., 2014). Other outcomes studied were learners' satisfaction or behavioural indicators, for example,

participation in discussion forums (Wintrup et al., 2015). Similarly, with different types of social interactions, a variety of constructs and metrics of assessment can be used to measure its effects, each possibly showing a different effect on attrition rates.

1.2 The current study

This review aims to synthesise research on how social interactions in MOOCs can be supported in an effective way considering participants' learning outcomes. The secondary aim of this review is to summarise the learning outcomes that were investigated in these studies on social interactions in MOOCs. Therefore, we have reviewed how learning outcomes were measured and what metrics were used. Dropout and completion rates in MOOCs are a common evaluation metric used by researchers (Onah et al., 2014; Jordan, 2015). However, dropout and completion rates are but a proxy for measuring change in constructs such as motivation, learning engagement and knowledge attained, to name a few. With each construct, there might be multiple ways of measuring them. For example, learning engagement could be operationalised in multiple ways such as self-ratings, duration of time spent on learning materials or the number of learning content viewed.

The main research question posed in this systematic review is: how can social interaction be supported effectively to improve MOOC participants' learning outcomes? Two sub-questions were addressed to answer the main research question. (A) What are ways in which MOOCs have incorporated social interactions? (B) What learning outcomes have been investigated with regard to social interactions and how are they measured?

Identifying effective methods of implementing interaction elements from past MOOCs can guide MOOC designers on appropriate ways of integrating peer interaction into the MOOCs so participants can have more fruitful learning experience. By also summarising the constructs and measurement metrics used along with suggestions for social interaction implementation ideas, future MOOC designers will not only be able to integrate social learning elements but also select appropriate evaluation methods to match the social interaction elements implemented.

2 Method

A systematic literature review method was used to identify empirical articles related to MOOC courses incorporating social interaction elements. The PRISMA framework (D Moher et al., 2009, p. 8) and recommendations by Cooper (2015) were used to ensure a systematic way to identify and assess the quality of the articles. Figure 1 presents the four phases of the PRISMA framework: identification, screening, eligibility and inclusion.

In phase 1, the identification phase or literature search stage (Cooper, 2015), a literature search was conducted from 3rd May to 13th May 2023. Literature was sourced from the following databases: Scopus and ERIC. Additional studies included were identified from the reference list of research papers. The database search was limited to title, abstract and author keywords. Keywords used were 'MOOC', 'massive open online course' and words relating to 'social' like 'peer' and 'group'. An asterisk was used as a wildcard at the end of a keyword

to include different forms of the word and broaden the search (e.g., interact* also searches for 'interaction' and 'interactions'). The databases were searched using the following query:

'("MOOC*" OR "Massive Open Online Course*") AND ("social" OR "group*" OR "peer" OR "communit*") AND ("influence" OR "learning" OR "interaction" OR "communication" OR "learning outcomes) NOT "review"'.

The search yielded 180 records. The records were then screened for duplicates. 39 duplicate items were removed.

In phase 2, the screening phase or data evaluation phase (Cooper, 2015), the title and abstracts were screened for the remaining 141 articles and filtered with the following inclusion criteria: (1) articles must be written in English; (2) articles must be about MOOCs; (3) articles must explore or examine the role of social interactions between learners; (4) articles must indicate learning outcome variables (5) research that are empirical in nature. No restriction was made on the types of empirical studies included as there was a limited sample of studies in this area. Thus, correlational and experimental studies were all included. 93 articles were filtered out as they did not meet the inclusion criteria.

In phase 3, the method, results and discussion sections of remaining 48 articles were filtered for and excluded using the following criteria: (1) the article did not explore or examine interaction between learners or how the interaction happened was not well defined; (2) the article did not report participants' learning outcomes clearly; (3) the research was not in a MOOC context; (4) the articles were opinion pieces or reviews; (5) the full-text articles were not available online; (6) articles are not peer reviewed. The filtering resulted in 28 articles being excluded and 20 articles being selected for the literature review.

In the phase 4, in the inclusion or analysis and interpretation stage (Cooper, 2015), the researchers worked collaboratively to ensure consistency and agreement. Descriptive information of the selected articles were extracted and organised as displayed in Table 1. Table 2 was also constructed to answer the two research questions—the ways in which MOOCs have incorporated social interactions, and the learning outcomes that have been investigated and measured. After multiple rounds of reading the articles, they were first sorted by the type of social interaction that was implemented in the MOOC. Then, four main categories of learning outcomes were identified and for each article, it was noted if any of the learning outcomes were included as an evaluation metric in some form.

3 Results

In this section, a descriptive overview of the included studies will first be reported. Then, the findings of the selected studies will be synthesised in three sub-sections: (a) how social interaction takes place in MOOCs; (b) constructs researched in MOOCs with social interactions and how the constructs they were measured; (c) an overarching synthesis of the findings, answering the main research question of how social interaction can be supported more effectively to improve MOOC participants' learning outcomes.

Of the 20 studies selected for this review, 14 were published relatively recently; within the last 5 years at the time of writing. Most ($n = 14$) MOOCs were between 4 to 8 weeks in duration. MOOCs varied and included topics like business and entrepreneurship,

FIGURE 1
PRISMA diagram of article selection process.
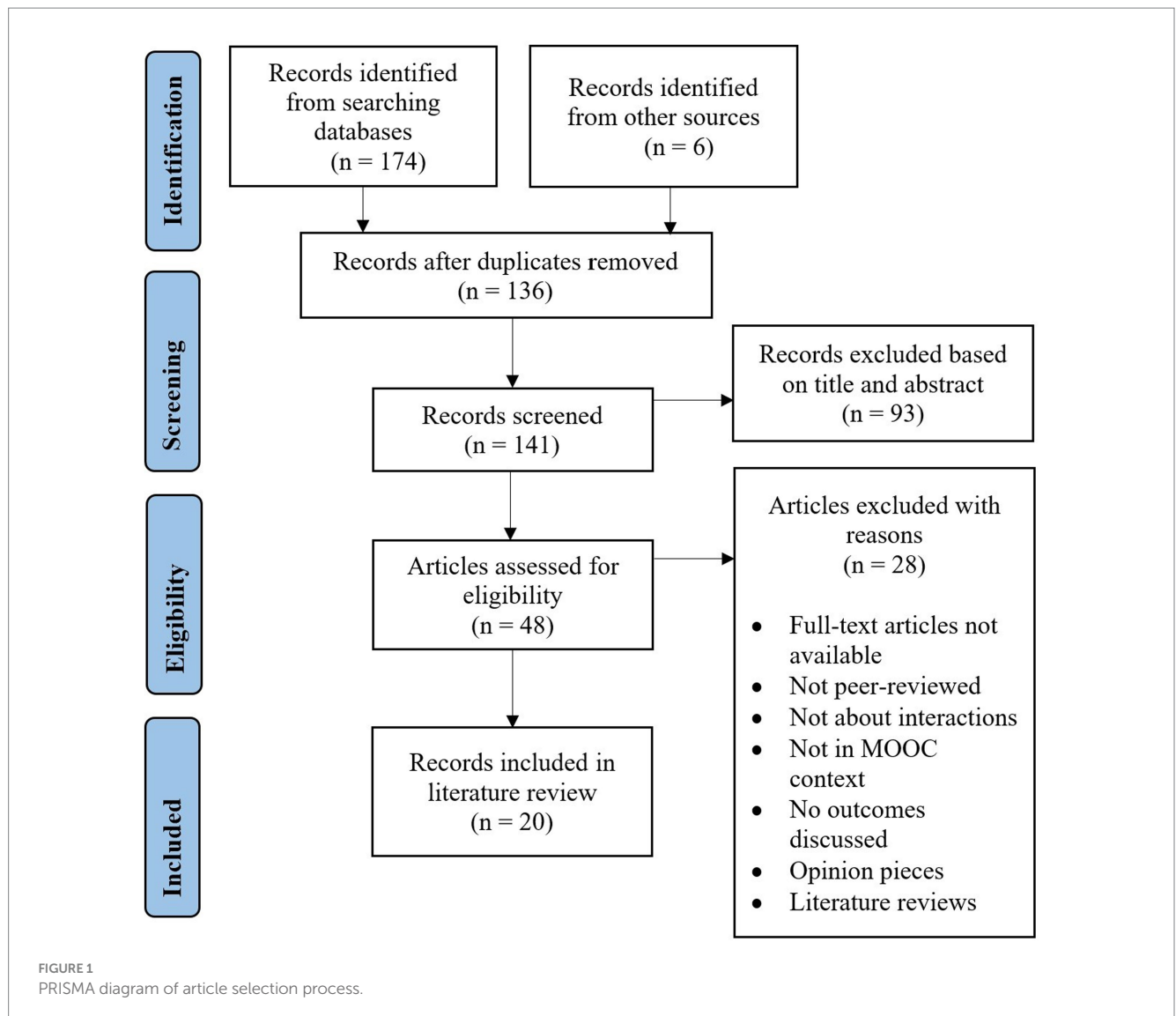
education planning, improving creativity, health care, and programming, to name a few. For study design, seven were case studies, four were mixed-method, six were experimental and three were correlational.

3.1 Sub-RQ1: social interaction in MOOCs

Comparing and contrasting the type of interactions in the 20 articles selected for the review, how social interactions are incorporated in MOOCs can be grouped into three categories differing in degrees of interpersonal closeness: (1) discussion forums (2) groups (3) peer (one-to-one interaction). A distinction must be made between the 'discussion forum' and 'group' categories since discussion forums can be viewed as a large group containing all individuals in the MOOC. The 'group' category contained studies with individuals who were matched with peers whom they interacted with regularly. The difference in interpersonal closeness between interacting individuals was the main criterion used to separate the two categories.

The research by Zhang et al. (2017) included three studies. Experiments 1 and 2 were focused on social interactions in discussion forums while experiment 3 focused on social interactions between peers. Thus, the study by Zhang et al. (2017) was counted in both the 'discussion forum' and 'group' categories.

Of the nine studies researching social interaction on discussion forums, five studies investigated the relationship between discussion forum posting and a variety of outcomes including co-construction of knowledge (Kellogg et al., 2014), motivation (Barak et al., 2016), dropout rate, participation rate (Sunar et al., 2017), content understanding (Gillani and Eynon, 2014), social engagement and content engagement (Shi et al., 2019). Four studies differed slightly from the rest. Zhang et al. (2017) looked at prompts to promote social interaction on discussion forums while Xu et al. (2019) compared the difference between on-topic and off-topic posting on learning outcomes. On topic posts were posts and messages that were relevant to the course content while off topic posts were posts that had no relation to course content. Anderson et al. (2020) researched how social media affected learning experiences by encouraging

TABLE 1 Descriptive information of included articles.

Authors	Year	Study design	MOOC topic / title	MOOC duration
Anderson et al.	2020	Case study	School for Health and Care Radicals	5 weeks
Barak et al.	2016	Mixed-methods	Nanotechnology and Nanosensors	10 weeks
Bouchet et al.	2017	Experimental	Project management	2–3 months
Gamage	2021	Mixed-methods	Object-Oriented Programming in Java	6 weeks
Gamage and Whiting	2021	Experimental	Creative problem solving	4 weeks
Gillani and Eynon	2014	Case study	Business strategy	6 weeks
Kellogg et al.	2014	Mixed-methods	Planning for the Digital Learning Transition in K-12 Schools	6 weeks
Krasny et al.	2018	Case study	Environmental Education: Transdisciplinary Approaches to Addressing Wicked Problems	–
Mayende et al.	2017	Case study	Success - Unleash Yourself	8 weeks
Pin-ju and Chen	2022	Case study	An Introduction to Marketing	6 weeks
Razmerita et al.	2020	Correlational	Social Entrepreneurship	12 weeks
Sanz-Martínez et al.	2019	Experimental	How to translate economy and finance texts from Spanish to English	8 weeks
Shi et al.	2019	Correlational	Literature and Mental Health	6 weeks
Su et al.	2016	Case study	Computer networks	20 weeks
Sunar et al.	2017	Case study	Developing Your Research Project	8 weeks
Wichmann et al.	2016	Experimental	Computer-mediated communication in teaching and learning	14 weeks
Xu et al.	2019	Correlational	Educational data mining and the analysis of big data in education	8 weeks
Yang and He	2022	Mixed-methods	Quadratic function in math for grade-10 students	–
Zhang et al.	2017	Experimental	Business operations strategy	5 weeks
Zhang et al.	2016	Experimental	Creativity, Innovation, and Change	6 weeks

participants to share knowledge and have discussion through social media tools such as Facebook and Twitter. Yang and He (2022) evaluated the effectiveness of a tool simulating a pseudo-synchronous atmosphere in bridging learners.

Of the 10 studies included in the group category, three studies examined how groups can be optimally created (Wichmann et al., 2016; Zhang et al., 2016; Sanz-Martínez et al., 2019) by considering participants' communication preferences and interaction with course content. Two studies examined learning groups led by volunteers (Krasny et al., 2018; Gamage, 2021). Two studies researched different tools with affordances that enhanced collaboration and learning outcomes in groups (Su et al., 2016; Gamage and Whiting, 2021). Mayende et al. (2017) investigated learner experiences and ways in which groups were beneficial. Pin-Ju and Chen (2022) studied communication differences between face-to-face and online group discussions. Razmerita et al. (2020) researched how learners' attitudes and communal influence shape their engagement in collaboration and impact learning and behaviour.

In the peer category, the two included studies were one by Bouchet et al. (2017) which looked at ways in which peers can be matched up and experiment 3 by Zhang et al. (2017) examining the effect of prompting students to have one-to-one peer discussions.

3.2 Sub-RQ2: learning outcomes and methods of assessment in MOOCs

In this section, the categories identified previously will be collapsed so the most and least common outcomes investigated can be identified. The learning outcomes will first be grouped into general categories before focusing on the different ways they were operationalised and measured. Table 2 displays a summary table of learning outcomes examined by each study.

3.2.1 Knowledge

Almost half of the included articles ($n = 9$) evaluated participants' knowledge as a learning outcome. This included some scoring via tests or assessments aimed at gauging whether content of the MOOC was understood and applied by the participants. Evaluation methods also included self-reports and interviews. The most common method of assessing participants' knowledge being the use of quiz scores and final grades, as was done by Gillani and Eynon (2014), Zhang et al. (2017) and Xu et al. (2019). Such metrics are built into most MOOC platforms (e.g. Coursera), making it an easy metric to use for evaluation. The link between interacting on discussion boards and final course grades is unclear. Gillani and Eynon (2014) concluded that they are significantly related while Zhang et al. (2017) found no

TABLE 2 Learning outcomes examined by each study.

Authors	Year	Learning outcomes			
		Knowledge	Social engagement	Learning engagement	Learner experience
Discussion forum					
Anderson et al.	2020	x			x
Barak et al.	2016				x
Gillani and Eynon	2014	x			
Kellogg et al.	2014		x		
Shi et al.	2019		x	x	
Sunar et al.	2017			x	
Xu et al.	2019			x	
Yang and He	2022		x		x
Zhang et al. (study 1 & 2)	2017	x			
Groups					
Gamage	2021		x		
Gamage and Whiting	2021		x		x
Krasny et al.	2018		x		x
Mayende et al.	2017				x
Pin-ju and Chen	2022				x
Razmerita et al.	2020				x
Sanz-Martínez et al.	2019		x	x	x
Su et al.	2016	x	x		
Wichmann et al.	2016	x		x	
Zhang et al.	2016	x			
Peer					
Bouchet et al.	2017		x		
Zhang et al. (study 3)	2017	x		x	

significant relationship. Xu et al. (2019) reported that participants who contributed to on-topic or a mix of on-topic and off-topic posts in discussion forums made up a higher proportion of students with non-zero final grades as compared to students with off-topic posts only. With one-to-one interaction between peers, Zhang et al. (2017) found that it led to an improvement in quiz scores. However, the results must be interpreted with caution as only a small fraction of the total students invited had these interactions.

Another method of assessment is a pre- and post-course survey such as the one used by Anderson et al. (2020). In their research, survey items used a Likert scale, measuring components that the course aimed to improve (e.g. purpose and motivation, theoretical understanding, ability to connect with others and support change initiatives). A quantitative analysis showed no significant difference between pre- and post-survey for most components assessed. However, there was a significant relationship between participation as part of a team and scores for the component of 'maintaining collaborative relationships'.

Qualitative methods such as interviews were also used. In the research by Anderson et al. (2020), participants volunteered to do a

semi-structured telephone interview with researchers. From the interview data, participants mentioned that being in groups enhanced their perseverance during the learning.

Participants could also be scored on tasks like in the research by Su et al. (2016) where they tested a new social searching tool that enabled students to share their problem-solving and searching process. Participants were assigned searching tasks and assessed on their searching process and searching abilities. A coding scheme was used to score relevant processes like 'specifying search terms', and 'evaluating the search result'. Results of the study showed that the problem-solving abilities of students improved and students were able to learn search strategies from other students.

Written submissions were also a method of assessing participants' knowledge. The research by Zhang et al. (2016) focused on a MOOC about creativity and innovation where participants did creative exercises on their own and submitted reflections but were not graded on them. Wichmann et al. (2016) had participants submit assessments as a group and assessed knowledge by counting the number of domain concepts included in the text. In their study investigating grouping participants by learning engagement homogeneity, significant

differences were found for different grouping homogeneity and the number of concepts included in text submissions.

3.2.2 Social engagement

Social engagement during a MOOC was an outcome measured by about half of the selected articles ($n = 9$). Included in this category are quantitative or qualitative social engagement measures related to participants' interaction with peers. A common measurement of a participant's social engagement was the frequency of messages posted on discussion boards or messages sent in a group. Shi et al. (2019) tracked the number of comments that participants made each week in the course. Other researchers like Yang and He (2022) used a survey for participants to self-report their social engagement, having participants indicate how regularly they used the social interaction tool provided to them.

In group settings, Krasny et al. (2018) surveyed and interviewed group community leaders to check the number of meetings and the attendance rate of participants. Sanz-Martínez et al. (2019) compared groups formed based on different homogeneity levels and compared them based on active participants in the group, messages posted by each participant and the number of students that participated in the collaborative assignments. A case study by Bouchet et al. (2017) tested a peer recommender system widget. Social engagement measurements used were participants who opened the widget at least once, participants who opened a discussion thread at least once and participants who sent at least one message to a discussion thread. In a similar case study by Su et al. (2016), to evaluate the effectiveness of a tool in aiding interaction and learning within a group, sharing time and sharing frequency were recorded to indicate the proportion of members that used the tool. Groups were identified as 'collaborative' or 'individual' based on these metrics.

Another way in which researchers looked at social engagement was by collecting participants' forum posts and messages and categorising them before analysis. Yang and He (2022) researched social presence by using a scheme to group posts into different categories: emotional expression, open communication, group cohesion, co-presence, and relationship development. Gamage and Whiting (2021) researched a similar topic of social presence but in groups. They labelled behaviours with different indicators according to the Communities of Inquiry model. Examples include an affect indicator for emotional content and an interactive indicator for acknowledgement or questions by peers. To analyse messages for knowledge building, Gamage (2021) used the Epistemic Network Analysis, a framework that categorised a conversation as 'cognitive task', 'social task', 'social non-task' or 'cognitive-non-task'. Analysis of conversations in CollabSpace (which incorporated a grouping framework) showed a strong association between social tasks and non-social tasks, and also non-social tasks and cognitive tasks. This indicated social presence being influential to learning in the community. Kellogg et al. (2014) used the Interaction Analysis Model to measure the extent to which social engagement resulted in co-construction of knowledge. Discussion threads were categorised into five different phases with the fifth being most indicative of knowledge not only being co-constructed but able to be applied by participants. Results from the research showed that participants' interactions were mainly in the first phase (providing observations, opinions or examples that support or extend prior statements). Some interactions reached Phase 2 (identifying areas of agreement or disagreement) and Phase 3 (exploring common ground in views and seeking to integrate ideas), however, few interactions went beyond this phase.

3.2.3 Learning engagement

A few articles measured participants' learning engagement during the MOOC as an outcome ($n = 4$). Learning engagement measures included participant's activity relating to course materials, tasks or assignments. In the research by Shi et al. (2019) and Sunar et al. (2017), steps were basic learning items in each week of the course, which could include articles, images or videos. For each week of the course, the step visit rates—the number of participants who visited at least one step—as well as completion rates—the number of learners that completed at least one step were recorded. In both research, findings show that participants who contributed to discussion threads completed more steps than passive participants.

