frontiers | Frontiers in Education

TYPE Original Research
PUBLISHED 26 September 2022
DOI 10.3389/feduc.2022.912837

Check for updates

OPEN ACCESS

EDITED BY
José Sánchez-Santamaría,
University of Castilla-La Mancha, Spain

REVIEWED BY
Ariel Mariah Lindorff,
University of Oxford, United Kingdom
Timothy Fukawa-Connelly,
Temple University, United States

*CORRESPONDENCE
Xin Liu
Xin.Liu@UGent.Be

SPECIALTY SECTION
This article was submitted to Assessment, Testing and Applied Measurement, a section of the journal Frontiers in Education

RECEIVED 04 April 2022
ACCEPTED 24 August 2022
PUBLISHED 26 September 2022

CITATION
Liu X, Valcke M, Yang Hansen K and De Neve J (2022) Exploiting the linked teaching and learning international survey and programme for international student assessment data in examining school effects: A case study of Singapore.
Front. Educ. 7:912837.
doi: 10.3389/feduc.2022.912837

Exploiting the linked teaching and learning international survey and programme for international student assessment data in examining school effects: A case study of Singapore

Xin Liu[1]*, Martin Valcke[1], Kajsa Yang Hansen[2] and Jan De Neve[3]

[1]Department of Educational Studies, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium, [2]Department of Education and Special Education, Faculty of Education, University of Gothenburg, Gothenburg, Sweden, [3]Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

This paper attempts to demonstrate the usefulness of the linkage data from two international large-scale assessment studies, Teaching and Learning International Survey 2013 (TALIS) 2013 and Programme for International Student Assessment (PISA) 2012, in examining the effects of schools. Data from seven educational systems are used to link, and four critical issues with five selection criteria are applied to the data selected. The linking dataset facilitates the investigation of mathematics performance while considering individual learner characteristics, mathematics teacher variables in the classroom environment and the school-level variables. We extend the new avenue of research by developing a linked database geared to the specific mathematics teaching and learning domain to reflect the school mathematics educational environment. The case study using Singapore linkage data demonstrated the feasibility and potential of exploring school effectiveness. In Singapore, schools with teachers of a higher level of education and self-efficacy in teaching mathematics related to a higher level of school mathematics performance. The study offers a guideline and inspiration to the research community to exploit the rich information in both TALIS and PISA studies to facilitate school effectiveness studies.

KEYWORDS

TALIS and PISA, mathematics achievement, educational effectiveness, school climate, multi-level perspectives

Introduction

Educational effectiveness research – Dynamic model of educational effectiveness

Given the growing globalization of education policy and practice, evaluation research focusing on "efficiency" and "effectiveness" of educational outcomes has grown rapidly. Many studies search for the factors playing a role at different levels in the school context (e.g., student background characteristics, quality of instruction, school leadership) as well as at the level of the educational system or regional context (e.g., educational policy). These factors are expected to be associated with students' learning outcomes (e.g., cognitive, affective, psychomotor, and metacognitive) see (Creemers and Scheerens, 1994; Scheerens and Bosker, 1997; Opdenakker and Van Damme, 2000; Creemers and Kyriakides, 2008; Reynolds et al., 2014; Chapman et al., 2015; Kyriakides et al., 2020). Educational Effectiveness Research (EER) takes into account that students are nested within classrooms, that classrooms are nested within schools, and that schools are nested in the region/country context. Student learning outcomes are associated with variables at these multiple levels.

Scholars describe EER as a dynamic process in which multiple levels of the educational system interact, and teaching and learning constantly adapt to changing demands and opportunities, e.g., (Opdenakker and Van Damme, 2006a,b, 2007; Creemers and Kyriakides, 2008; Scheerens, 2013). Over the years, educational researchers tested and developed a more advanced EER model, labeled the "Dynamic Model of Educational Effectiveness" (Creemers and Kyriakides, 2008; Kyriakides et al., 2020).

The Dynamic Model of Educational Effectiveness (DMEE) situates education effectiveness at four nested levels: student, classroom/teacher, school, and system/context. **Figure 1** depicts this DMEE levels hierarchy, which attempts to describe the direct and indirect effects of related factors on a range of student outcomes.

Since teaching and learning are mainly situated at the student and classroom/teacher level, the DMEE also models the interrelationships between student factors (e.g., student background characteristics) and teaching practices. This implies that teachers adjust and apply teaching practices based on the characteristics of students to adapt the teaching to their needs. School factors influence teaching and learning through the implementation of, e.g., a school policy and by the creation of an optimal school learning environment for all. Nonetheless, students, teachers, and schools are agencies within a system or context that is defined by educational policies implemented in their countries, regions, or other functions operating above the school level (Kyriakides et al., 2017). For instance, in highly centralized or decentralized educational systems, the degrees of freedom in defining the learning environment, options for school leaders, or the degrees of freedom in opting for teaching styles, depend on the restrictions imposed by the supra-school level.

Available studies at the national and international levels and meta-analyses tested the validity of the DMEE, with a focus on variables at the different levels, a focus on the measuring dimensions, and a focus on the associations between variables and the learning outcomes, e.g., (White, 1982; Driessen, 2002; Sirin, 2005; Kyriakides et al., 2010, 2013, 2014; Van Damme et al., 2010; Antoniou and Kyriakides, 2013; Scheerens, 2013; Muijs et al., 2014; Panayiotou et al., 2014, 2016). These empirical studies shed light on specific factors that are associated with effective teaching and learning and provide insights to improve educational effectiveness research.

Dynamic Model of Educational Effectiveness (DMEE) highlights that micro-, meso- and macro-level factors are critical when analyzing learning outcomes. Additionally, the model helps in conceptualizing the nature of instructional quality. In this dissertation, DMEE will be applied as the theoretical framework to study mathematical instructional quality and to help in explaining mathematics performance by looking at associated variables at the student level and the school level.

The rise of international large-scale assessments (ILSAs) helped in providing reliable evidence to support policy development and implementation, and can be used to analyze the long-term implications of earlier decisions (Rutkowski et al., 2013; Wagemaker, 2014, 2020). The core feature of ILSAs is the generation of hierarchical data about the home, student, teacher, school, and societal factors to evaluate educational outcomes, develop country profiles, and foster comparison between educational systems (Rutkowski et al., 2013; Wagemaker, 2014). ILSAs provide multiple indicators, covering the student, (teacher) school, and system level of the DMEE, which allow for a decomposition of the variation in outcome measures. The ILSAs contribute to an investigation of educational outcomes both within and across countries and help policymakers learn from other countries (Klieme, 2013). Current ILSAs examples include the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS) conducted by the International Association for the Evaluation of Educational Achievement (IEA), and the Program for International Student Assessment (PISA) conducted by the Organization for Economic Co-operation and Development (OECD).

Some large-scale effectiveness studies have already applied the DMEE to measure educational quality and equity at the classroom, school, and system levels. Nilsen and Gustafsson (2016) explained variance linked to school climate when examining the relationship between school climate, teacher quality, and student's learning outcomes in eight-grade