The other studies that included learning engagement as an outcome variable were studies in group contexts examining the effects of grouping participants by learning engagement homogeneity. Wichmann et al. (2016) measured the quantity of text contributed by each participant in a group and the quantity of text submitted by the whole group to determine group productiveness while Sanz-Martínez et al. (2019) measured learning engagement by looking at the number of groups that completed tasks and submitted assignments. Results from both studies suggest that homogeneous grouping of students led to more learning engagement as measured by assignment text quantity, task completions and assignment submissions.

3.2.4 Learner experience

The other learning outcomes evaluated by researchers could be encompassed in the broad category of learner experience. This category included participants' perception of the grouping structure, satisfaction with the MOOC and their motivation. Approximately half of the selected articles ($n = 9$) had explored learner experience as an outcome.

Surveys were the most commonly used method to gather participants' learning experiences. The focus of survey items varied considerably between different researchers. Sanz-Martínez et al. (2019) and Razmerita et al. (2020) asked participants to rate their satisfaction with their group and group collaboration experiences with a Likert scale. Gamage and Whiting (2021) and Mayende et al. (2017) looked at participants' perceptions of different aspects of group learning by asking them if they agreed or disagreed with statements like, "Our team was effective," "Our team was supportive" and "I received positive feedback." Learners were also asked what the group helped with, from questions about motivation, understanding the content, and technical support. The case study by Yang and He (2022) involved participants being provided with a learning tool and survey questions centred around ease of use and if the learning tool fostered the social presence of others and a sense of belongingness. The survey by Gamage and Whiting (2021) was similar except that questions were gauging participants' sense of belonging with a six-item scale. Participants were asked if being in the group made them feel understood, connected, welcomed, if they felt that they were respected by others, and if they were happy in the group.

Participants' learning experience was also collected using interviews. Pin-ju and Chen (2022) recorded interviews with participants and held focus group meetings. Anderson et al. (2020)

held a semi-structured interview with volunteers. The interviews allowed participants to elaborate more on their experiences and for researchers to reflect on aspects that were potentially outside the researchers' considerations. Additionally, in the research by Anderson et al. (2020), the qualitative data showed a contrast to the quantitative evaluations. Interviewees expressed that support and encouragement were felt from interactions with other participants and that they had a sense of shared purpose with others.

Another outcome related to learner experience was participants' motivation, as examined by three studies. All three studies used surveys. Krasny et al. (2018) also interviewed group leaders to understand their motivations for volunteering. Barak et al. (2016) compared pre and post-course motivation ratings, finding that participants working alone on the final project had relatively low means for motivation to learn as compared to participants working in groups of four or five. Barak et al. (2016) conclude that participants' motivation increases in small groups as these small-group discussions stimulate interest in the learning content. This effect of group learning on motivation is echoed by Mayende et al. (2017) however they attribute motivation gain to peers frequently interacting in the group, providing feedback and exhibiting high commitment to making progress in the MOOC.

3.3 Main research question: how can social interaction be supported in MOOCs?

In this section, we synthesise findings from the literature on how we can support social interaction in MOOCs. The findings will be broken down into different categories of interaction levels discussed above: discussion forums, groups, and one-to-one interactions with peers.

3.3.1 Supporting social interactions on discussion forums

From the studies above, sending prompts for participants to use discussion forums has been shown to be effective in encouraging the use of it (Zhang et al., 2017). There are consistent findings to support the benefits of social interactions on discussion forums including a higher completion rate of MOOC learning activities, course completion and higher course grades (Gillani and Eynon, 2014; Manli, 2014; Zhang et al., 2017; Shi et al., 2019). There is also evidence to suggest that discussions move beyond the phase of sharing information and statements of agreement, with participants co-constructing knowledge by expressing dissonance over understanding of learning materials (Kellogg et al., 2014; Manli, 2014).

Participating in both on and off-topic discussions early in the MOOC increases participants' course engagement (Sunar et al., 2017; Xu et al., 2019). Thus, MOOC designers can encourage participants to participate in off-topic chats initially, so connections are built between individuals that can serve as a vital source of support as the MOOC progresses and gradually encourage on-topic discussion as the course progresses. For off-topic chats, participants can introduce themselves and their interests, which could help them find matching participants with mutual interests. On-topic posts that can be encouraged include sharing notes, asking questions and answering other participants' questions.

MOOC learning is asynchronous which can lead to participants feeling isolated. To remedy this, MOOC designers can look into tools such as Danmaku, a tool tested by Yang and He (2022) to foster a pseudo-synchronous learning environment where participants feel as if they are watching lectures with their peers at the same time. The study's results highlight the tool's usefulness in facilitating deeper connectedness with peers and promoting help-seeking interactions which can serve participants well in overcoming difficulties with course materials.

3.3.2 Supporting social interactions in groups

The research detailing strategies for supporting social interactions in learning groups is grouped into three broad categories: (1) optimal grouping methods; (2) leader-led groups; and (3) assistive tools.

Regarding grouping participants, from the studies reviewed, grouping participants based on their learning and social engagement leads to better learning outcomes as compared to grouping by communication preferences. MOOC administrators can group participants 2 to 3 weeks into the course and not at the beginning course. From the beginning of the course, data should be collected on participant's engagement with the MOOC, including their page views, submitted assignments and the number of forum messages. Expected learning outcome gains from this method of grouping include higher group productivity and assignment quality (Wichmann et al., 2016), increased peer interactions and increased satisfaction with collaboration (Sanz-Martínez et al., 2019). Not grouping participants at the outset has the added benefit of allowing the initial drop-off in participants to level off before the formation of groups, decreasing the chance of learners being demotivated by inactive group members.

MOOC designers and administrators can also consider leader-led groups with volunteers. Krasny et al. (2018) invited volunteers who had previously completed the course while Gamage (2021) and Gamage and Whiting (2021) assigned individuals who indicated their interest as a leader during the grouping process. Social presence was felt in leader-led groups and participants also mentioned feeling a sense of belonging (Gamage, 2021; Gamage and Whiting, 2021). Individuals who had completed the MOOC and volunteered to lead groups assisted in dividing tasks and learning content among participants, facilitated meeting sessions and explained difficult topics (Krasny et al., 2018).

With regards to assistive tools, MOOC designers can look into tools such as PeerCollab, a community-building tool used by Gamage and Whiting (2021) that creates communities in forum spaces. Gamage and Whiting (2021) suggested that most MOOC discussion forums are populated with cognitive presence but limited in social presence. A tool like PeerCollab can help learners find community groups with shared learning goals and engage in leader-driven community activities.

Another tool that MOOC designers can look into is a social searching system that Su et al. (2016) did a case study on. The tool is akin to a group search diary and activity log, helpful for tasks where participants are searching and evaluating information. Also, it is especially helpful for participants who are limited to asynchronous collaboration. In the system, there is a search window, group history, search suggestions, a web annotation tool, and a discussion room. Participants can discuss their searching and problem-solving process and view their peers' activity history. A social collaborative tool like this being integrated into a MOOC platform would allow MOOC

participants to not only interact but also learn and collaborate more seamlessly.

3.3.3 Supporting one-to-one peer interactions

While one-to-one peer interactions are not focused on as frequently, the experimental research by Zhang et al. (2017) shows some evidence that discussions with a peer improve participants' quiz scores and completion of course activities. However, the study's findings are limited as only a small proportion of overall participants followed through after indicating their desire and being matched with a peer for a discussion. A possible explanation is that more effort is required to coordinate a meeting with a peer as compared to using the discussion forum. Another potentially related issue is participants' fear of judgement preventing them from reaching out to their peers, as noted by Bouchet et al. (2017). More research is necessary to understand how one-to-one interactions with peer affects learning outcomes and overcoming barriers to one-to-one interactions.

4 Discussion

This review set out to examine studies done on social interaction in MOOC contexts to understand how to effectively promote social interactions among MOOC participants. Two sub-questions were posed to help answer the main research question. The first question was, what are ways in which MOOCs have incorporated social interactions? To conclude, our results showed that the types of social interaction differed in interpersonal closeness. Interpersonal closeness ranged from most to least number of individuals that could participate in a discussion topic. For example, in discussion forums there are usually a large number of participants which makes the interpersonal closeness less, whereas in one-to-one interactions the interpersonal closeness would be high. Interactions usually occurred in the form of messages in discussion forums (Gillani and Eynon, 2014; Kellogg et al., 2014; Manli, 2014; Barak et al., 2016; Sunar et al., 2017; Zhang et al., 2017; Shi et al., 2019; Xu et al., 2019; Anderson et al., 2020; Yang and He, 2022), in groups (Su et al., 2016; Wichmann et al., 2016; Zhang et al., 2016; Mayende et al., 2017; Krasny et al., 2018; Sanz-Martínez et al., 2019; Razmerita et al., 2020; Gamage, 2021; Gamage and Whiting, 2021; Pin-Ju and Chen, 2022) or one-to-one with another peer (Bouchet et al., 2017; Zhang et al., 2017).

The second question sub-question was what learning outcomes have been investigated with regard to social interactions and how are they measured Our results showed that the learning outcomes investigated in the studies that were reviewed fall into four broad categories: knowledge, social engagement, learning engagement and learner experience. For knowledge, researchers operationalised it using quiz scores (Zhang et al., 2017), final course grades (Gillani and Eynon, 2014; Zhang et al., 2017; Xu et al., 2019), self-reported ratings (Anderson et al., 2020), assessments (Su et al., 2016) and written text submissions (Zhang et al., 2016). Social engagement was measured by the number of messages posted on discussion boards, in groups, and participant attendance rates in group meetings (Krasny et al., 2018). Messages and interactions were also categorised and analysed for themes such as group cohesion, co-presence, emotional content (Gamage and Whiting, 2021; Yang and He, 2022), on-task relatedness (Gamage, 2021; Gamage and Whiting, 2021) and co-construction of

knowledge (Kellogg et al., 2014). For learning engagement, researchers looked at the number of activities viewed and completed by participants in the duration of the course (Sunar et al., 2017; Shi et al., 2019), and the quantity and quality of text in written assignments (Zhang et al., 2016). Learner experience was mainly centred around participants' satisfaction with different aspects of working in groups (Sanz-Martínez et al., 2019; Razmerita et al., 2020) and the degree to which they felt connected with other group members (Mayende et al., 2017; Gamage and Whiting, 2021; Yang and He, 2022). Additionally, participants' motivational outcomes were also collected (Barak et al., 2016; Mayende et al., 2017; Krasny et al., 2018).

Based on our review we can conclude that multiple favourable learning outcomes can be expected when MOOCs include some element of social interactions among peers. Outcomes include improving participants' quiz scores (Gillani and Eynon, 2014; Zhang et al., 2017; Xu et al., 2019) and building students' problem solving process (Su et al., 2016). Social interaction was also related to a higher number of domain concepts included in participants' written text assessments (Wichmann et al., 2016). Participants who were active on discussion boards interacted more with course materials (Sunar et al., 2017; Shi et al., 2019). When grouping of participants were effective, participants completed more tasks, submitted more assignments and turned in assignments of higher quality (Wichmann et al., 2016; Sanz-Martínez et al., 2019). Social interaction in groups stimulated interest in course materials (Barak et al., 2016) and were also a source of motivation to persevere in the MOOC (Barak et al., 2016; Mayende et al., 2017; Anderson et al., 2020).

Concerning our first research question on the type of social interaction in MOOCs, this review substantiates the existing literature's findings that social interactions usually happen in discussion forums but also occur outside of it (Veletsianos et al., 2015). The findings of this review extend this insight by extracting unique implementations of social interaction in discussion forums as well as other forms of effective online social interaction. When designing a MOOC to include social interactive elements, MOOC designers can use this categorisation as a starting point to identify the type of interaction they which to promote. For each of the categories, different strategies can be applied to promote interactions between participants.

This review is intended to help MOOC designers and administrators know how to support peer interactions and thus, the recommendations for each category of interaction will be summarised. For discussion forums, MOOC administrators can send messages via email or via the MOOC platform to encourage learners to use discussion forums (Zhang et al., 2017). Messages in the prompt could include getting learners to share what they have learned after watching a lecture, ask questions about assignments and answer others' questions (Zhang et al., 2017; Shi et al., 2019; Xu et al., 2019). Especially at the start of the course, designers should also encourage learners to participate in off-topic discussions so learners can potentially form connections that can serve to motivate and assist them later in the course. Alea et al. (2023) showed that students' exchanging their personal stories in an online course contributed to the teaching and learning outcomes. To continue, tools such as Danmaku can also be considered as it creates a pseudo-synchronous learning environment and foster deeper connectedness among peers. Tools that also offer discussion prompts present a wide range of perspectives to the participants and encourage active participation (Gao et al., 2013; Lieu et al., 2022).

Should MOOC designers choose to implement a grouping structure in the course, participants can be grouped based on their learning engagement for the first few weeks of the course or so (Sanz-Martínez et al., 2019). A homogeneous group seems to be the most effective overall at bringing about a range of different learning outcomes (Wichmann et al., 2016; Sanz-Martínez et al., 2019). Groups with volunteers taking on a leadership role are also effective in community building and for participants to overcome difficulties (Krasny et al., 2018; Gamage, 2021). However, from the studies done, recruitment of volunteers for a leadership role might not be scalable (Gamage, 2021), casting doubt on the viability of having volunteers as group leaders. One strategy to attract participants to volunteer could be to communicate the positive features associated with a leading role, e.g., offering a hands-on facilitation experience, ownership in directing a discussion (Hew, 2015). Assistive tools to look into include PeerCollab for creating community groups in MOOC forums and a social search tool like the one studied by Su et al. (2016) which provides participants with a seamless collaboration experience.

With promoting one-to-one peer interactions, there is insufficient research in this review for recommendations to be made. While benefits for peer interaction were noted (Zhang et al., 2017), more apparent issues that should be tackled first are learners having a fear of judgement (Bouchet et al., 2017), possibly explaining their not being open to interaction (Bouchet et al., 2017) and a low number of participants not following through on contacting their peers after being matched up (Zhang et al., 2017). Future research should delve more into reasons for this fear of judgement and how barriers to peer interaction can be removed.

Concerning the second research question on the type of learning outcomes investigated in relation to social interaction in MOOCs, the findings in this review align with constructivist theories of learning which propose that interactions among peers are integral to learning (Vygotsky, 1978; Salomon and Perkins, 1998). That is, from the review of the studies, multiple favourable learning outcomes can be expected when MOOCs include some element of social interactions among peers. Elements can include encouraging discussion board use (Zhang et al., 2017), grouping participants according to engagement metrics (Wichmann et al., 2016; Sanz-Martínez et al., 2019), and using tools with extra affordances (Su et al., 2016; Gamage and Whiting, 2021; Yang and He, 2022). This conclusion is supported across all categories of interaction, however, one-to-one peer interaction requires more research support.

Salomon and Perkins (1998), Vygotsky (1978) and Brown et al. (1989) proposed that the construction of knowledge is effectively built on the understanding of peers. This sentiment is evident in multiple studies selected for this review, including the one by Kellogg et al. (2014) where some co-construction of knowledge between participants was found and also the study by Mayende et al. (2017) where participants indicated that the groups helped them understand course content. Siemen's theory of Connectivism which emphasised tapping on information sources for learning was exemplified by Krasny et al. (2018) where participants who had previously completed the course were recruited to be group leaders. The group leaders fostered a supportive environment between peers, facilitated discussions and scaffolded participants' learning when course materials were challenging. The results from answering this second research question helps MOOC designers identify the learning outcomes that they wish to target and the existing ways of evaluating these outcomes.

There are some limitations to this review. Firstly, although researchers worked collaboratively to ensure consistency and agreement in the analyses of the selected studies, the selection was done by one of the researchers and no interrater reliability score was calculated. It is recommendable to have two researchers perform the selection and analysis of the studies in a review to avoid bias and ensure objectivity. Secondly, while there is strong evidence that social interaction is related to a multitude of favourable learning outcomes, it is arguable that there is insufficient evidence to conclude that there is a causal effect of social interaction on learning outcomes. Only about a quarter of the studies selected ($n=6$) in this review were experimental in research design. It is possible that participants who achieve favourable outcomes seek out social interactions or that there is an unknown third variable influencing both learning achievement and tendency to seek social interaction (DeVries et al., 2018). Ideas for future rese arch include experimental studies manipulating what individuals post on discussion forums and their frequency of posting to better discern their effects on learning outcomes. With learning groups, experimental research manipulating different activities in group settings can be done to identify which specific activities and group processes contribute most to learning outcomes. The results of those studies will build upon what has been discovered in this review, helping MOOC designers know what exactly to prompt learners to post and also how to proceed after creating an optimal grouping of participants.

Evidence also suggests that other factors mediate the relationship between social interaction and learning outcomes. These factors include learning group identification, immersive experience and satisfaction of relatedness needs (Fang et al., 2019). Previous research in a non-MOOC setting has shown evidence of social closeness affecting learning outcomes like satisfaction (Feng et al., 2022). With the categories of interaction in this study differing in degrees of closeness, perhaps it is worth investigating whether these different categories of interaction have differential impacts on the above-mentioned mediating factors. Recommendations for future research are to carry out comparative studies for the different categories of interaction and investigate if there are differences in feelings of group identification, immersiveness of a MOOC, and satisfaction of relatedness needs.

Finally, while this review show that interactions at different levels are effective, promoting all three might not be ideal as a participant could then be spending too much time on interactions. If a lack of time is cited as a common reason for dropout (Onah et al., 2014), then MOOC designers should be cautious not to overdo the promotion of interactions as this could lead to a diminishing return on beneficial learning outcomes. The suitable type of interaction to promote in a MOOC may depend on factors such as course length, course difficulty and course topics. For example, with a shorter course length, a discussion forum might suffice for satisfying relatedness needs and feelings of social immersion. Future research should look into understanding how these factors could potentially mediate the effects of different social interactions on learning outcomes. This is likely to help MOOC designers pick the best type of social interaction to promote for MOOCs with specific characteristics.

In conclusion, in a MOOC learning context, interactions among participants have been shown to have a variety of beneficial learning outcomes such as knowledge gained, learning and social engagement and learner experience. MOOC designers should consider strategies

to promote peer interactions at different levels, in discussion forums, in learning groups, or in facilitating one-to-one peer interactions. While actionable steps have been outlined in this review, more research can be done to optimise the proposed strategies to maximise the benefits of social interactions for participants in MOOCs.

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.

Author contributions

HL: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. GM: Writing – original draft, Writing – review & editing. CM: Writing – original draft, Writing – review & editing. MB: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

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

Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., and Alalwan, N. (2019). Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *J. Comput. High. Educ.* 32, 429–454. doi: 10.1007/s12528-019-09241-y

Alea, N., Jawitz, S., Adams, P., and Yang, P. (2023). The use and functions of students' personal stories in online discussion forums. *Int. J. Teach. Learn. Higher Educ.* 35, 101–112.

Al-Samarraie, H. (2019). A scoping review of videoconferencing systems in higher education. *Int. Rev. Res. Open and Distributed Learn.* 20, 121–140. doi: 10.19173/irrodl. v20i4.4037

Andersen, K., Thorsteinsson, E. S., Thorbergsson, H., and Gudmundsson, S. K. (2019). Using MOOCs to supplement reading courses: An instructor's view. *2019 IEEE learning with MOOCS (LWMOOCS)*, Milwaukee, WI, USA.

Anderson, V., Gifford, J., and Wildman, J. (2020). An evaluation of social learning and learner outcomes in a massive open online course (MOOC): A healthcare sector case study. *Hum. Resour. Dev. Int.* 23, 208–237. doi: 10.1080/13678868.2020.1721982

Azmat, M., and Ahmad, A. (2022). Lack of social interaction in online classes during COVID-19. *J. Materials and Environ. Sci.* 13, 185–196.

Baber, H. (2021). Social interaction and effectiveness of the online learning – A moderating role of maintaining social distance during the pandemic COVID-19. *Asian Educ. Develop. Stud.* 11, 159–171. doi: 10.1108/aeds-09-2020-0209

Bandura, A. (1978). Social learning theory. *Contemp. Sociol.* 7:84. doi: 10.2307/2065952

Barak, M., Watted, A., and Haick, H. (2016). Motivation to learn in massive open online courses: examining aspects of language and social engagement. *Comput. Educ.* 94, 49–60. doi: 10.1016/j.compedu.2015.11.010

Boggs, H., Forero-Hernandez, P., Laboissiere, M., and Neher, K. (2021). Scaling online education: Five lessons for colleges. McKinsey & Company. Available at: https://www. mckinsey.com/industries/education/our-insights/scaling-online-education-five-lessons-for-colleges

Bouchet, F. R., Labarthe, H., Bachelet, R., and Yacef, K. (2017). "Who wants to chat on a mooc? Lessons from a peer recommender system" in *Lecture notes in computer science*, eds. D. K. Carlos, J. Patrick, P-. S. Mar, T. S. Daniel and W. Su (US: Springer Science+Business Media), 150–159.

Brown, J. M., Collins, A. J., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educ. Res.* 18, 32–42. doi: 10.3102/0013189x018001032

Coffrin, C., Corrin, L., De Barba, P., and Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. *LAK '14: Learning analytics and knowledge conference 2014*, Indianapolis, Indiana, USA.

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach* (*2*). US: Sage publications.

D MoherLiberati, A., Tetzlaff, J., and Altman, D. G.PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

De Felice, S., De Vigliocco, G. C., and Hamilton, A. F. (2021). Social interaction is a catalyst for adult human learning in online contexts. *Curr. Biol.* 31, 4853–4859.e3. doi: 10.1016/j.cub.2021.08.045

Deng, R., Benckendorff, P., and Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Comput. Educ.* 129, 48–60. doi: 10.1016/j. compedu.2018.10.019

DeVries, J. M., Rathmann, K., and Gebhardt, M. (2018). How does social behavior relate to both grades and achievement scores? *Front. Psychol.* 9, 1–8. doi: 10.3389/ fpsyg.2018.00857

El Said, G. R. (2016). Understanding how learners use massive open online courses and why they drop out. *J. Educ. Comput. Res.* 55, 724–752. doi: 10.1177/073563 3116681302

Eriksson, T., Adawi, T., and Stöhr, C. (2016). "Time is the bottleneck": A qualitative study exploring why learners drop out of MOOCs. *J. Comput. High. Educ.* 29, 133–146. doi: 10.1007/s12528-016-9127-8

Fang, J., Tang, L., Yang, J., and Peng, M. (2019). Social interaction in MOOCs: the mediating effects of immersive experience and psychological needs satisfaction. *Telematics Inform.* 39, 75–91. doi: 10.1016/j.tele.2019.01.006

Feng, S., Qiu, S., Gibson, D. R., and Ifenthaler, D. (2022). "The effect of social closeness on perceived satisfaction of collaborative learning" in *Cognition and exploratory learning in the digital age*, Eds. I. Dirk, G. S. Demetrios and I. Pedro (Springer). 101–113.

Gamage, D. (2021) Scaffolding social presence in MOOCs. *CHI '21: CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan.

Gamage, D., and Whiting, E. M. (2021). Together we learn better: Leveraging communities of practice for MOOC learners. *Asian CHI Symposium 2021, virtual conference*.

Gameel, B. G. (2017). Learner satisfaction with massive open online courses. *Am. J. Dist. Educ.* 31, 98–111. doi: 10.1080/08923647.2017.1300462

Gao, F., Zhang, T., and Franklin, T. (2013). Designing asynchronous online discussion environments: recent progress and possible future directions. *Br. J. Educ. Technol.* 44, 469–483. doi: 10.1111/j.1467-8535.2012.01330

Gillani, N., and Eynon, R. (2014). Communication patterns in massively open online courses. *Internet High. Educ.* 23, 18–26. doi: 10.1016/j.iheduc.2014.05.004

Gomez-Zermeno, M. G., and De La Garza, L. Y. A. (2016). Research analysis on MOOC course dropout and retention rates. *Turkish Online J. Distance Educ.* 17, 3–14. doi: 10.17718/tojde.23429

Greene, J. A., Oswald, C., and Pomerantz, J. H. (2015). Predictors of retention and achievement in a massive open online course. *Am. Educ. Res. J.* 52, 925–955. doi: 10.3102/0002831215584621

Hew, K. F. (2014). Promoting engagement in online courses: what strategies can we learn from three highly rated MOOCs. *Br. J. Educ. Technol.* 47, 320–341. doi: 10.1111/ bjet.12235

Hew, K. F. (2015). Student perceptions of peer versus instructor facilitation of asynchronous online discussions: further findings from three cases. *Instr. Sci.* 43, 19–38. doi: 10.1007/s11251-014-9329-2

Hew, K. F., Hu, X., Qiao, C., and Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Comput. Educ.* 145:103724. doi: 10.1016/j.compedu.2019.103724

Holotescu, C., Grosseck, G., Cretu, I. V., and Antoanela, N. (2014). Integrating MOOCs in blended courses. The *10th eLearning and software for education conference - eLSE 2014*, Bucharest, Romania.

Hood, N., Littlejohn, A., and Milligan, C. (2015). Context counts: how learners' contexts influence learning in a MOOC. *Comput. Educ.* 91, 83–91. doi: 10.1016/j.compedu.2015.10.019

Ivanec, T. P. (2022). The lack of academic social interactions and students' learning difficulties during COVID-19 faculty lockdowns in Croatia: the mediating role of the perceived sense of life disruption caused by the pandemic and the adjustment to online studying. *Sociol. Sci.* 11:42. doi: 10.3390/socsci11020042

Joo, Y. H., So, H., and Kim, N. (2018). Examination of relationships among students' self-determination, technology acceptance, satisfaction, and continuance intention to use K-MOOCs. *Comput. Educ.* 122, 260–272. doi: 10.1016/j.compedu.2018.01.003

Jordan, K. (2015). Massive open online course completion rates revisited: assessment, length and attrition. *Int. Rev. Res. Open and Distributed Learn.* 16, 341–358. doi: 10.19173/irrodl.v16i3.2112

Jung, Y., and Lee, J. M. (2018). Learning engagement and persistence in massive open online courses (MOOCS). *Comput. Educ.* 122, 9–22. doi: 10.1016/j.compedu.2018.02.013

Kellogg, S., Booth, S., and Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *Int. Rev. Res. Open and Distributed Learn.* 15, 263–289. doi: 10.19173/irrodl.v15i5.1852

Krasny, M. E., DuBois, B., Adameit, M., Atiogbe, R., Baih, L., Bold-Erdene, T., et al. (2018). Small groups in a social learning MOOC (slMOOC): strategies for fostering learning and knowledge creation. *Online Learning* 22, 119–139. doi: 10.24059/olj.v22i2.1339

Lieu, E., Cole, J., and Watkins, C. (2022). Bring something to the potluck: A system for inclusive and reciprocal online discussion. In *35th International BCS Human-Computer Interaction Conference* (pp. 1–6), Newcastle, UK.

Liyanagunawardena, T. R., Adams, A. B., and Williams, S. (2013). MOOCs: A systematic study of the published literature 2008-2012. *Int. Rev. Res. Open and Distributed Learn.* 14:202. doi: 10.19173/irrodl.v14i3.1455

Manli, L. (2014). MOOCs and college English teaching. 2014 *2nd international conference on advances in social science, humanities, and management*, Guangzhou, China.

Mayende, G., Prinz, A., Isabwe, G. M. N., and Muyinda, P. B. (2017). Learning groups in MOOCs: lessons for online learning in higher education. *Int. J. Engineer. Pedagogy (iJEP).* 7, 109–124. doi: 10.3991/ijep.v7i2.6925

Narayan, R., Rodriques, C., Araujo, J., Shaqlaih, A., and Moss, G. (2013). "Constructivism—constructivist learning theory" in *The handbook of educational theories.* eds. J. I. Beverly, B. Genevieve, L.–A. Rafael and J. Shirley (Charlotte, North Carolina, United States of America: Information Age Publishing, Inc.), 169–183.

Onah, D. F. O., Pang, E. L. L., Sinclair, J., and Uhomoibhi, J. (2018). "Learning analytics for motivating self-regulated learning and fostering the improvement of digital MOOC resources" in *Advances in intelligent systems and computing.* eds. E. A. Michael and T. Thrasyvoulos (US: Springer Nature), 14–21.

Onah, D. F. O., Sinclair, J., and Boyatt, R. (2014). Dropout rates of massive open online courses: behavioural patterns. *6th international conference on education and new learning technologies*, Barcelona, Spain.

Pérez-Sanagustín, M., Hilliger, I., Alario-Hoyos, C., Kloos, C. D., and Rayyan, S. (2017). H-MOOC framework: reusing MOOCs for hybrid education. *J. Comput. High. Educ.* 29, 47–64. doi: 10.1007/s12528-017-9133-5

Piaget, J. (1964). Cognitive development in children: development and learning. *J. Res. Sci. Teach.* 2, 176–186. doi: 10.1002/tea.3660020306

Pin-Ju, C., and Chen, Y. (2022). Massive open online course study group: interaction patterns in face-to-face and online (Facebook) discussions. *Front. Psychol.* 12, 1–15. doi: 10.3389/fpsyg.2021.670533

Ramesh, A., Goldwasser, D., Huang, B., Daumé, H., and Getoor, L. (2014). Uncovering hidden engagement patterns for predicting learner performance in MOOCs. *L@S 2014: First (2014) ACM Conference on Learning @ Scale*, Atlanta, Georgia, USA.

Razmerita, L., Kirchner, K., Hockerts, K., and Tan, C. (2020). Modeling collaborative intentions and behavior in digital environments: the case of a massive open online course (MOOC). *Acad. Manag. Learn. Educ.* 19, 469–502. doi: 10.5465/amle.2018.0056

Reich, J., and Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science* 363, 130–131. doi: 10.1126/science.aav7958

Salomon, G., and Perkins, D. L. (1998). Individual and social aspects of learning. *Rev. Res. Educ.* 23:1. doi: 10.2307/1167286

Sanz-Martínez, L., Er, E., Martínez-Monés, A., Dimitriadis, Y., and Bote-Lorenzo, M. L. (2019). Creating collaborative groups in a MOOC: A homogeneous engagement grouping approach. *Behav. Inform. Technol.* 38, 1107–1121. doi: 10.1080/0144929x.2019.1571109

Shapiro, H., Lee, C. N., Roth, N. W., Li, K., Etinkaya-Rundel, M., and Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: an examination of attitudes, motivations, and barriers. *Comput. Educ.* 110, 35–50. doi: 10.1016/j.compedu.2017.03.003

Shi, L., Cristea, A. I., Toda, A. M., and Oliveira, W. (2019). Social engagement versus learning engagement an exploratory study of future learn learners. *2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Dalian, China.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *Int. J. Instructional Technol. Distance Learn.* 2, 3–10.

Su, Y., Huang, C. S. J., and Ding, T. (2016). Examining the effects of MOOCs learners' social searching results on learning behaviors and learning outcomes. *Eurasia J. Mathematics, Sci. Technol. Educ.* 12, 2517–2529. doi: 10.12973/eurasia.2016.1282a

Sunar, A. S., White, S., Abdullah, N. H., and Davis, H. C. (2017). How learners' interactions sustain engagement: A MOOC case study. *IEEE Trans. Learn. Technol.* 10, 475–487. doi: 10.1109/tlt.2016.2633268

Tullis, J. G., and Goldstone, R. L. (2020). Why does peer instruction benefit student learning? *Cogn. Res.: Principles and Implications* 5:15. doi: 10.1186/s41235-020-00218-5

University of Technology Sydney. (2023). Global education statistics 2023. Retrieved October 6, 2023, from https://studyonline.uts.edu.au/blog/education-statistics

Urrutia, M. L., Cano, E. V., and Meneses, E. L. (2017). MOOC learning analytics using real-time dynamic metrics. *@Tic: Revista D'Innovació Educativa.* 38–47. doi: 10.7203/attic.18.10022

Veletsianos, G., Collier, A., and Schneider, E. (2015). Digging deeper into learners' experiences in MOOCs: participation in social networks outside of MOOCs, notetaking and contexts surrounding content consumption. *Br. J. Educ. Technol.* 46, 570–587. doi: 10.1111/bjet.12297

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Available at: https://ci.nii.ac.jp/ncid/BA03570814

Wang, Y., and Baker, R. S. (2015). Content or platform: why do students complete MOOCs? *MERLOT J. Online Learn. Teach.* 11, 17–30.

Wang, W., Guo, L., He, L., and Wu, Y. J. (2018). Effects of social-interactive engagement on the dropout ratio in online learning: insights from MOOC. *Behav. Inform. Technol.* 38, 621–636. doi: 10.1080/0144929x.2018.1549595

Wichmann, A., Hecking, T., Elson, M., Christmann, N., Herrmann, T., and Hoppe, H. U. (2016). Group formation for small-group learning. *OpenSym '16: The international symposium on open collaboration*, Berlin, Germany.

Williams, E. A., Zwolak, J. P., Dou, R., and Brewe, E. (2019). Linking engagement and performance: the social network analysis perspective. *Phys. Rev.* 15, 020150-1–020150-15. doi: 10.1103/physrevphyseducres.15.020150

Wintrup, J., Wakefield, K., and Davis, H. (2015). *Engaged learning in MOOCs: A study using the UK engagement survey*. York: Higher Education Academy.

Wut, T., and Xu, J. (2021). Person-to-person interactions in online classroom settings under the impact of COVID-19: a social presence theory perspective. *Asia Pac. Educ. Rev.* 22, 371–383. doi: 10.1007/s12564-021-09673-1

Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., and Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. DOAJ (DOAJ: Directory of Open Access Journals). Available at: https://doaj.org/article/4d80003de7b44a29b81aede0ab5661dd

Xu, Y., Gitinabard, N., Lynch, C. F., and Barnes, T. (2019). What you say is relevant to how you make friends: measuring the effect of content on social connection. In Educational data mining. Available at: https://dblp.uni-trier.de/db/conf/edm/edm2019.html#XuGLB19

Yang, B., and He, H. (2022). Learning alone yet together: Enhancing between-learner social connectivity at scale. *L@S '22: Ninth (2022) ACM Conference on Learning @ Scale*, New York City, NY, USA.

Zhang, D. J., Allon, G., and Van Mieghem, J. A. (2017). Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses. *Manuf. Serv. Oper. Manag.* 19, 347–367. doi: 10.1287/msom.2016.0615

Zhang, Q., Peck, K. L., Hristova, A., Jablokow, K. W., Hoffman, V., Park, E., et al. (2016). Exploring the communication preferences of MOOC learners and the value of preference-based groups: is grouping enough? *Educ. Technol. Res. Dev.* 64, 809–837. doi: 10.1007/s11423-016-9439-4

*CORRESPONDENCE
Wei Chen
✉ bluerwei@163.com

The impact of flipped classroom on English proficiency of first-year Chinese urban and rural pre-service teachers

Wei Chen*

School of Foreign Languages, Hainan Normal University, Haikou, China

Introduction: In recent years, China has made strides in adopting student-oriented teaching approaches, particularly in tertiary English education, through the integration of enhanced technology. This study aimed to investigate the impact of flipped classroom on the English proficiency of first-year pre-service teachers at a Chinese normal university. It also sought to determine whether educational background (urban or rural) interacted with the teaching approach (flipped or traditional) in affecting the language proficiency of the learners.

Methods: A quasi-experimental design was utilized with two treatments: a flipped classroom approach and a traditional teacher-centered teaching approach. Both approaches were implemented in the Integrated English Course over a 12-week semester. Two randomly selected classes, consisting of 60 pre-service teachers in each class, were assigned to either the experimental or control group. Data were collected from pretest and post-test assessments and analyzed using two-way ANOVA.

Results: The results revealed a positive impact of the flipped classroom and a significant interaction between educational background and teaching approach on English proficiency. Specifically, urban pre-service teachers achieved higher English proficiency than their rural peers when taught in the flipped classroom, but not in the traditional classroom. Furthermore, urban pre-service teachers in the flipped classroom outperformed their urban peers in the traditional classroom, while rural pre-service teachers did not show any significant difference in their performance between the two classes.

Discussion: The findings suggested that the flipped classroom approach was more effective than the traditional approach for Chinese pre-service teachers, particularly those with an urban educational background. However, it is important to ensure that rural learners receive sufficient support to benefit equally from this innovative teaching approach. Accordingly, implications and recommendations for future research are discussed.

KEYWORDS

flipped classroom, English language teaching, pre-service teachers, English proficiency, educational background

1 Introduction

Learning English as a foreign language (EFL) is a challenging task for students in a dominant first language setting (Abdullah et al., 2019). In China, tertiary EFL learners are confronted with several difficulties that hamper their language achievements (Xiu, 2021).

Among them, the traditional teaching approach, which prioritizes teacher-centered instruction and content regurgitation, is often criticized for contributing to the difficulties encountered by Chinese learners (Wang, 2020). This conventional approach is still commonly employed by language educators in Chinese universities, impeding students' acquisition of English knowledge, skills, and associations they develop with their English learning (Yang and Chen, 2020). Moreover, educational backgrounds can significantly impact upon Chinese EFL learners. Due to conspicuous regional disparities in development between urban and rural areas of China (Wu et al., 2021), students from rural areas face more challenges in EFL learning than their urban peers (Liu and Wang, 2021). Hence, supporting Chinese students of diverse educational backgrounds with potentially more effective teaching approaches seems to be needed.

In the last decade, the advert of technology and its application have brought about tremendous changes in EFL teaching and learning (Haghi, 2021). Quite a few language educators worldwide have adopted gradual steps to integrate technologies into their teaching approaches (Basal, 2015). As an active-learning, student-centered, and flexible pedagogical approach incorporating digital technology, flipped classroom has gained growing prominence, especially at the tertiary level (Kurt, 2017). Its increasing popularity is attributed to research suggesting that flipped classroom may contribute to positive language learning outcomes (Sukerti et al., 2020; Gok et al., 2021; Pang, 2022; Liu et al., 2023). Nevertheless, critical evaluations concerning its effectiveness and contextual applicability has also been raised (Alhamami and Costello, 2019; Lakarnchua et al., 2020; Altas and Enisa, 2021; Rajabi et al., 2021).

Currently, the implementation of flipped classroom in EFL education in China has revolutionized the traditional teaching model and invigorated modern classroom teaching (Yang and Chen, 2020). In response to the Chinese higher education informatization reform, an increasing number of front-line teachers and researchers are adopting the flipped classroom approach to promote English learning (Wang et al., 2021). While several studies have identified the benefits of flipped EFL classroom in the Chinese context, its effectiveness is still subject to debate (Zhang and Guo, 2018; Chen, 2019). Some critics argue that this approach may lead to unequal language achievements among urban and rural students, emphasizing the need to consider students' educational background as a context-specific factor (Lv and Wang, 2016). So far, the use of flipped classroom in EFL teacher education in China is still under researched. Therefore, this study aimed to fill the gap by investigating the implementation and impact of flipped classroom on English proficiency of Chinese pre-service teachers while considering the potential effect of their educational backgrounds. It answered the following the research questions.

1) Is there a significant difference in English proficiency between first-year Chinese pre-service teachers exposed to the flipped classroom and those engaged in the traditional teacher-centered instruction?
2) Does educational background (urban or rural) significantly interact with teaching approach (flipped or traditional) in influencing English proficiency among the pre-service teachers?

Accordingly, two null hypotheses were tested.

H01: There is no significant difference in English proficiency between first-year Chinese pre-service teachers exposed to the flipped classroom and those engaged in the traditional teacher-centered instruction.

H02: There is no significant interaction between educational background and teaching approach in terms of English proficiency of the pre-service teachers.

2 Literature review

2.1 EFL learning and influencing factors

EFL learning refers to the systematic acquisition of English language proficiency through formal instruction, typically in a language classroom setting. In many contexts where English is not the dominant native language, such as Russia, Egypt, Thailand, Japan, and China, EFL learners have minimal exposure to English outside of the classroom (Jenkins, 2003). As a result, they face significant challenges in achieving proficiency in English. To overcome these challenges, it is crucial for both educators and learners to recognize and understand potential factors influencing EFL learning. Research has pinpointed a range of internal factors, spanning personality, motivation, cognitive style, attitude, self-confidence, anxiety, and previous educational backgrounds (Pariyanto and Pradipta, 2020). Furthermore, external factors, including teaching approaches, learning materials, environment, and social elements such as peer groups, play a pivotal role in shaping the EFL learning experience (Getie, 2020). While these factors appear intricately intertwined and jointly impactful in EFL learning, some researchers emphasize the significance of specific learner and teacher factors, such as previous educational background and teaching approach (Cuong, 2021; Ömer and Akçayoğlu, 2021). This study, therefore, sought to investigate the possible influences of these factors on the EFL learning of pre-service teachers in China and propose implications accordingly.

2.2 Educational background and urban-rural disparities in EFL education

Educational background encompasses both formal and informal education, as well as any ongoing learning pursuits (Glassdoor Team, 2021). It not only reflects geographic features, but also embodies historical, cultural, and educational characteristics of a local area, exerting an impact on individual subjects (Evans and Savage, 2015). In this study, educational background, as a contextually specific factor, was defined as an urban or rural area where a student grew up and attended his/her secondary school, categorized according to China's national statistics on urban and rural districts (National Bureau of Statistics, 2020).

Previous research has revealed that EFL learners with rural educational backgrounds tend to possess lower linguistic attainment, inadequate learning motivations, and higher language anxiety compared to their urban peers (Cuong, 2021). To address these disparities, some researchers argue that language teachers should

identify students' educational backgrounds and modify teaching methodologies accordingly (Deepa, 2021), while also leveraging technology as an educational equalizer (Li et al., 2018).

In China, urban–rural inequality in education has been a persistent problem across all levels of education (Rural Education Action Program, 2018; Wu and Zhang, 2021), primarily resulting from the structural duality in social and economic systems (Feng et al., 2019). In recent years, the Chinese government has made significant efforts in scaling up education in rural areas (Yue et al., 2018), reflecting its commitment to achieving educational equality. In furtherance of this goal, the Communist Party of China Central Committee and The State Council (2018) released the National Rural Vitalization Strategic Plan (2018–2022), with a focus on promoting education in rural areas. Moreover, the Chinese Ministry of Education (2012) issued "Plan for ICT in Education 2011–2020," advocating adoption of information technology and technology-assisted teaching approaches. Despite progress in alignment with these policies, disparities persist, particularly in the domain of EFL education (Liu and Wang, 2021). Previous studies have shown that rural students achieve significantly lower English scores in standardized tests compared to their urban peers (Ke, 2016), and face challenges and problems in communication (Liu, 2021), foreign language anxiety (Chen, 2020), classroom participation (Ou, 2017), learning motivation (Liu and Wang, 2021), digital literacy, and use of cognitive, meta-cognitive and emotional regulation strategies (Zhang, 2022). It seems that much remains to be done to bridge the regional gaps in EFL education and the implementation of innovative teaching approaches tailored to the specific needs of Chinese EFL learners from different educational backgrounds requires further exploration.

2.3 Flipped classroom

Flipped classroom was first defined by Lage et al. (2000) as an instructional approach in which "events that have traditionally taken place inside the classroom now take place outside the classroom and vice versa" (pp. 30–43). This broad definition is a starting point to provide teachers freedom and flexibility to adopt the flipped classroom with various methods (Eppard and Rochdi, 2017). However, the broad definition is later on restricted by some researchers who limit the concept of flipped classroom by inclusion of specific methods like instructional videos outside of class as well as interactive group learning inside of the class (Bishop and Verleger, 2013). The narrow definition suggests that information transmission occurs out of the classroom, while the assimilation takes place in the classroom (Talbert, 2012).

The flipped classroom is a promising instructional approach with theoretical underpinnings in both cognitivism and constructivism. Cognitivism supports the flipped classroom through a process that involves students in lower levels of cognitive work prior to class, such as recalling and understanding concepts, and engages them in higher levels of cognitive activities to apply, analyze, synthesize, and evaluate during in-class time (Bloom, 1984). The revised Bloom's Taxonomy is even more relevant to flipped classroom, as it shows a dynamic cognitive process from information transmission, occurring independently outside of class, to information assimilation, requiring more critical reasoning under the guidance of a teacher during class and after class (Talbert, 2012). Constructivist paradigms, including

cognitive and social constructivism, are also key theoretical foundations for the flipped classroom. Teachers in flipped classes need to adjust students' learning process with their cognitive stages, facilitate problem-solving skills in an active learning environment, and encourage collaborative tasks that help students construct knowledge through social interaction. In constructivist flipped classes, learners become active participants and information constructors, building up their subjective representations of knowledge based on prior experience (Bergmann and Sams, 2012a). To facilitate students' knowledge construction, teachers should organize interactive discussions, create active learning environments based on questioning, and offer immediate feedback and differentiated support to scaffold students' flipped learning.

2.4 Impact of flipped classroom on English proficiency

Flipped classroom is a teaching approach that utilizes digital technologies to facilitate independent learning outside of the classroom, and interactive learning activities inside the classroom (Bergmann and Sams, 2012b). This technology-assisted teaching approach has been reported to produce a number of positive outcomes in EFL learning at various levels (Strelan et al., 2020), such as improved language proficiency (Sukerti et al., 2020), ameliorated foreign language anxiety (Gok et al., 2021), promoted critical thinking skills (Pang, 2022), boosted self-confidence, motivation and autonomy (Ghufron and Nurdianingsih, 2019), and enhanced learner engagement and satisfaction (Wu and Zhang, 2021). Among these positive outcomes, improved English proficiency has been underlined by many researchers (Roth and Suppasetseree, 2016; Loucky, 2017; Kırmızıa and Kömeç, 2019; Sukerti et al., 2020). For example, Kırmızıa and Kömeç (2019) found that Turkish high school students who engaged in flipped learning achieved significantly higher scores in English vocabulary quizzes than those taught by the traditional method. Similarly, Zhang et al. (2016) discovered that first-year English majors in a Chinese university were able to produce more language output, actively solved their own problems in a more engaging way, and increased their interest in language learning through the use of the flipped classroom approach.

However, the effectiveness of the flipped classroom on improving English proficiency is still open to debate. Some studies have found no significant difference in language proficiency of students taught by flipped and traditional teaching approaches (Suranakkharin, 2017; Yang, 2017; Alhamami and Khan, 2019), and there are also concerns about the potential impact of educational backgrounds on the effectiveness of this model (Lv and Wang, 2016; Liu and Wang, 2021). For instance, Lv and Wang (2016) found that rural Chinese students with less digital skills had higher levels of language anxiety and obtained lower scores in flipped classes compared to their urban peers. Therefore, while the flipped classroom shows promise in improving EFL education outcomes, further research is needed to explore its effectiveness in different contexts, taking into account the influence of cultural and educational backgrounds on language learning outcomes.

Despite the growing interest in implementing the flipped classroom in EFL classes, limited research has been conducted on its effectiveness in EFL teacher education, and there have been mixed results (García-Sánchez and Santos-Espino, 2017; Karaaslan and

Çelebi, 2017). One study by García-Sánchez and Santos-Espino (2017) found that the flipped classroom approach empowered Spanish pre-service EFL teachers to improve their academic performance, while other studies by Adnan (2017) and Cabi (2018) found no significant difference in English learning outcomes between flipped and non-flipped classes. It seems that more research conducted in different contexts is needed to clarify the effectiveness of the flipped classroom on pre-service EFL teachers. Hence, this study may add to existing literature by providing valuable empirical evidence and instructive insights into the feasibility of this teaching approach in a specific Chinese context.

3 Methodology

3.1 Course context

The utilization of the flipped classroom was examined in Integrated English Course (IEC), which is a mandatory course for undergraduate English majors in China. According to the Teaching Guide for Undergraduate English Majors issued by the Ministry of Education of China in 2020, IEC is not solely a language course developed to equip students with basic English knowledge and skills to enhance their language proficiency (Jiang, 2019), but also a competence-building course aimed at promoting students' lifelong learning skills such as problem-solving, communication, and collaboration (Qu, 2017; Wen, 2019). This course is offered to undergraduate English majors for the first two years (semester 1–4) of their university study. There are two classes per week, each lasting one and a half hours.

3.2 Research design

This research utilized a quasi-experiment with a 2×2 non-equivalent control group design to investigate whether the flipped classroom approach or traditional teacher-centered teaching approach had a significant impact on the English proficiency of first-year Chinese pre-service teachers with urban and rural educational backgrounds.

3.3 Participants

Participants involved in the study were 120 students from the Faculty of English Language Teaching at a normal university in China. All the participants were pre-service teachers in their first year of undergraduate studies and attending Integrated English Course over a 12-week semester in academic year 2022.

Before enrollment, these students, aged from 18 to 20, had learned English for at least 9 years in urban or rural schools, and most of them had no prior flipped classroom experience. They were randomly picked from nine classes and assigned to an experimental group and a control group. Each group comprised 60 pre-service EFL teachers (30 urban and 30 rural students). The male-to-female student ratio was approximately 1:3 in every group.

First-year pre-service teachers were chosen as the target participants due to their possible need of guidance and help as freshmen in their university study as well as positive outcomes the

flipped classroom might bring to the cohort in their subsequent tertiary EFL learning and in their future career as EFL teachers.

Two carefully selected instructors were invited to teach in the experimental and control groups. Both were females, close in age and in working years, and similar in their educational backgrounds. The instructor assigned to the flipped classroom had four-year experience with it. Despite this, she still underwent a training session to familiarize with specific procedures required for the flipped course instruction.

3.4 Intervention

Two teaching approaches were implemented by the two language teachers, respectively. The control group received traditional teacher-centered teaching, which was the typical style of teaching in the EFL classroom. This approach focused more on the teacher than on the students, and the majority of class time was spent helping students remember and understand key concepts. In contrast, the experimental group was taught using the flipped classroom approach, which emphasized the importance of students being active in their technology-assisted learning supported and guided by the teacher.

In the traditional classroom, the teaching process followed the normal convention. Before class, students were asked to review textbook materials while the instructor prepared PowerPoint presentations. In class, new topics were explained in detail using lecture slides, supplemented with exercises to enhance comprehension. Occasionally, students were invited to share their work and opinions. At the end of class, the instructor summarized the topic, assigned homework, and shared additional materials online for self-study after class. If needed, students could seek assistance from the instructor outside of class if needed.

Common activities performed in the control group included teacher-led lectures, fast-reading, pair-work, group discussions and presentations. Lectures, consuming a a substantial amount of class time, utilized multimedia tools for content delivery. Fast-reading involved independent skimming, scanning, pair discussions, and class sharing. Group discussions and presentations emphasized collaboration, requiring students to collectively apply knowledge. These activities targeted integrated language skills, focusing on the cognitive domain. Due to lecture-focused time constraints, higher-order thinking activities were often conducted after class.

The flipped classroom was implemented in the following steps. Before each class, students were tasked with independently mastering key concepts at their own pace and completing a pre-class quiz. They could also post questions on an online forum within a Learning Management System (LMS) or via WeChat or QQ online study groups. The teacher then collated students' results, selected key questions to be addressed during class, and designed relevant activities to help students apply what they had learned. The face-to-face class began with a brief lecture, followed by feedback from students using a voting system in the LMS. Individual and group activities were then conducted, with the teacher monitoring performance and providing guidance as needed. Both types of activities aimed to promote positive interactions among students and with the teacher. At the end of class, the teacher summarized the topic and unresolved issues, and students reflected on their learning and provided suggestions for future activities. After class, students continued to apply their knowledge by

completing further exercises and engaging in discussions on the topic. The teacher offered ongoing guidance and adjusted lesson designs based on student feedback.

In the flipped classroom, a variety of individual activities (such as polling, individual problem-solving, brainstorming, crowd-sourcing, and individual presentations) and group-based activities (such as three-step-interviews, cumulative brainstorming, think-pair-share, and group presentations) were carefully designed and carried out to enhance learning outcomes progressively. For instance, the three-step-interview required students to work in pairs, interviewing each other about their gains and questions, encouraging independent or collaborative information seeking and problem-solving. This relatively low-difficulty activity aimed to help them remember and gain a better understanding of specific learning content. Following that, the think-pair-share with a certain level of complexity could be conducted to promote higher-level cognitive development. This activity required students to work independently with guidance from the instructor or collaboratively with their peers to gather and synthesize key points, create a concept map, and present their work in class. The activity not only aimed to help students better organize their knowledge and illustrate complex conceptual associations but also intended to improve their language skills, enhance their communication abilities, spark their interest in learning, and foster positive interpersonal relationships among classmates. Undoubtedly, various other activity options were utilized in the flipped classroom. These activities, used individually or in combination for each lesson, were tailored to the expectations and evolving needs of the pre-service teachers in the specific context.

In the experimental group, various technological tools were adopted to facilitate the flipped teaching and learning process. Common hardware like laptops, tablets, and smartphones supported the entire process, while public computers in the library were available for pre-service teachers with limited device access. Multimedia facilities, including interactive whiteboards, projectors, computers, and video recorders, were used for face-to-face instruction. During the preparation phase, software such as search engines, video editing tools, and text editing tools was employed. The LMS played a central role in content delivery, progress monitoring, and communication. Additionally, social media apps like WeChat and QQ complemented the LMS, serving as alternative platforms for material sharing, pre-class discussions, and feedback collection. These tools were strategically utilized throughout the in-class and post-class phases, effectively guiding the flipped teaching and learning.

For ethical considerations, teaching content for both groups was identical, and all resources provided to the experimental group were made available to the control group after the end of the study. Furthermore, the same intervention was implemented with the control group afterward.

3.5 Instrument and procedures

The instrument used in the study was two IELTS (International English Test System) tests selected from the IELTS Student's Book published by Cambridge University (Cambridge, 2019). The two papers were modified and utilized as the pretest and the post-test.

The proficiency tests were chosen due to their widely recognized status for reliability and validity. The reliability, measured through Cronbach's Alpha, averaged 0.88 for reading and listening modules (University of Cambridge Local Examinations Syndicate, 2007). While speaking and writing modules lacked reported reliability coefficients, their dependability was ensured through examiner training and standardization (Fazel and Ahmadi, 2011). The IELTS homepage attested to excellent internal consistency (coefficient: 0.95, SEM: 0.21) across all modules (Hashemi and Daneshfar, 2018). Past studies also affirmed predictive and construct validity, with IELTS tests predicting academic language performance (Ingram and Bayliss, 2007) and correlating positively with GPA (Yen and Kuzma, 2009). Despite construct validity requiring further exploration, satisfactory levels were reported across reading, listening, writing, and speaking modules, affirming precision of IELTS tests in reflecting language abilities (Quaid, 2018; Noor, 2020).

Moreover, the IELTS tests, specifically the General Training version, were selected because these tests closely aligned with the expected learning outcomes of the IEC in assessing language abilities. The adoption of the General Training version was based on its suitability for evaluating the language proficiency of participants, who were expected to enhance and demonstrate their English skills in practical daily contexts.

In terms of test components, each test paper consisted of four modules: listening, reading, writing, and speaking. The listening module included a recorded monolog and a daily conversation with 20 questions in total. This module measured students' ability to understand opinions, purposes, and attitudes of speakers as well as ability to follow detailed information and idea development. The reading module contained two sections (20 questions) with factual texts on general topics. The reading module, focusing on real-world situations, evaluated students' ability to understand main ideas, details, and implied meanings as well as ability to recognize authors' arguments, attitudes, and purposes. The writing module was about a letter writing task requiring students to write at least 150 words. This part examined whether students were capable of engaging in personal correspondence to elicit information, express, and justify opinions, outline a problem, and present related solutions. And the speaking module, comprised of general questions about some familiar topics, a short presentation on a given topic, and a two-way discussion, assessed a wide range of speaking skills, including organizing ideas, justifying views, and analyzing and discussing issues. Each of the four test modules accounted for a quarter (25 points) of the total score (100 points). The first three modules, including listening, reading, and writing, lasted one and a half hours (30 min for each module) without a break. And the speaking module, given on the same day, lasted for 6–8 min for each student. The performance of students was evaluated by two examiners who had received standardized IELTS assessment training as scorers from a British Council's IELTS test center in China. Using answer keys and band descriptors as reference points, these examiners provided scores for all four modules to all 120 participants. To further ensure the instrument's validity and reliability, a pilot study was done to assess content validity, internal consistency, and inter-rater reliability of IELTS speaking and writing modules using analysis of intra-class correlation coefficients (ICC). A panel of five experts in English language education and assessment conducted content validation. They reviewed and scored each item, ensuring relevance and representation. Results showed high agreement (95%), confirming

content validity. For internal consistency, another group of first-year pre-service teachers ($n = 38$) took an IELTS test, yielding a Cronbach's alpha of 0.702, indicating acceptable consistency. Inter-rater reliability for subjective scoring in speaking and writing modules was assessed, resulting in excellent ICC values of 0.92 and 0.98, meeting established guidelines (Koo and Li, 2016). This comprehensive analysis ensured the instrument's reliability and validity.

Before the intervention, a pretest was administered to both groups. Participants in the experimental group underwent a one-week training session, introducing them to components, requirements, procedures, activities, and potential benefits of the flipped classroom, given their lack of previous experience in this approach. Subsequently, they engaged in a two-week practice of the flipped classroom to become familiar with expectations both in and out of class. Following this, both groups were taught the same content from the English course, and the intervention lasted for 8 weeks. At the end of the treatment, a post-test was given to both groups. Test papers were collected, and data were analyzed to compare the English proficiency of the pre-service teachers in the groups. Moreover, during the pretest and post-test, demographic information about students' class and educational background was collected for subsequent data analysis.

4 Results

All the data were analyzed quantitatively using the Statistical Packages for the Social Science (version 26).

First, a test of homogeneity of variances was conducted to examine whether all the subgroups were homogeneous in terms of their English proficiency at the beginning of the intervention. The assumption of homogeneity across all levels of the independent variables was met, indicated by the Levene's test of equality of error variances, $F(3, 116) = 2.61$, $p = 0.055$ for the IELTS pretest score.

After confirmation of homogeneity of groups, assumption testing of the two-way ANOVA was followed. An inspection of the assumption of univariate normality demonstrated that the dependent variable was normally distributed for all the sub-groups ($ps > 0.05$), supporting the assumption of normality. There was just one mild univariate outlier and no extreme outliers for the dependent variable indicated by the boxplots. The assumption of homogeneity of variances was tenable, indicated by the Levene's test of equality of error variances, $F(3, 116) = 1.87$, $p = 0.138$ for the scores. As all the assumptions were met, an analysis was conducted.

Table 1 reports the descriptive statistics of the IELTS post-test scores of the flipped classroom (FC) and traditional teacher-centered teaching classroom (TT), including the mean, standard deviation of values for each combination of the groups of the independent variables, as well as the number of participants for each cell.

The results of the ANOVA (see Table 2) showed a significant main effect of teaching approach on English proficiency, $F(1,116) = 29.41$, $p < 0.001$, $\eta_p^2 = 0.20$. Based on Cohen's (1988) benchmarks, the effect size was large. It indicated that teaching approach explained 20% of the variance in the dependent variable. The pre-service teachers in the FC ($M = 64.80$, SD = 6.97) scored significantly higher on the test than those in the TT ($M = 58.10$, SD = 7.65). Besides, there was a significant main effect of educational background on English proficiency, $F(1,116) = 12.02$, $p = 0.001$, $\eta_p^2 = 0.09$. Educational background accounted for 9% of the variance in the English proficiency. The urban

TABLE 1 Descriptive statistics.

Teaching approach (TA)	Educational background (EB)	Mean	Std. deviation	N
TT	Rural	57.90	7.81	30
	Urban	58.30	7.61	30
	Total	58.10	7.65	60
FC	Rural	60.72	5.32	30
	Urban	68.88	6.00	30
	Total	64.80	6.97	60
Total	Rural	59.31	6.77	60
	Urban	63.59	8.64	60
	Total	61.45	8.02	120

pre-service teachers ($M = 63.59$, SD = 8.64) scored significantly higher than the rural ones ($M = 59.31$, SD = 6.77). Additionally, there was a significant interaction effect between teaching approach and educational background on the English proficiency, $F(1,116) = 9.88$, $p = 0.002$, $\eta_p^2 = 0.08$, indicating that the effect of educational background on English proficiency were different in the flipped and traditional groups. The interaction explained 8% of the variance.

Since the interaction was significant (Figure 1), simple effect analysis was performed. Referring to the pairwise differences based on estimated marginal means, urban pre-service teachers in the FC achieved significantly higher English proficiency than their urban peers in the TT (MD = 10.58, $p < 0.001$). For rural pre-service teachers, no significant difference was observed between the two groups (MD = 2.82, $p = 0.110$). In the FC, the urban pre-service teachers significantly outperformed their rural classmates (MD = 8.17, $p < 0.001$). However, in the TT, no significant difference was found in the English proficiency of urban and rural pre-service teachers (MD = 0.40, $p = 0.819$). Overall, the pre-service teachers who received the combined effect of flipped classroom and urban educational background presented substantially higher English proficiency than those under other conditions.

5 Discussion

This study aimed to investigate the impact of flipped classroom on English proficiency of first-year Chinese pre-service teachers. Additionally, it sought to determine whether educational background interacted with teaching approach in affecting the language proficiency of the learners. In an attempt to reach the research objectives, the quantitative data were gathered with the IELTS tests. The results indicated that there was a statistically significant difference in English proficiency among the pre-service teachers, with the flipped group outperforming the traditional group. The first null hypothesis was rejected. The results also revealed a significant interaction between educational background and teaching approach in terms of the pre-service teachers' English proficiency. The second null hypothesis was rejected.

Contrary to the first null hypothesis, the study findings supported the positive impact of the flipped classroom on English proficiency, which was consistent with some previous studies (Arslan, 2020; He,

TABLE 2 Tests of between-subjects effects.

Dependent variable: IELTS post-test						
Source	Type III sum of squares	Degree of freedom	Mean square	F	Sig.	Partial η^2
Corrected model	2349.52[a]	3	783.17	17.11	0.000	0.31
Intercept	453132.30	1	453132.30	9896.96	0.000	0.99
TA	1346.70	1	1346.70	29.41	0.000	0.20
EB	550.41	1	550.41	12.02	0.001	0.09
TA * EB	452.41	1	452.41	9.88	0.002	0.08
Error	5311.06	116	45.79			
Total	460792.88	120				
Corrected total	7660.58	119				

[a]$R^2 = 0.307$ (Adjusted $R^2 = 0.289$).

FIGURE 1
Interaction plot for IELTS post-test.
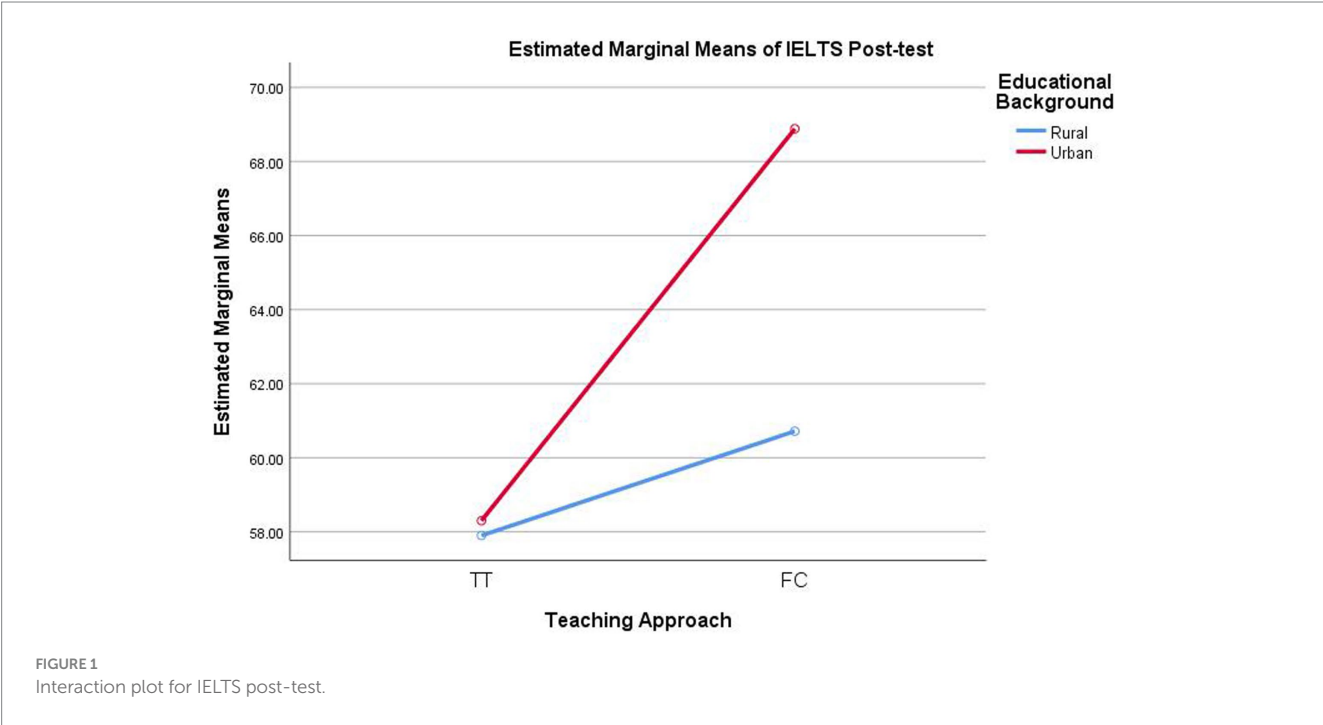
2020; Turan and Akdag-Cimen, 2020; Kusuma, 2022). The significant difference between the two groups in terms of English proficiency could be attributed to various factors. The higher English proficiency of the pre-service teachers in the flipped group may be due to more opportunities for in-depth processing, frequent practice and reinforcement of higher-level cognitive skills, as well as the collaborative, scaffolding, and constructivist flipped learning environment. In contrast, the lower level of English proficiency of those in the traditional group might be attributable to the passive learning style among the Chinese EFL learners in the teacher-centered classroom environment.

To be specific, the flipped classroom may support the pre-service teachers to progressively achieve higher levels in Bloom's taxonomy, starting with lower-order cognitive tasks before class and then advancing to more complex cognitive tasks during class hours. The higher-level cognitive activities in flipped classes, as suggested by

Djamàa (2020), could facilitate the integration of knowledge and prompt corrective actions, resulting in higher English proficiency levels for the EFL learners. Moreover, the flipped classroom, which aligns with socio-constructivist theories, emphasized the active construction of knowledge through social interactions and authentic learning environments. By engaging in communicative and collaborative tasks, the pre-service teachers could actively build their own knowledge, scaffolded and mediated by their peers and instructors. It was very likely that the communicative, collaborative, and active learning atmosphere extended numerous opportunities outside and inside the classroom for the pre-service teachers to have linguistically rich interactions with learning materials, peers, and the instructor, resulting in significantly enhanced English proficiency levels. In contrast, the traditional classroom might intensify the passivity of the Chinese EFL learners influenced by the traditional Confucian culture. As remarked by Karjanto and Simon (2019), many

East Asian students who grow up in the Confucian Heritage Culture are often portrayed as quiet and passive learners and may be accustomed to passively listening to their teachers' instructions, leading to reticence (Chang and Lin, 2019) and a lack of participation in class activities (Suranakkharin, 2017; Lee and Wallace, 2018; Karjanto, 2021). These typical learning features, which might exacerbate in the traditional teacher-centered classroom, probably hindered the social construction of language knowledge and development of associated skills of the pre-service teachers, leading to lower English proficiency levels compared to their counterparts.

Furthermore, as opposed to the second null hypothesis, the results revealed a significant interaction effect between educational background and teaching approach, indicating that the impact of educational background on English proficiency differed in the flipped and traditional groups. Specifically, in the flipped classroom, urban pre-service teachers achieved higher English proficiency than their rural counterparts, whereas no significant difference was found in the traditional classroom. Additionally, the urban pre-service teachers in the flipped classroom outperformed their urban peers in the traditional classroom, while the rural pre-service teachers did not differ significantly in the two classes.

The finding concerning the higher English proficiency of urban pre-service teachers when compared with their rural classmates in the flipped classroom corroborated previous research by Lv and Wang (2016) that reported polarized achievements of urban and rural students in flipped language classes. These researchers argued that the flipped classroom may bring more positive learning outcomes to urban students than rural ones, since those from rural areas may lack in sufficient learner autonomy and information technology capacities to quickly adapt to the new teaching model. This study supported this explanation, as the significant difference in English proficiency between urban and rural pre-service teachers may be attributed to their different adaptability levels in the flipped learning environment. As critically commented by Adnan (2017), technology in education can be a "gift" or a "curse" for students, depending on their capacities to utilize digital tools and resources. As a technology-embedded teaching approach, the flipped classroom may favor urban pre-service teachers, who may possess higher digital abilities to utilize technology devices, and disadvantage rural pre-service teachers. The latter might face significant challenges not only in self-paced learning prior to class but also in collaborative social interactions during technology-aided classes. Compared to their rural peers, urban pre-service teachers may better integrate the two key components of the flipped classroom (preparatory work outside of class and interactive activities inside the classroom) with more skillful use of technological tools. Therefore, they may more easily and quickly adjust to and benefit from their flipped learning, ultimately achieving higher English proficiency.

Moreover, the study found no significant difference between the English proficiency of urban and rural pre-service teachers in the traditional classroom. This new finding was contrary to the results of previous studies that reported significant achievement gaps between Chinese urban and rural EFL learners in conventional language classrooms due to economic and educational disparities between urban and rural areas (Gao et al., 2019; Liu and Wang, 2021). In this study, although urban pre-service teachers performed slightly better than their rural counterparts in the teacher-centered environment, there was no conspicuous gap in their English proficiency. This may be due to the strong influence of traditional Confucian and Taoist culture on both urban and rural pre-service teachers, causing them to feel hesitant and anxious while expressing themselves and communicating with their peers, and to show deference to their teacher's authority. The high affective filter of these learners may have resulted in limited improvement in English proficiency, regardless of their educational backgrounds. These findings suggested that the traditional teaching approach, which emphasized the authoritative and central role of the instructor, may not be an effective instructional approach for stimulating the enhancement of English proficiency for both urban and rural pre-service teachers in this context.

Furthermore, the study revealed higher English proficiency of the urban pre-service teachers in the flipped group than their urban peers in the traditional group. This intriguing finding might be attributed to greater opportunities afforded by the flipped approach for the urban learners to utilize their strengths in technology and digital literacy. As digital natives (Prensky, 2001), these urban pre-service teachers might have been well-resourced with digital devices and grown up with a relatively high level of technical proficiency. Therefore, it was likely that they could benefit from the frequent use of digital devices in the flipped classroom to deepen their understanding of key concepts prior to class and engage in advanced cognitive development via active social interactions with their peers and instructor. In contrast, the urban pre-service teachers in the traditional classroom had fewer chances to employ their digital skills and tools to support their learning, as most of the class time was devoted to teacher-centered instruction and content recall, which limited their cognitive development. Further, they were left to tackle complex cognitive tasks independently after class, with less support and scaffolding available to their peers in the interactive and technology-aided flipped classroom. Overall, these findings suggested that the flipped classroom seemed to be a promising approach to enhance the English proficiency of urban pre-service teachers by leveraging technology and their digital capacities to facilitate advanced cognitive development.

Additionally, the new finding about the insignificant difference of the rural pre-service teachers in the two groups might be due to the inadequacy of learner autonomy of the rural pre-service teachers in the flipped classroom. While the utilization of digital tools and technology-assisted flipped teaching benefited the urban pre-service teachers in their study, it may not have resulted in the same outcomes for their rural peers. Even if offered a good number of opportunities to direct their learning in and out of the flipped classroom, the rural pre-service teachers, who might have fewer digital skills due to their economic constraints (Liu and Wang, 2021; Zhang, 2022), may encounter multiple challenges to handle their flipped learning with necessary autonomous abilities. As noted by Shahid et al. (2020), the lack of learner autonomy can be a significant hurdle impeding the development of EFL learners' English proficiency. Thus, it was not surprising that the rural pre-service teachers in the flipped classroom who might be deficient in their learner autonomy could not achieve significantly higher English proficiency than their rural peers in the traditional classroom. This finding highlighted the great need to provide adequate support to Chinese rural pre-service teachers to help them maximize the potential benefits of the flipped classroom.

6 Conclusion

The present study made a significant contribution to the literature by examining the impact of the flipped classroom on English proficiency of first-year pre-service teachers in China with urban and rural educational backgrounds. The findings suggested that the flipped classroom approach was more effective in improving the English proficiency of the pre-service teachers, especially those with an urban educational background. However, the results also indicated that rural pre-service teachers may need additional support to gain equal benefits from this approach. To address this issue, teacher educators should provide sufficient training opportunities for flipped classroom, allow for an extended period of time for practice and adaptation of flipped learning and the use of technological tools, gradually delegate learning responsibilities to the pre-service teachers, and take on a supportive role in guiding the learners, particularly those with rural educational backgrounds, in their flipped language learning process. Additionally, it is recommended that educational institutions and authorities in the context provide adequate administrative, financial, and technical support to encourage the implementation of the flipped classroom, such as providing flipped classroom and technology training to both teachers and students, promoting staff collaboration and communication, increasing investment to teacher educational institutions to improve ICT infrastructure, and initiating programs to promote technology-assisted EFL education in rural secondary schools.

The current study had several limitations that warrant acknowledgement. First, the experiment's duration was relatively short, and the sample size was small, as the participants were recruited from a single university. Second, the research instrument was limited, resulting in an insufficient amount of data collected. To address these limitations, future research could benefit from conducting longitudinal studies to examine the long-term effects of the flipped classroom on pre-service teachers' English proficiency. Additionally, delving into the cultivation of key competencies crucial for sustainability of the learners, such as critical thinking, problem-solving, and collaboration, as highlighted by Karjanto and Acelajado (2022), would be beneficial. Expanding the sample size and scope of the English courses could lead to more generalizable results. Employing a variety of instruments to collect both quantitative and qualitative data would enhance the study's comprehensiveness. Investigating different factors, such as technology abilities, influencing the outcomes would offer a deeper understanding of the subject matter. Further, considering possible gender differences in learning within the flipped classroom, it is recommended to take into account gender characteristics when applying and further improving the pedagogy, potentially through algorithm-based machine learning. Lastly, probing into specific challenges encountered by pre-service teachers in flipped English classes, especially within Confucian Heritage Culture contexts, holds promise. Identifying these hurdles and proposing targeted solutions could pave the way for more effective and culturally sensitive pedagogical strategies.

Despite these limitations, this study contributed valuable empirical evidence and practical insights into the effectiveness of the flipped classroom approach in enhancing the English proficiency of first-year Chinese pre-service teachers. The findings of this study established a solid groundwork for subsequent research in China and in educational contexts with similar characteristics.

Data availability statement

The datasets generated for this study are available on reasonable request to the corresponding author.

Ethics statement

The study involving humans was approved by the Human Ethics Committee, Taylor's University. The study was conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

WC: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing.

Funding

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

References

Abdullah, M. Y., Hussin, S., and Ismail, K. (2019). Investigating the effects of the flipped classroom model on Omani EFL learners' motivation level in English speaking performance. *Educ. Inf. Technol.* 24, 2975–2995. doi: 10.1007/S10639-019-09911-5

Adnan, M. (2017). Perceptions of senior-year ELT students for flipped classroom: a materials development course. *Comput. Assist. Lang. Learn.* 30, 204–222. doi: 10.1080/09588221.2017.1301958

Alhamami, M., and Costello, H. (2019). Pre-service EFL teachers' expectations, needs, and challenges in a language learning and technology course. *J. Lang. Teach. Res.* 10, 593–602. doi: 10.17507/JLTR.1003.23

Alhamami, M., and Khan, M. R. (2019). Effectiveness of flipped language learning classrooms and students' perspectives. *J. English Foreign Lang.* 9, 71–86. doi: 10.23971/JEFL.V9I1.1046

Altas, E. A., and Enisa, M. (2021). The impact of flipped classroom approach on the writing achievement and self-regulated learning of pre-service English teachers. *Turk. Online J. Distance Educ.* 22, 66–88. doi: 10.17718/TOJDE.849885

Arslan, A. (2020). A systematic review on flipped learning in teaching English as a foreign or second language. *J. Lang. Linguist. Stud.* 16, 775–797. doi: 10.17263/JLLS.759300

Basal, A. (2015). The implementation of a flipped classroom in foreign language teaching. *Turk. Online J. Distance Educ.*, 28–37. doi: 10.17718/TOJDE.72185

Bergmann, J., and Sams, A. (2012a). Before you flip, consider this. *Phi Delta Kappan* 94:25. doi: 10.1177/003172171209400206

Bergmann, J., and Sams, A. (2012b). *Flip your classroom: Reach every student in every class every day*. Washington, DC: International Society for Technology in Education.

Bishop, J., and Verleger, M. A. (2013). "The flipped classroom: a survey of the research" in *2013 ASEE Annual Conference & Exposition*. 23.1200.1201–1223.1200.1218

Bloom, B. S. (1984). *Taxonomy of educational objectives*. London: Pearson.

Cabi, E. (2018). The impact of the flipped classroom model on students' academic achievement. *Int. Rev. Res. Open Distance Learn.* 19, 202–221. doi: 10.19173/IRRODL.V19I3.3482

Cambridge, E. S. O. L. (2019). *Cambridge IELTS 9 student's book with answers: Authentic examination papers from Cambridge ESOL*. Cambridge: Cambridge University Press.

Chang, C., and Lin, H. C. K. (2019). Classroom interaction and learning anxiety in the IRS-integrated flipped language classrooms. *Asia-Pacific Educ. Res.* 28, 193–201. doi: 10.1007/S40299-018-0426-X

Chen, W. (2019). Fanzhuan ketang moshixia yingyu zhuanye xuesheng yingyu xuexi jiaolv de diaocha yu yanjiu [A survey of English learning anxiety of English majors under flipped classroom]. *Chifeng xueyuan xuebao* 40, 124–129.

Chen, Z. (2020). A contrastive study on the causes of foreign language classroom anxiety of high school students in urban and rural areas. Master's thesis, Hainan Normal University.

Chinese Ministry of Education (2012). *Plan for ICT in Chinese Ministry of Education 2011–2020*. Available at: http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/s3342/201203/xxgk_133322.html (Accessed November 30, 2023).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Communist Party of China Central Committee and The State Council (2018). *National rural revitalization plan (2018–2022)*. Available at: http://www.gov.cn/zhengce/2018-09/26/content_5325534.htm (Accessed November 30, 2023).

Cuong, P. H. (2021). English language education in rural areas: current issues, complexities and ways forward. *VNU J. Sci. Educ. Res.* 37, 39–48. doi: 10.25073/2588-1159/VNUER.4538

Deepa, R. (2021). A study on rural and urban learners in learning English language and their difficulties in sentence formation in English. *Int. J. Multidiscip. Educ. Res.* 10, 105–110.

Djamàa, S. (2020). Lecture in the living room, homework in the classroom: the effects of flipped instruction on graduate EFL students' exam performance. *Comput. Sch.* 37, 141–167. doi: 10.1080/07380569.2020.1795513

Eppard, J., and Rochdi, A. (2017). "A framework for flipped learning" in *International Association for Development of the information society (IADIS) international conference on mobile learning* (International Association for Development of the Information Society).

Evans, W., and Savage, J. (2015). *Developing a local curriculum: Using your locality to inspire teaching and learning*. London: Routledge.

Fazel, I., and Ahmadi, A. (2011). On the relationship between writing proficiency and instrumental/integrative motivation among Iranian IELTS candidates. *Theory Pract. Lang. Stud.* 1, 747–757. doi: 10.4304/tpls.1.7.747-757

Feng, W., Liu, Y., and Qu, L. (2019). Effect of land-centered urbanization on rural development: a regional analysis in China. *Land Use Policy* 87:e104072. doi: 10.1016/j.landusepol.2019.104072

Gao, Y., Li, R., Li, S., and Wang, M. G. (2019). Nongcun zhongxue yingyu jiaoyu weiguan shengtai shizheng yanjiu [An empirical study on English education in rural areas from educational micro-ecosystem perspective]. *Donghua ligong daxue xuebao* 38, 62–65.

García-Sánchez, S., and Santos-Espino, J. M. (2017). Empowering pre-service teachers to produce ubiquitous flipped classes. *PROFILE Issues Teach. Prof. Dev.* 19, 169–185. doi: 10.15446/PROFILE.V19N1.53857

Getie, A. S. (2020). Factors affecting the attitudes of students towards learning English as a foreign language. *Cogent Educ.* 7:e1738184. doi: 10.1080/2331186X.2020.1738184

Ghufron, M. A., and Nurdianingsih, F. (2019). Flipped teaching with call in EFL writing class: how does it work and affect learner autonomy? *Eur. J. Educ. Res.* 8, 983–997. doi: 10.12973/EU-JER.8.4.983

Glassdoor Team (2021). *Answering questions about educational background*. Available at: https://www.glassdoor.com/blog/guide/educational-background/ (Accessed November 30, 2023).

Gok, D., Bozoglan, H., and Bozoglan, B. (2021). Effects of online flipped classroom on foreign language classroom anxiety and reading anxiety. *Comput. Assist. Lang. Learn.* 36, 840–860. doi: 10.1080/09588221.2021.1950191

Haghi, M. (2021). The effective integration of flipped classroom in ELT contexts: a review of recent literature. *Arab World English J.* 2, 120–132. doi: 10.24093/AWEJ/MEC2.9

Hashemi, A., and Daneshfar, S. (2018). A review of the IELTS test: focus on validity, reliability, and washback. *Indonesian J. English Lang. Teach. Appl. Linguist.* 3, 39–52. doi: 10.21093/ijeltal.v3i1.123

He, J. (2020). Research and practice of flipped classroom teaching mode based on guidance case. *Educ. Inf. Technol.* 25, 2337–2352. doi: 10.1007/S10639-020-10137-Z

Ingram, D., and Bayliss, A. (2007). IELTS as a predictor of academic language performance (part I). *IELTS Res. Rep.* 7, 1–68,

Jenkins, J. (2003). *World Englishes*. London: Routledge.

Jiang, H. X. (2019). Xinshidai waiyu jiaoyu gaige de jidian gouxiang [Some thoughts on reform of foreign language education in the new era]. *Waiyujie* 1, 13–16.

Karaaslan, H., and Çelebi, H. (2017). ELT teacher education flipped classroom: an analysis of task challenge and student teachers' views and expectations. *J. Lang. Linguist. Stud.* 13, 643–666,

Karjanto, N. (2021). "Active participation and student journal in Confucian heritage culture mathematics classrooms" in *Proceedings of the international conference on mathematics, geometry, statistics, and computation (IC-MaGeStiC 2021)* (Dordrecht: Atlantis Press), 89–91.

Karjanto, N., and Acelajado, M. J. (2022). Sustainable learning, cognitive gains, and improved attitudes in college algebra flipped classrooms. *Sustain. For.* 14:12500. doi: 10.3390/su141912500

Karjanto, N., and Simon, L. (2019). English-medium instruction calculus in Confucian-heritage culture: flipping the class or overriding the culture? *Stud. Educ. Eval.* 63, 122–135. doi: 10.1016/J.STUEDUC.2019.07.002

Ke, Z. (2016). Kecheng gaige yu nongcun xueye chenggong jihui- Jiyu A'shi banian zhongkao shuju de fenxi [Curriculum reform and the opportunity of academic success of rural students - Based on eight-year data of senior high school entrance examination from city A]. *Jiaoyu yanjiu*, 10, 95–105,

Kırmızıa, O., and Kömeç, F. (2019). The impact of the FC on receptive and productive vocabulary learning. *J. Lang. Linguist. Stud.* 15, 437–449. doi: 10.17263/jlls.586096

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/J.JCM.2016.02.012

Kurt, G. (2017). Implementing the flipped classroom in teacher education: evidence from Turkey. *Educ. Technol. Soc.* 20, 211–221.

Kusuma, I. P. I. (2022). How does a TPACK-related program support EFL pre-service teachers' flipped classrooms? *Learn. J. Lang. Educ. Acquis. Res. Netw.* 15, 300–325.

Lage, M. J., Platt, G. J., and Treglia, M. (2000). Inverting the classroom: a gateway to creating an inclusive learning environment. *J. Econ. Educ.* 31, 30–43. doi: 10.1080/00220480009596759

Lakarnchua, O., Balme, S., and Matthews, A. (2020). Insights from the implementation of a flipped classroom approach with the use of a commercial learning management system. *Turk. Online J. Distance Educ.* 21, 63–76. doi: 10.17718/TOJDE.762027

Lee, G., and Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: outcomes and perceptions. *TESOL Q.* 52, 62–84. doi: 10.1002/TESQ.372

Li, G., Jee, Y., and Sun, Z. (2018). Technology as an educational equalizer for EFL learning in rural China? Evidence from the impact of technology-assisted practices on teacher-student interaction in primary classrooms. *Lang. Lit.* 20, 159–184. doi: 10.20360/LANGANDLIT29415

Liu, J. L. (2021). Zai nongcun chuzhong yingyu jiaoxue zhong tuijin fanzhuan ketang de shijian yanjiu [Practical research on promoting flipped classroom in rural junior middle school English teaching]. *Jiazhang* 21, 122–123.

Liu, X., Gu, J., and Xu, J. (2023). The impact of the design thinking model on pre-service teachers' creativity self-efficacy, inventive problem-solving skills, and technology-related motivation. *Int. J. Technol. Des. Educ.* 34, 167–190. doi: 10.1007/S10798-023-09809-X

Liu, Y., and Wang, H. (2021). English education in rural secondary schools in Southwest China: problems and countermeasures under the rural vitalization strategy. *Int. J. English Lang. Teach.* 8, 32–40. doi: 10.5430/IJELT.V8N2P32

Loucky, J. P. (2017). "Studies of flipping classes with Asian students," in *Flipped instruction methods and digital technologies in the language learning classroom*, eds. J. P. Loucky and Ware, J. L. Hershey, Penn: IGI Global, 64–90.

Lv, T. T., and Wang, N. (2016). Jiyu SPOC+ shuzihua jiaoxue ziyuan pingtai de fanzhuan ketang jiaoxue moshi yanjiu- Yi daxue yingyu weili [A study on the

establishment and effect of the flipped classroom mode for SPOC+ teaching resource platform as applied in college English teaching]. *Zhongguo dianhua jiaoyu* 5, 85–90.

National Bureau of Statistics (2020). *Announcement on the renewal of the national statistical zoning code and urban and rural zoning code*. Available at: http://www.stats.gov.cn/tjsj/tjbz/tjyqhdmhcxhfdm/2020/index.html (Accessed November 30, 2023).

Noor, H. A. (2020). A probe into the different aspects of 'validity' and 'reliability' of IELTS writing test. *Int. J. English Literat. Soc. Sci.* 5, 968–972. doi: 10.22161/ijels.54.21

Ömer, Ö., and Akçayoğlu, D. İ. (2021). Examining the roles of self-efficacy beliefs, self-regulated learning and foreign language anxiety in the academic achievement of tertiary EFL learners. *Particip. Educ. Res.* 8, 357–372. doi: 10.17275/per.21.43.8.2

Ou, C. (2017). A review on language learner autonomy research in China (2006-2016): based on 12 key domestic journals. *English Lang. Teach.* 10, 76–86. doi: 10.5539/ELT.V10N11P76

Pang, Y. (2022). The role of web-based flipped learning in EFL learners' critical thinking and learner engagement. *Front. Psychol.* 13:1008257. doi: 10.3389/FPSYG.2022.1008257

Pariyanto, P., and Pradipta, B. (2020). Factors influencing an EFL learner's proficiency: an English teacher's perspective. *J. Lang. Lit. Cult. Stud.* 2, 89–97. doi: 10.30996/ANAPHORA.V2I2.3369

Prensky, M. (2001). Digital natives, digital immigrants part 1. *Horizon* 9, 1–6. doi: 10.1108/10748120110424816

Qu, P. (2017). Jiyu weixin pingtai de zonghe yingyu kouyu fanzhuan ketang moshi tanjiu [A flipped classroom model of oral practice in integrated English teaching based on WeChat platform]. *Ningbo Jiaoyu xueyuan xuebao* 17, 23–26.

Quaid, E. D. (2018). Reviewing the IELTS speaking test in East Asia: theoretical and practice-based insights. *Lang. Test. Asia* 8, 1–9. doi: 10.1186/s40468-018-0056-5

Rajabi, P., Mahmoodi, K., and Hosseini, S. A. (2021). Flipped classroom model and its impact on Iranian EFL learners' classroom anxiety and listening performance. *Comput. Lang. Learn. Electron. J.* 22, 1–16.

Roth, C., and Suppasetseree, S. (2016). "Flipped classroom: can it enhance English listening comprehension for pre-university students in Cambodia" in *Proceedings of classic: Learning in and beyond the classroom: Ubiquity in foreign language education*, 255–264.

Rural Education Action Program (2018). *Computers as tutors: Leveraging PCs to advance learning in China's rural schools*. Available at: https://reap.fsi.stanford.edu/sites/default/files/REAP116-EN.pdf (Accessed November 30, 2023).

Shahid, C., Ong, E. T., Wong, K. T., and Perveen, A. (2020). Expectations and reality of learner autonomy and communicative competence in Pakistani higher education institutions: a review. *Int. J. Educ. Psychol. Couns.* 5, 91–101. doi: 10.35631/IJEPC.534007

Strelan, P., Osborn, A., and Palmer, E. (2020). The flipped classroom: a meta-analysis of effects on student performance across disciplines and education levels. *Educ. Res. Rev.* 30:e100314:100314. doi: 10.1016/j.edurev.2020.100314

Sukerti, G. N. A., Rudiastari, E., and Susana, K. Y. (2020). The effectiveness of flipped learning in teaching writing. *Soshum J. Sos. dan Hum.* 10, 78–92. doi: 10.31940/soshum.v10i1.1634

Suranakkharin, T. (2017). Using the flipped model to foster Thai learners' second language collocation knowledge. 3L Lang. *Linguist. Lit.* 23, 1–20. doi: 10.17576/3L-2017-2303-01

Talbert, R. (2012). Inverted classroom. *Colleagues* 9, 89–95.

Turan, Z., and Akdag-Cimen, B. (2020). Flipped classroom in English language teaching: a systematic review. *Comput. Assist. Lang. Learn.* 33, 590–606. doi: 10.1080/09588221.2019.1584117

University of Cambridge Local Examinations Syndicate (2007). *IELTS handbook 2007*. Cambridge: University of Cambridge Local Examinations Syndicate.

Wang, H., Mansor, N. S., Wang, X., Hoon, L., and Darmi, R. (2021). Application of flipped classroom for English education based on network information technology. *J. Balk. Tribol. Assoc.* 27, 321–331.

Wang, Y. (2020). A study on college English high-efficiency class based on blended teaching mode of flipped classroom. *Theory Pract. Lang. Stud.* 10, 1066–1071. doi: 10.17507/tpls.1009.08

Wen, Q. F. (2019). Xinzhongguo waiyu jiaoyu 70 nian: Chengjiu yu tiaozhan [Foreign language education in China in the past 70 years: Achievements and challenges]. *Waiyu jiaoxue yu yanjiu* 5, 735–745.

Wu, S., Xu, Y., Han, J., and Jiang, M. (2021). Detecting the factors affecting the learning performance of students with different learning styles in flipped learning. *ICIC Exp. Lett. Part B Appl.* 12, 1137–1144.

Wu, X. C., and Zhang, H. M. (2021). Fangyi qijian shiping xianshang ketang dui xuesheng xuexi xingqu yingxiang de ge'an yanjiu- Yi yingyu zhuanye zonghe yingyu weili [The influence of online classroom teaching model on students' satisfaction and learning interests under the epidemic COVID-19: a case of integrated English course]. *Zhongguo duomeiti yu wangluo jiaoxue xuebao* 3, 51–54.

Xiu, G. (2021). Xinshidai waiyu zhuanye jiaoyu de xinzhuanbian [Changes and innovation in the education of English majors in the new era]. *Zhongguo waiyu* 1, 21–22.

Yang, C. C. R. (2017). An investigation of the use of the 'flipped classroom' pedagogy in secondary English language classrooms. *J. Inf. Technol. Educ. Innov. Pract.* 16, 1–20. doi: 10.28945/3635

Yang, C. C. R., and Chen, Y. (2020). Implementing the flipped classroom approach in primary English classrooms in China. *Educ. Inf. Technol.* 25, 1217–1235. doi: 10.1007/s10639-019-10012-6

Yen, D., and Kuzma, J. (2009). Higher IELTS score, higher academic performance? The validity of IELTS in predicting the academic performance of Chinese students. *Worcester J. Learn. Teach.* 3, 1–7.

Yue, A., Tang, B., Shi, Y., Tang, J., Shang, G., Medina, A., et al. (2018). Rural education across China's 40 years of reform: past successes and future challenges. *China Agric. Econ. Rev.* 10, 93–118. doi: 10.1108/CAER-11-2017-0222

Zhang, S. (2022). Chengshi, xiancheng ji nongcun xiaoxuesheng yingyu xuexi qinggan duibi yanjiu [A comparative study of English learning emotions of primary school students in cities, counties and rural areas]. *Jiaoxue yu guanli* 12, 27–32.

Zhang, J. Q., and Guo, L. (2018). Benkesheng ketang jiaolv yanjiu: Waiyu ketang jiaolv liangbiao (FLCAS) de xinxiaodu zaijianyan [Study on language anxiety of non-English majors: An examination of FLCAS's reliability and construct validity]. *Huabei ligong daxue xuebao* 18, 105–112.

Zhang, H., Li, J., Jiao, L., Ma, W., and Guan, C. (2016). The adjustment and effects of vocabulary teaching strategies in flipped classroom. *Creat. Educ.* 7, 1966–1973. doi: 10.4236/ce.2016.714199

The impact of teaching presence on students' online learning experience: evidence from 334 Chinese universities during the pandemic

Wen Li[1] and Weiping Wang[2]*

[1]School of Public Administration, Southwestern University of Finance and Economics, Chengdu, China, [2]Institute of China Innovation and Entrepreneurship Education, Hangzhou Normal University, Hangzhou, China

Teaching presence is the core role of teachers in online education and is the most promising mechanism for developing online learning communities. Drawing on the theoretical framework of teaching presence, and based on data from an online survey of university students from 334 Chinese universities, this study constructs a framework for teaching presence and compares the differences in teaching presence among different student groups, and further explores the impact of teaching presence on students' online learning experience and its heterogeneity in terms of gender and discipline. The study finds that teaching presence includes three dimensions: instructional design and organization, facilitating interaction, and direct instruction, and there are differences among different groups. Teaching presence has a significant impact on the online learning experience, among which facilitating interaction is the most important influencing factor. Heterogeneity examination reveals that the direct instruction dimension has a greater impact on the online learning experience of female students and students in science and engineering, while the facilitating interaction dimension has a greater impact on the experience of male students and students in science and engineering. For the development of online teaching, it is necessary to help online teachers comprehensively improve teaching presence, adopt different teaching strategies and improvement measures for different student groups, and pay attention to collecting and analyzing student behavioral data for teachers to reflect on and improve teaching.

KEYWORDS

online learning, teaching presence, group heterogeneity, Chinese universities, online teaching

1 Introduction

Because of the COVID-19 pandemic, instructors had to swiftly transition to online courses. This sudden shift has prompted the rapid advancement of online teaching. However, a substantial number of online courses struggle to effectively address students' requirements and achieve the intended course objectives (Rovai and Downey, 2010; Allen and Seaman, 2014). Scholars have undertaken extensive research to explore factors contributing to the enhancement of interaction quality within online courses. Notably, one crucial area that

requires further investigation concerns students' engagement with their online instructors, commonly referred to as teaching presence (Garrison et al., 1999; Khalid and Quick, 2016).

Teaching presence is a concept derived from the "Community of Inquiry" (COI) model, which also includes cognitive presence and social presence (Garrison et al., 1999). Three presences in COI framework are interrelated and interdependent constructs rooted in collaborative constructivist learning to guide online and blended learning (Armellini and De Stefani, 2016; Amemado and Manca, 2017). Teaching presence is the primary catalyst for formation of social presence and cognitive presence (Shea and Bidjerano, 2009) and has been interpreted as effective instructional strategies during the learning processes (Akyol et al., 2010). The work of Akyol and Garrison (2008) provides further support by highlighting the crucial role of teaching presence in facilitating student perceived learning and satisfaction. Students' satisfaction and learning outcomes together can represent a better understanding of online learning experience. Learning outcomes measure whether students achieve competencies in their learning (Weinert, 2001). And learning satisfaction represents an attitudinal construct and measures the affective aspect (Goh et al., 2017).

China ranks first globally in terms of both the number of online Massive Open Online Courses (MOOCs) offerings and the learner population. As of November 2022, the total count of MOOCs is projected to surpass 61,900. Furthermore, there will be an estimated 402 million registered users, 979 million learners, and 352 million MOOC credits earned by current students (Ministry of Education of the People's Republic of China, 2023). A great number of studies have revealed that teaching presence is crucial for ensuring student satisfaction in online courses. But the effects of the instructor's teaching presence on students' online learning experience are unclear. There is a need for continuing research studies related to specific areas, such as pedagogical strategies to promote learners' online learning experience, the impact of learner characteristics on learner's online learning experience (Cereijo et al., 1999; Hartley and Bendixen, 2001).

Hence, the purpose of this study is to construct a complete theoretical framework of teaching presence and explore these variables that may influence student's satisfaction in mainland China. Concurrently, it also determines the association between teaching presence, gender, discipline and student online learning experience.

2 Literature review

2.1 Teaching presence

Teaching presence has been confirmed as the core role of online teachers and one of the most promising mechanisms for developing online learning communities. Teaching presence is the design, facilitation, and direction of cognitive and social processes for the realization of personally meaningful and educationally worthwhile learning outcomes (Anderson et al., 2001). It mainly involves three key roles: instructional design and organization, facilitating interaction, and direct instruction (Garrison and Arbaugh, 2007). Despite there are many studies on the teaching presence, its measurement framework is still to be explored. For example, Wang et al. (2012) designed teaching presence as five factors: course content organization, instructional teaching, assessment, teaching activity

organization, and promoting discourse. Wang et al. (2021) investigated 408 Chinese college students and scaled teaching presence with five dimensions: design and organization, discourse facilitation, direct instruction, assessment, and technological support.

Meanwhile, existing research on the perception of teaching presence and its relationship with learners' characteristics lacks consensus. Shea et al. (2006) found no association between learners' characteristics (such as gender, age, employment status, distance from the campus, student registration status, reasons for taking online courses, and course duration) and their perception of teaching presence. However, others have discovered that both young (18–22 years old) and older (48–62 years old) respondents tended to equate teacher guidance and intentions with cognitive outcomes (Akyol et al., 2011). Despite the extensive research conducted on teaching presence, there is still a need for further exploration and development of its measurement framework (Rourke and Kanuka, 2009; Garrison et al., 2010).

2.2 The impact of teaching presence on online learning

The important impact and role of teaching presence on learners' online learning experience have been noticed and valued by scholars. The evaluation of learning effect perception and satisfaction in the context of an online learning experience occurs at the "result" level. This level involves capturing learners' psychological feedback and their overall assessment of the attained learning outcomes and results of the course (Liu et al., 2016). The study by Khalid and Quick (2016) found a significant positive correlation between teaching presence and learning satisfaction and can be used to predict learners' performance in learning persistence. Szeto (2015) found through case studies that compared to social presence and cognitive presence, learners' learning outcomes are more dependent on the performance of teaching presence. Teaching presence, as indicated by Caskurlu (2018), can influence students' participation and engagement. But in the research conducted by Zhao and Sullivan (2017), it was observed that an increase in the level of teaching presence corresponded to a decrease in students' participation and interaction. The effects of teaching presence on students' online learning experience are unclear.

The validation of the influence of particular aspects of the theoretical framework regarding teaching presence on the results of online learning is yet to be confirmed. Since they only conducted surveys based on a single course and a small scale, no unified conclusions have been reached. For instance, Wang et al. (2012) found in their study on the performance of teaching presence in online courses that there is a significant correlation between the level of teaching presence and learning achievements, and learners' learning achievements can be predicted through the performance of teaching presence. Wu et al. (2017) believe that direct teaching is the most important factor in teaching presence that affects the effectiveness of online learning, while Wang et al.'s (2012) concluded that only course content organization has a significant impact on the ability factors of learning performance. Last but not least, there is not a consensus on the effects of teaching presence on online learning experience (Wang and Liu, 2020), further research is needed in understanding and

interpreting the relationship between teaching presence and student learning experience.

2.3 Heterogeneity in the impact of teaching presence on online learning

Previous research not only focuses on the impact of teaching presence on online learning experience but also further explores contextual factors (e.g., course length, instructor facilitation, participants' college level) as variables that influence students' perceived teaching presence and satisfaction. For instance, Akyol et al. (2011) found that students' perception of teaching presence and satisfaction were higher in short-term courses than longer-term courses. In the study conducted by Epp et al. (2017), it was observed that students' perceived learning was more pronounced in instructor-led courses and longer-duration courses. However, the effect size for facilitation was found to be even more substantial in influencing students' perceived learning outcomes.

Past studies have also investigated that discipline and gender differences on the effects of student evaluations. In terms of learning engagement, according to the research of Lu et al. (2014), liberal arts students perceive more student-centered teaching methods and teacher-student communication than students in other disciplines. But in a recent study, Lim (2019) used sequential mixed-methods and found no significant difference in students' teaching presence levels across soft-pure, soft-applied, hard-pure, and hard-applied disciplines. Gender also has a significant impact on indicators such as academic challenge, active cooperative learning, and richness of educational practices in learning engagement (Yang and Zhang, 2016). Considering the gender differences in learning engagement and the characteristics of disciplines, even though the vast majority of previous research has shown that both teaching presence and dimensions of teaching presence are associated with student outcomes in online learning, research on how democratic variables could impact students' online learning experiences showed mixed results. Therefore, it is particularly important to explore the heterogeneity of the impact of teaching presence on learning experience.

In summary, the theoretical model of teaching presence is relatively mature, and small-scale surveys have been used to verify the impact of teaching presence on learning outcomes. However, since current studies have not conducted large-scale surveys, there has been no synthesis of studies that provide quantitative evidence to support the relationship between teaching presence and students' online learning experience. Based on this, this study will be based on the theoretical model of teaching presence, using large-scale undergraduate self-assessment data from Chinese universities, to investigate the impact of teaching presence on learning experience and its heterogeneity. Specifically, three research questions addressed:

1. What are the characteristics of college students' perception of teaching presence?
2. Does teaching presence have a significant impact on student learning experiences?
3. Are there any differences in the impact of teaching presence on online learning experience?

3 Data and methods

The data used in this study comes from a survey of college students' online teaching conducted in March 2020 by the Center for Teaching and Learning Development in Xiamen University (Wang and Li, 2022; Wu and Li, 2022). The survey collected 251,929 student samples from 334 universities in Mainland China. The questionnaire covers three parts: students' basic information, perceived teaching presence by students, online learning experience. Among them, students' basic information includes dimensions such as gender, age, discipline, institutional level, and grade. According to the research purpose, we excluded student samples that have not participated in online learning and deleted samples with missing values in key variables (case-wise deletion method). In the end, the effective sample entering the final analysis of this study is 223,092.

Online learning experience is the dependent variable in this study. Drawing on the conceptual definitions of Liu et al. (2016), this article mainly uses two items, online learning satisfaction and online learning outcome, to depict students' online learning experience. In the questionnaire, both are measured using a 5-point Likert scale. In the data analysis, drawing from the *refine approach* proposed by DiStefano et al. (2009), we used a regression method to create a factor scores (McNeish and Wolf, 2020), and then using the min-max normalization method, they are rescaled into 0–1 (Devlieger et al., 2016; Andersson and Yang-Wallentin, 2021), and finally they were converted into 0–100 (see Table 1).

The independent variable in this study is teaching presence. Through exploratory factor analysis and confirmatory factor analysis, it is also confirmed that the teaching presence of online learning for college students in China has three dimensions, namely, instructional design and organization, direct instruction, and facilitating discourse. The measurement items and reliability and validity test results of each dimension are shown in Table 2. It is particularly worth mentioning that, because the chi-square value is very sensitive to the impact of sample size in validity testing. The chi-square test is generally only suitable for cases with a sample size of 100–300. When the sample size is large, the chi-square value often becomes large. It is easy to reject the null hypothesis at this time. Given that our sample size exceeds 220,000, it is appropriate to relax the value standards of traditional model fit indicators, such as the chi-square value and RMSEA (Wu, 2009).

In the selection of control variables (see Table 1), this study not only includes dimensions such as gender, grade, and age that are involved in existing literature but also includes variables such as the region where the university is located, institutional level, type of university, and discipline, considering the characteristics of vertical and horizontal stratification in higher education.

4 Empirical results

4.1 Differences in teaching presence among different groups

This study measures differences in categorical variables through two methods of group comparison. The first is the T-test method used for binary variables (0–1 variables). The second is the variance test method (F-test) used for multi-classification variables. Since variance analysis cannot specifically show more detailed

differences between groups, researchers further use the Bonferroni multiple testing technique for pairwise post-hoc comparisons based on variance testing. Finally, we used the effect size to measures the real strength of mean comparison by Stata command (esize, esizei).

Table 3 displays the differences in teaching presence among different student groups. Overall, the group differences in teaching presence are diverse. First, except for the facilitating discourse dimension, females perform better than males in teaching presence in the other two dimensions. In terms of training and using experience, students with online learning experience and training experience have better teaching presence. In terms of the region where the school is located, universities in the eastern region perform best in all three dimensions of teaching presence, followed by the central region, and the western region is the worst. In terms of institutional level, the higher the institutional level, the better the performance in the instructional design and organization dimension, while in the direct instruction and facilitating discourse dimensions, vocational and technical colleges perform the best, and research universities perform

the worst. In terms of disciplines, in the three dimensions of instructional design and organization, direct instruction, and facilitating discourse, there is a trend that natural sciences and engineering majors perform the worst, while humanities perform the best. In terms of age, except for the facilitating discourse dimension, the other two dimensions of teaching presence show a trend: the younger the age, the better the performance. By grade, in the instructional design and organization dimension, juniors perform the best, and seniors perform the worst. In the direct instruction dimension, freshmen perform the best, and seniors perform the worst, showing a trend that the higher the grade, the worse the performance.

4.2 The impact of teaching presence on online learning

In this section, we will use the multiple linear regression (MLR) method to explore the impact of teaching presence on online learning experience, and then use the dominance analysis to identify the

TABLE 1 Descriptive analysis of variables.

Variable	N	Mean	Standard deviation	Min	Max
Online learning outcome	223,092	51.35	25.8	0	100
Online learning satisfaction	223,092	64.98	21.22	0	100
Instructional design and organization	223,092	70.63	14.27	0	100
Direct instruction	223,092	42.82	19.14	0	100
Facilitating discourse	223,092	58.78	15.5	0	100
Grade	223,092				
Freshman	88,823	39.81		1 = Freshman; 2 = sophomore; 3 = Junior; 4 = Senior	
Sophomore	70,731	31.7			
Junior	54,195	24.29			
Senior	9,343	4.19			
Institutional level	223,092	2.03		1 = research university; 2 = undergraduate colleges and universities; 3 = vocational and technical colleges	
Research Universities	4,053	1.82			
Undergraduate colleges and universities	208,975	93.67			
Vocational and technical colleges	10,064	4.51			
Discipline	223,092			1 = humanities; 2 = social science; 3 = natural science; 4 = engineering technology; 5 = medicine	
Humanities	47,018	21.08			
Social science	69,054	30.95			
Natural science	30,215	13.54			
Engineering technology	67,219	30.13			
Medicine	9,586	4.3			
Male (0 = female; 1 = male)	223,092	0.42	0.49	0	1
Public colleges and universities (0 = private; 1 = public)	223,092	0.77	0.42	0	1
Age	223,092			1 = after 80; 2 = after 90; 3 = after 00	
Post80s	608	0.27			
Post-90s	99,935	44.8			
Post-00s	122,549	54.93			
Using experience (0 = no; 1 = yes)	223,092	0.46	0.5	0	1
Training experience (0 = no; 1 = yes)	223,092	0.38	0.49	0	1

For categorical variables and 0–1 variables, the values in the third column represent percentages.

TABLE 2 Reliability, validity test and measurement methods of teaching presence.

Dimension	Item	Scoring	Reliability (Cronbach'α)	Validity
Instructional design and organization	The appropriate online course content	1–5 (1 means strongly disagree, 5 means strongly agree)	0.8716	$\chi 2 = 773.457; p = 0.046; RMSEA = 0.074;$ CFI = 0.988; TLI = 0.975
	Teachers' teaching attitude			
	Teaching assistants			
	Online technical service support			
	Help students develop good online learning behavior habits			
Direct instruction	Teachers keep abreast of students' learning status		0.9086	
	Teachers keep abreast of students' knowledge			
	Feedback on student concerns			
	Teacher on-site guidance and supervision			
Facilitating discourse	Timeliness of teacher-student interaction		0.7792	
	Communication and collaboration among students			
	Students communicate with teachers and choose learning content according to their needs			
Overall			0.8204	

TABLE 3 Differences of teaching presence in different groups of students.

Dimensions (T-test)	Gender	Public universities	Using experience	Training experience	/
Instructional design and organization	2.799($p = 0.000$; Cohen's $d = 0.193$)	−1.498($p = 0.000$; Cohen's $d = -0.109$)	−1.010($p = 0.000$; Cohen's $d = -0.057$)	−2.376($p = 0.000$; Cohen's $d = -0.150$)	/
Direct instruction	2.231($p = 0.000$; Cohen's $d = 0.152$)	0.166($p = 0.086$; Cohen's $d = -0.008$)	−0.229($p = 0.005$; Cohen's $d = -0.009$)	−1.352($p = 0.000$; Cohen's $d = -0.076$)	/
Facilitating discourse	0.104 ($p = 0.115$; Cohen's $d = 0.015$)	1.1453 ($p = 0.000$; Cohen's $d = 0.080$)	−2.370 ($p = 0.000$; Cohen's $d = -0.155$)	−5.557 ($p = 0.000$; Cohen's $d = -0.380$)	/

Dimensions (One-way ANOVA test)	Region	Institutional level	Discipline	Age	Grade
Instructional design and organization	$F = 4.20$ ($p = 0.0149$; $\eta^2 = 0.0000$)	$F = 333.45$; ($p = 0.000$; $\eta^2 = 0.0029$)	$F = 43.25$ ($p = 0.000$; $\eta^2 = 0.0007$)	$F = 9.22$; ($p = 0.000$; $\eta^2 = 0.0001$)	$F = 32.84$ ($p = 0.000$; $\eta^2 = 0.0003$)
Direct instruction	$F = 35.08$ ($p = 0.000$; $\eta^2 = 0.0004$)	$F = 5.17$; ($p = 0.000$; $\eta^2 = 0 0.0001$)	$F = 33.87$($p = 0.000$; $\eta^2 = 0.0008$)	$F = 71.28$; ($p = 0.000$; $\eta^2 = 0.0009$)	$F = 63.98$ ($p = 0.000$; $\eta^2 = 0 0.0014$)
Facilitating discourse	$F = 246.87$; ($p = 0.000$; $\eta^2 = 0.0020$)	$F = 47.44$ ($p = 0.000$; $\eta^2 = 0 0.0003$)	$F = 16.51$ ($p = 0.000$; $\eta^2 = 0.0005$)	$F = 0.40$ ($p = 0.6713$; $\eta^2 = 0.0000$)	$F = 162.27$($p = 0.000$; $\eta^2 = 0.0022$)

In the T test, the values represent the difference between the means of the two groups. The specific calculation method is to subtract the group with high assignment from the group with low assignment.

relative importance ranking of specific variables.[1] The study found that, keeping other variables constant, for every unit increase in the dimension of instructional design and organization, learning outcome improve by 0.409 points; for every unit increase in direct instruction,

learning outcome improve by 0.706 points; for every unit increase in facilitating discourse, learning outcome improve by 14.670 points (see Table 4).

Continuing to examine from the perspective of learning satisfaction (Table 4), this study also confirms that teaching presence can significantly affect students' experiences of online learning. In the satisfaction model, keeping other variables constant, for every unit increase in the dimension of instructional design and organization, learning outcome improve by 4.668 points; for every unit increase in the direct instruction dimension, learning satisfaction improves by 0.902 points; for every unit increase in the facilitating discourse dimension, learning satisfaction improves by 15.454 points.

1 To fully check the five main assumptions underlying multiple regression models (Osborne and Waters, 2019), we used the Stata command (e.g., estat imtest, white; estat vif) to check whether the assumptions of MLR is satisfied. This method is also used in section 4.3. We found that the assumptions of statistical test are basically met.

In the control variable section, regardless of whether it is the learning outcome or satisfaction model, males have higher evaluations than females. The higher the grade, the higher the student's evaluation. Public universities have lower evaluations than private colleges. Students with online learning experience have better effects and satisfaction with online learning. Humanities and social sciences students have better online learning outcome and satisfaction than science, engineering, and medical majors. In addition, students' online learning experiences also show differences in other dimensions. For example, in terms of institutional level, in the learning outcome model, the higher the institutional level, the better the learning outcome. However, in the satisfaction model, the higher the institutional level, the lower the student's satisfaction. In terms of grade, in the learning outcome model, the lower the grade, the worse the effect; while in the satisfaction model, there is no difference between students of different ages. In terms of online learning using experience, in the learning outcome model, students with using experience have higher evaluations; but in the satisfaction model, there is no significant relationship between using experience and satisfaction.

Although the above has confirmed that teaching presence will significantly improve the outcome and satisfaction of online learning, further analysis is needed on the impact of specific dimensions on effects and satisfaction, as well as the relative importance ranking in learning outcome and satisfaction. This is because the size of each coefficient cannot be used for direct comparison. Traditional practices are stepwise regression, significance testing, and coefficient standardization techniques. However, the order of introducing explanatory variables in stepwise regression is very subjective, and coefficient standardization does not know the relative importance of each dimension. Therefore, this study will introduce the Dominance Analysis proposed by Israeli (2007). This method aims to compare pairs of predictors across all subsets of the predictors in a model to determine the additional contribution that each predictor makes to the prediction model. In fact, the contribution to the coefficient of determination also reflects the contribution of different explanatory variables to the variance of the dependent variable.

In columns 3 and 6 in Table 4, the study also found that whether in the learning outcome or satisfaction model, facilitating discourse is the most important influencing factor, followed by instructional design and organization, and direct instruction dimensions. Moreover, relatively speaking, the importance of facilitating discourse is greater in the satisfaction model. Specifically, in the learning outcome model, the contribution of the facilitating discourse dimension is 0.274, which can be interpreted as the marginal contribution of this variable to the goodness of fit is 0.274. That is to say, in this linear regression, relative to the dimensions of instructional design and organization and direct instruction, the facilitating discourse dimension has a stronger explanatory power for the variance change of the dependent variable online learning outcome. In the learning satisfaction model, the contribution of the facilitating discourse dimension is 0.452. Therefore, in this linear regression, relative to the dimensions of instructional design and organization and direct instruction, the facilitating discourse dimension still has a stronger explanatory power for the variance change of the dependent variable online learning outcome.

In summary, the empirical results indicate that teaching presence has a significant impact on the online learning experience. Among the dimensions of teaching presence, facilitating discourse is the most

critical factor affecting both learning outcomes and satisfaction. The study also highlights the importance of instructional design and organization as well as direct instruction. The results can be used to guide the development of online education, especially in understanding the different dimensions of teaching presence and their relative importance in enhancing the online learning experience.

4.3 Heterogeneity of the impact of teaching presence on online learning: gender and discipline

In this part, we will use the multiple linear regression and interactions in regression to the examine whether the relationship between the teaching presence and the online learning experience changes depending on the value of another independent variable (e.g., gender and discipline). Among the initial characteristics of learners such as student gender, grade, age, discipline, region, and institutional level, only gender and discipline show significance in the impact of teaching presence on online learning experience (see Table 3). Combining the characteristics of female learning investment and discipline mentioned in the literature review section, the researchers will explore the group heterogeneity of the impact of teaching presence on online learning experience from the perspectives of gender and discipline.

Gender heterogeneity of the impact of teaching presence on online learning is shown in Table 5. The examination of heterogeneity mainly focuses on the value and direction of the main effects and interaction terms. In the learning outcomes model, the study found that, on the one hand, the direct instruction dimension has opposite effects on females and males. Specifically, for females, for every unit increase in direct instruction, learning outcomes improve by 1.717 points. For males, for every unit increase in direct instruction, learning outcomes decrease by 0.27 points (1.717–1.987). On the other hand, the facilitating discourse dimension has a greater impact on males. Specifically, for females, for every unit increase in facilitating discourse, learning outcomes improve by 13.737 points. For males, for every unit increase in the facilitating discourse dimension, learning outcomes improve by 15.540 points (13.737 + 1.803).

In the learning satisfaction model, the study also found that the impact of teaching presence on online learning varies by gender. First, instructional design and organization have a greater impact on males. Specifically, for females, for every unit increase in instructional design and organization, learning satisfaction improves by 4.530 points. For males, for every unit increase in the instructional design and organization dimension, learning satisfaction improves by 4.756 points (4.530 + 0.226). Secondly, direct instruction has a greater impact on females. Specifically, for females, for every unit increase in instructional design and organization, learning satisfaction improves by 1.239 points. For males, for every unit increase in direct instruction, learning satisfaction improves by 0.609 points (1.239–0.630). Finally, facilitating discourse has a greater effect on males. Specifically, for females, for every unit increase in the facilitating discourse dimension, learning satisfaction improves by 14.844 points. For males, for every unit increase in direct instruction, learning satisfaction improves by 16.077 points (14.844 + 1.233).

Regarding the discipline heterogeneity of the impact of teaching presence on online learning, for ease of interpretation of

TABLE 4 The influence of teaching presence on students' online learning experience.

	Learning outcome			Learning satisfaction		
	Coeff.	Standard error	Relative importance	Coeff.	Standard error	Relative importance
Independent variable						
Instructional design and organization	0.409***	(0.049)	0.001	4.668***	(0.033)	0.0546
Direct instruction	0.763***	(0.048)	0.0009	0.902***	(0.033)	0.0019
Facilitating discourse	14.670***	(0.052)	0.2747	15.454***	(0.035)	0.452
Control variable						
Male	2.622***	(0.101)		0.130*	(0.069)	
Grade (reference group: freshman)						
Sophomore	1.098***	(0.119)		0.018	(0.081)	
Junior	1.164***	(0.162)		1.066***	(0.110)	
Senior	3.452***	(0.266)		2.332***	(0.180)	
Institutional level (reference group: research universities)						
Undergraduate colleges and universities	0.141	(0.348)		−3.382***	(0.236)	
Vocational and technical college	0.893**	(0.409)		−4.625***	(0.277)	
Disciplines (reference group: humanities)						
Social science	−0.791***	(0.131)		−0.427***	(0.089)	
Natural science	−0.729***	(0.165)		−0.338***	(0.112)	
Engineering technology	−1.300***	(0.140)		−0.268***	(0.095)	
Medicine	−4.041***	(0.250)		−0.930***	(0.170)	
Public universities	−1.264***	(0.114)		−0.152**	(0.077)	
Age (reference group: post-80s)						
Post-90s	−3.586***	(0.887)		−0.474	(0.602)	
Post-00s	−3.431***	(0.891)		0.537	(0.604)	
Using experience	2.042***	(0.096)		0.084	(0.065)	
Training experience	1.860***	(0.099)		2.245***	(0.067)	
Constant	52.974***	(0.967)		67.333***	(0.656)	
Number	223,092			223,092		
R square	0.286			0.514		

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

the results, the five-category discipline is converted into two categories: humanities and social sciences and science and engineering. Table 5 reports the results of the discipline heterogeneity of the impact of teaching presence on online learning. In the learning outcomes model, there is no significant difference between arts and science and engineering in the impact of teaching presence. In the satisfaction model, there is a significant difference between arts and science and engineering in the impact of teaching presence. On the one hand, direct instruction has a greater impact on science and engineering. Specifically, for science and engineering students, for every unit increase in direct instruction, learning satisfaction improves by 0.048 points. For humanities and social sciences majors, for every unit increase in direct instruction, learning satisfaction improves by 0.039 points (0.048–0.009). On the other hand, the facilitating discourse dimension also has a greater effect on science and engineering students. Specifically, for science and engineering students, for every unit increase in the facilitating discourse dimension, learning

satisfaction improves by 0.907 points. For humanities and social sciences majors, for every unit increase in direct instruction, learning satisfaction improves by 0.895 points (0.907–0.012).

In conclusion, this section of the study reveals that the impact of teaching presence on online learning experience varies among different groups, particularly in terms of gender and discipline. For gender, the study found that the dimensions of teaching presence have different impacts on males and females. Specifically, direct instruction is more beneficial for females, while facilitating discourse is more beneficial for males in terms of learning outcomes. In terms of learning satisfaction, instructional design and organization have a greater impact on males, while direct instruction has a greater impact on females, and facilitating discourse has a greater effect on males.

For discipline, the study did not find significant differences in the impact of teaching presence on learning outcomes between humanities and social sciences and science and engineering. However, in terms of learning satisfaction, direct instruction and facilitating discourse have

a greater impact on science and engineering students compared to humanities and social sciences students.

5 Discussion and implications

5.1 Discussion

This study found that the teaching presence model is applicable to online teaching and learning in Chinese universities, consistent with previous research findings (Garrison and Arbaugh, 2007). For instance, our data analysis revealed that teaching presence positively influences student satisfaction and learning outcomes, corroborating established literature on the subject. Additionally, this study verified that teaching presence comprises three dimensions, a finding that has been subject to debate in past literature (Hoskins and Van Hooff, 2005; Shea et al., 2006). By conducting rigorous statistical analyses and quantitative assessments, we were able to confirm the distinct roles of design, facilitation, and direct instruction in shaping teaching presence, thereby providing empirical support for this conceptual framework. Moreover, our findings further revealed significant differences in online teaching activities among various student groups, challenging existing research conclusions (Piccoli et al., 2001; Arbaugh and Hwang, 2006). For example, while previous studies may have suggested uniformity in online teaching effectiveness across demographic groups (Zhao and Sullivan, 2017; Caskurlu, 2018), our data indicate that

females generally outperform males, and students with prior usage experience and training exhibit stronger teaching presence. Additionally, our study uncovered regional and institutional disparities in teaching presence performance. Contrary to prevailing assumptions, universities in the eastern region demonstrated superior performance in online teaching activities. Furthermore, our analysis revealed variations in instructional design effectiveness across institutional types, with higher-level institutions excelling in this aspect, while vocational and technical colleges demonstrated strengths in direct instruction and facilitating discourse. Lastly, our findings underscored the significant impact of student discipline, age, and grade on teaching presence. Discipline-specific differences were observed, highlighting the importance of developing tailored approaches to adapt teaching presence strategies to accommodate diverse learner characteristics. By incorporating these detailed data examples and comparisons with existing research, this paper strengthens its arguments and provides a more comprehensive understanding of the implications of the findings for educational practice and future research.

Furthermore, the study results revealed that the facilitating discourse dimension is the most important influencing factor, impacting teaching presence on the online learning experience. This outcome contrasts with the findings of other previously published investigations (Qiao et al., 2021) in China based on small-scale investigations (Wang et al., 2012). In the satisfaction model, the facilitating discourse dimension holds greater importance. Although Chen et al. (2020) examined learning outcomes, their conclusions

TABLE 5 Group heterogeneity of the effect of teaching presence on students' online learning.

	Gender			Discipline	
	Learning outcome	Learning satisfaction		Learning outcome	Learning satisfaction
Instructional design and organization	0.362***	4.530***	Instructional design and organization	0.031***	0.310***
	(0.068)	(0.046)		(0.005)	(0.003)
Male	2.566***	0.117*	Humanities and social sciences	1.649***	0.726*
	(0.101)	(0.069)		(0.609)	(0.414)
Instructional design and organization*male	−0.056	0.226***	Instructional design and organization*humanities and social sciences	−0.001	0.007
	(0.098)	(0.067)		(0.006)	(0.004)
Direct instruction	1.717***	1.239***	Direct instruction	0.040***	0.048***
	(0.065)	(0.044)		(0.003)	(0.002)
Direct instruction*male	−1.987***	−0.630***	Direct instruction*humanities and social sciences	−0.006	−0.009***
	(0.098)	(0.066)		(0.005)	(0.003)
Facilitating discourse	13.737***	14.844***	Facilitating Discourse	0.858***	0.907***
	(0.071)	(0.048)		(0.004)	(0.003)
Facilitating discourse*male	1.803***	1.233***	Facilitating discourse*humanities and social sciences	−0.007	−0.012***
	(0.103)	(0.070)		(0.006)	(0.004)
Control variable	Yes	Yes	Control variable	Yes	Yes
Constant	52.999***	67.371***	Constant	−2.768***	−10.208***
	(0.965)	(0.655)		(1.060)	(0.719)
Number	223,092	223,092	Number	223,092	223,092
R square	0.288	0.515	R square	0.284	0.513

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; Control variables: (1) For gender heterogeneity, the control variables are: age, discipline, institutional level, grade, age, university location, type of university, institutional level. (2) For discipline heterogeneity, the control variables are: age, gender, school level, grade, age, institutional location, type of university, institutional level.

align with this article, emphasizing that social interaction is the most significant factor in online learning. However, it should be noted that this conclusion conflicts with research also based on surveys of Chinese university students (Li and Jiang, 2009). In their study, the direct instruction dimension emerges as the most important influencing factor. Our findings present a different portrait of Chinese students, often characterized as passive, reproductive, and surface learners in the literature (Jones, 1999).

This study showed that teaching presence has a significant heterogeneous impact on college students' online learning outcome and satisfaction. Among students from different genders and disciplines, different dimensions of teaching presence have different impacts on students' learning experiences. The direct instruction dimension has a greater impact on the online learning outcome and satisfaction of females, while the facilitating discourse dimension has a greater impact on the online learning outcome and satisfaction of males. As Qin et al. (2022) found in online learning, males pay more attention to the external environment, while females pay more attention to the effectiveness and quality of learning. In the learning outcomes model, there is no disciplinary heterogeneity in the impact of teaching presence on online learning. However, in the learning satisfaction model, there is heterogeneity, and both direct instruction and facilitating discourse dimensions have a greater impact on the learning satisfaction of science and engineering students. The findings of this study were supported by a previous study in China (Lu et al., 2014). In summary, the study concludes that teaching presence, especially the facilitating discourse dimension, is crucial in online learning experiences. There are also significant differences in how teaching presence affects different groups, especially in terms of gender and discipline.

The findings of this study contribute significantly to the understanding of online teaching and learning dynamics, particularly within the context of Chinese universities. The dominance of the facilitating discourse dimension as the most influential factor impacting online learning experiences underscores the critical role of interaction and collaboration in virtual learning environments, aligning with contemporary pedagogical theories emphasizing social constructivism and collaborative learning in online education. However, the divergence from previous research conducted in China suggests the need for contextualized investigations considering cultural factors and institutional contexts, which may shape students' preferences and experiences regarding online interaction (Li and Jiang, 2009). Moreover, the gender and disciplinary heterogeneity in the impact of teaching presence on online learning outcomes and satisfaction reveal complex dynamics warranting further exploration. The observed differential effects among male and female students suggest underlying cognitive and socio-cultural factors influencing their engagement with online learning environments (Chen et al., 2020). Similarly, disciplinary variations (e.g., hard-applied disciplines vs. soft disciplines) underscore the importance of tailoring instructional approaches to specific requirements and conventions of different academic disciplines, highlighting the significance of discipline-specific pedagogical strategies in online education (Lim and Richardson, 2022).

5.2 Implications

Digital transformation is a slow process in education, which became an urgent topic in the spring of 2020 due to COVID-19 (Bogdandy et al., 2020). Our findings contribute to the emerging

literature on the teacher presence framework and provide insight into online education satisfaction, potentially setting further directions for research and practice. This study has constructed a localized teaching presence guiding framework as a breakthrough point to comprehensively assist teachers conducting online teaching in improving teaching quality and satisfaction.

For educators, the model of teaching presence offers a theoretical basis for strategic choices when transitioning to online teaching roles, validating these strategies through large-scale empirical surveys. This validation is crucial for educators' readiness for the digital transformation of education (Tóth et al., 2022). The teaching presence framework serves as a means to comprehend online learning and teaching methods, with its three dimensions offering both theoretical and practical guidance for strengthening teaching strategies in online environments. Our findings also underscore the importance for teachers to consider disciplinary and gender characteristics, advocating for the adoption of differentiated teaching strategies and improvement measures tailored to diverse student groups.

Institutions would benefit greatly from exploring ways to evaluate online learning. Our research results demonstrate that teaching presence has varying impacts on learners' outcomes and satisfaction across different genders and disciplines. This underscores the necessity for educational institutions and educators to acknowledge these differences and adapt their strategies accordingly to enhance the online learning experience for all students. Moreover, the data sources utilized in this study are essential for future AI-powered automation (Zarifis and Efthymiou, 2022), providing valuable references for teachers on improving student learning outcomes and supporting universities undergoing digital transformation.

Governments must recognize that in online teaching at central and western universities, teachers may tend to replicate traditional face-to-face teaching practices without fully transitioning from knowledge disseminators to designers and facilitators. Addressing the regional disparities in teaching presence is crucial, with governments needing to prioritize and support the perception of teaching presence in online teaching, particularly for students from central and western universities.

5.3 Limitations of the study

Certainly, this study still has third areas of deficiency awaiting further advancements in subsequent research. Firstly, dominance analysis, while useful for determining predictor importance in regression models, has limitations including assumptions of linearity and challenges in interpretation. To supplement it, machine learning techniques like feature importance from random forests (Mizumoto, 2023), permutation importance, SHAP values, and LIME can offer more flexibility and robustness, especially in handling non-linear relationships and high-dimensional data. Secondly, regarding the significant gender and subject heterogeneity in the impact of teaching presence on the online learning experience, this study aims to provide a theoretical foundation by reviewing the literature on gender learning and subject characteristics. This explanation partially accounts for the important moderating role of gender and discipline. However, it is acknowledged that the current data structure limits the depth of quantitative research to uncover detailed internal mechanisms,

concrete examples or case stud, and microscopic effects. While qualitative research could serve as a complementary approach, its expanded discussion is beyond the scope of this article, considering practical constraints related to article length and focus. Finally, future research should consider conducting comparisons between different countries or cultures to examine the heterogeneity of the impact of teaching presence on students' online learning. Especially noteworthy are the differences between Confucian-heritage culture and the Western university classroom teaching model (Biggs, 1998; Karjanto and Simon, 2019). These comparisons are crucial for understanding the integration of global online learning.

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

Ethical approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. 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

WL: Writing – review & editing, Writing – original draft, Funding acquisition, Data curation, Conceptualization. WW: Writing – original draft, Software, Methodology, Formal analysis.

Funding

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

Akyol, Z., and Garrison, D. R. (2008). The development of a community of inquiry over time in an online course: understanding the progression and integration of social, cognitive and teaching presence. *Online Learn.* 12, 3–22. doi: 10.24059/olj.v12i3.72

Akyol, Z., Ice, P., Garrison, R., and Mitchell, R. (2010). The relationship between course socio-epistemological orientations and student perceptions of community of inquiry. *Internet High. Educ.* 13, 66–68. doi: 10.1016/j.iheduc.2009.12.002

Akyol, Z., Vaughan, N., and Garrison, D. R. (2011). The impact of course duration on the development of a community of inquiry. *Interact. Learn. Environ.* 19, 231–246. doi: 10.1080/10494820902809147

Allen, I. E., and Seaman, J. (2014). Grade change: Tracking online education in the United States. Available at:http://www.onlinelearningsurvey.com/reports/gradechange.pdf

Amemado, D., and Manca, S. (2017). Learning from decades of online distance education: MOOCs and the community of inquiry framework. *J. E-Learn. Know. Soc.* 13, 21–32. doi: 10.20368/1971-8829/1339

Anderson, T., Liam, R., Garrison, D. R., and Archer, W. (2001). Assessing teaching presence in a Computer Conferencing Context. *Online Learn.* 5. doi: 10.24059/olj.v5i2.1875

Andersson, G., and Yang-Wallentin, F. (2021). Generalized linear factor score regression: a comparison of four methods. *Educ. Psychol. Meas.* 81, 617–643. doi: 10.1177/0013164420975149

Arbaugh, J. B., and Hwang, A. (2006). Does "teaching presence" exist in online MBA courses? *Internet High. Educ.* 9, 9–21. doi: 10.1016/j.iheduc.2005.12.001

Armellini, A., and De Stefani, M. (2016). Social presence in the 21st century: an adjustment to the community of inquiry framework. *Br. J. Educ. Technol.* 47, 1202–1216. doi: 10.1111/bjet.12302

Biggs, J. (1998). Learning from the Confucian heritage: so size doesn't matter? *Int. J. Educ. Res.* 29, 723–738. doi: 10.1016/S0883-0355(98)00060-3

Bogdandy, B., Tamas, J., and Toth, Z. (2020). Digital transformation in education during COVID-19: a case study. In 2020 11th IEEE international conference on cognitive infocommunications (CogInfoCom), 000173–000178.

Caskurlu, S. (2018). Confirming the subdimensions of teaching, social, and cognitive presences: a construct validity study. *Internet High. Educ.* 39, 1–12. doi: 10.1016/j.iheduc.2018.05.002

Cereijo, M. V. P., Young, J., and Wilhelm, R. W. (1999). Factors facilitating learner participation in asynchronous web-based courses. *J. Comput. Teach. Educ.* 18, 32–39.

Chen, T., Gong, Y. X., and Pu, Y. (2020). Exploring socialized meaning: interaction in online teaching and its impact on learning outcomes: based on a survey of online teaching in 334 universities. *High. Edu. Res.* 41, 72–81.

Devlieger, I., Mayer, A., and Rosseel, Y. (2016). Hypothesis testing using factor score regression: a comparison of four methods. *Educ. Psychol. Meas.* 76, 741–770. doi: 10.1177/0013164415607618

DiStefano, C., Zhu, M., and Mîndrilă, D. (2009). Understanding and using factor scores: considerations for the applied researcher. *Pract. Assess. Res. Eval.* 14.

Epp, C. D., Phirangee, K., and Hewitt, J. (2017). Student actions and community in online courses: the roles played by course length and facilitation method. *Online Learn. J.* 21, 53–77. doi: 10.24059/olj.v21i4.1269

Garrison, D. R., Anderson, T., and Archer, W. (1999). Critical inquiry in a text-based environment: computer conferencing in higher education. *Internet High. Educ.* 2, 87–105. doi: 10.1016/S1096-7516(00)00016-6

Garrison, D. R., and Arbaugh, J. B. (2007). Researching the community of inquiry framework: review, issues, and future directions. *Internet High. Educ.* 10, 157–172. doi: 10.1016/j.iheduc.2007.04.001

Garrison, D. R., Cleveland-Innes, M., and Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: student perceptions of the community of inquiry framework. *Internet High. Educ.* 13, 31–36. doi: 10.1016/j.iheduc.2009.10.002

Goh, C. F., Leong, C. M., Kasmin, K., Hii, P. K., and Tan, K. (2017). Students' experiences, learning outcomes and satisfaction in e-learning. *J. E-Learn. Know. Soc.* 13, 117–128. doi: 10.20368/1971-8829/144

Hartley, K., and Bendixen, L. D. (2001). Educational research in the internet age: examining the role of individual characteristics. *Educ. Res.* 30, 22–26. doi: 10.3102/0013189X030009022

Hoskins, S. L., and Van Hooff, J. C. (2005). Motivation and ability: which students use online learning and what influence does it have on their achievement? *Br. J. Educ. Technol.* 36, 177–192.

Israeli, O. (2007). A Shapley-based decomposition of the R-square of a linear regression. *J. Econ. Inequal.* 5, 199–212. doi: 10.1007/s10888-006-9036-6

Jones, A. (1999). *The Asian learner: An overview of approaches to learning*. Melbourne: The University of Melbourne.

Karjanto, N., and Simon, L. (2019). English-medium instruction Calculus in Confucian-heritage culture: flipping the class or overriding the culture? *Stud. Educ. Eval.* 63, 122–135. doi: 10.1016/j.stueduc.2019.07.002

Khalid, M. N. M., and Quick, D. (2016). Teaching presence influencing online Students' course satisfaction at an institution of higher education. *Int. Educ. Stud.* 9:62. doi: 10.5539/ies.v9n3p62

Li, H. Z., and Jiang, G. Z. (2009). Evaluation of teaching presence in online teaching: a case study of the online course 'Research methods in distance education. *China Dis. Edu.* 12, 44–47.

Lim, J. (2019). *Disciplinary differences in a community of inquiry*. West Lafayette: Purdue University.

Lim, J., and Richardson, J. C. (2022). Considering how disciplinary differences matter for successful online learning through the Community of Inquiry lens. *Comp. Educ.* 187:104551. doi: 10.1016/j.compedu.2022.104551

Liu, B., Zhang, W. L., and Jiang, Y. J. (2016). Online course learning experience: connotation, development, and influencing factors. *China Educ. Technol.* 10, 90–96.

Lu, G. S., Peng, Z. X., and Hu, W. (2014). Analysis of differences in learning experience of college students in different subjects. *J. Soochow Univ.* 2, 64–73.

McNeish, D., and Wolf, M. G. (2020). Thinking twice about sum scores. *Behav. Res. Methods* 52, 2287–2305. doi: 10.3758/s13428-020-01398-0

Ministry of Education of the People's Republic of China, Technology empowers education and shares university resources. (2023) Available at:http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/2023/2023_zt01/mtbd/202302/t20230212_1043924.html

Mizumoto, A. (2023). Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests. *Lang. Learn.* 73, 161–196. doi: 10.1111/lang.12518

Osborne, J. W., and Waters, E. (2019). Four assumptions of multiple regression that researchers should always test. *Pract. Assess. Res. Eval.* 8:2. doi: 10.7275/r222-hv23

Piccoli, G., Ahmad, R., and Ives, B. (2001). Web-based virtual learning environments: a research framework and a preliminary assessment of effectiveness in basic IT skills training. *MIS Q.* 25, 401–426. doi: 10.2307/3250989

Qiao, W. F., Liu, W. T., and Li, M. L. (2021). Online teaching in the eyes of students: behavior, effects, and challenges - based on a survey of Tsinghua University Students' online learning behavior during the COVID-19 pandemic. *Tsinghua Univ. Educ. Res.* 42, 57–66.

Qin, H. X., Fang, F., and Zhou, J. H. (2022). Gender differences in university Students' satisfaction and continuous use intention in online teaching. *Univ. Educ. Sci.* 1, 44–53.

Rourke, L., and Kanuka, H. (2009). Learning in communities of inquiry: a review of the literature. *J. Dist. Educ.* 23, 19–48.

Rovai, A. P., and Downey, J. R. (2010). Why some distance education programs fail while others succeed in a global environment. *Internet High. Educ.* 13, 141–147. doi: 10.1016/j.iheduc.2009.07.001

Shea, P., and Bidjerano, T. (2009). Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education. *Comp. Educ.* 52, 543–553. doi: 10.1016/j.compedu.2008.10.007

Shea, P., Li, C. S., and Pickett, A. (2006). A study of teaching presence and student sense of learning community in fully online and web-enhanced college courses. *Internet High. Educ.* 9, 175–190. doi: 10.1016/j.iheduc.2006.06.005

Szeto, E. (2015). Community of Inquiry as an instructional approach: what effects of teaching, social and cognitive presences are there in blended synchronous learning and teaching? *Comput. Educ.* 81, 191–201. doi: 10.1016/j.compedu.2014.10.015

Tóth, T., Virágh, R., Hallová, M., Stuchlý, P., and Hennyeyová, K. (2022). Digital competence of digital native students as prerequisite for digital transformation of education. *Int. J. Emerg. Technol. Learn.* 17, 150–166. doi: 10.3991/ijet.v17i16.31791

Wang, G. X., Bai, C. J., and Lu, H. (2012). An empirical analysis of teaching presence in network distance education courses. *China Educ. Technol.* 9, 42–47.

Wang, W. P., and Li, W. (2022). Regional differences in experiences of online learning among Chinese college students and influencing factors: based on analysis of the survey data of 334 institutions of higher education. *Chin. Educ. Soc.* 55, 384–403. doi: 10.1080/10611932.2023.2213155

Wang, Y., and Liu, Q. (2020). Effects of online teaching presence on students' interactions and collaborative knowledge construction. *J. Comput. Assist. Learn.* 36, 370–382. doi: 10.1111/jcal.12408

Wang, Y., Zhao, L., Shen, S., and Chen, W. (2021). Constructing a teaching presence measurement framework based on the Community of Inquiry Theory. *Front. Psychol.* 12:694386. doi: 10.3389/fpsyg.2021.694386

Weinert, F. E. (2001). *Concept of competence: A conceptual clarification*. US: Hogrefe & Huber Publishers.

Wu, M. L. (2009). *Structural equation modeling: Operations and applications of AMOS*. Chongqing: Chongqing People's Publishing House, 23–25.

Wu, X. G., Chen, X. H., and Wu, J. (2017). On the effect of presence on online learning. *Mod. Dis. Educ.* 2, 24–30.

Wu, D. G., and Li, W. (2022). The interim characteristics of large-scale online education at Chinese institutions of higher education: an empirical study based on a questionnaire survey of students instructors, and administrative personnel. *Chin. Educ. Soc.* 55, 330–383. doi: 10.1080/10611932.2023.2213140

Yang, L. J., and Zhang, W. (2016). Influencing factors and mechanism of college Students' learning engagement. *High. Educ. Develop. Eval.* 32, 49–61.

Zarifis, A., and Efthymiou, L. (2022). The four business models for AI adoption in education: giving leaders a destination for the digital transformation journey. In 2022 IEEE Global Engineering Education Conference (EDUCON), 1868–1872. doi: 10.1109/EDUCON52537.2022.9766687

Zhao, H., and Sullivan, K. P. (2017). Teaching presence in computer conferencing learning environments: effects on interaction, cognition and learning uptake. *Br. J. Educ. Technol.* 48, 538–551. doi: 10.1111/bjet.12383

Frontiers in
Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

frontiers

Frontiers in
Psychology

frontiers | Research Topics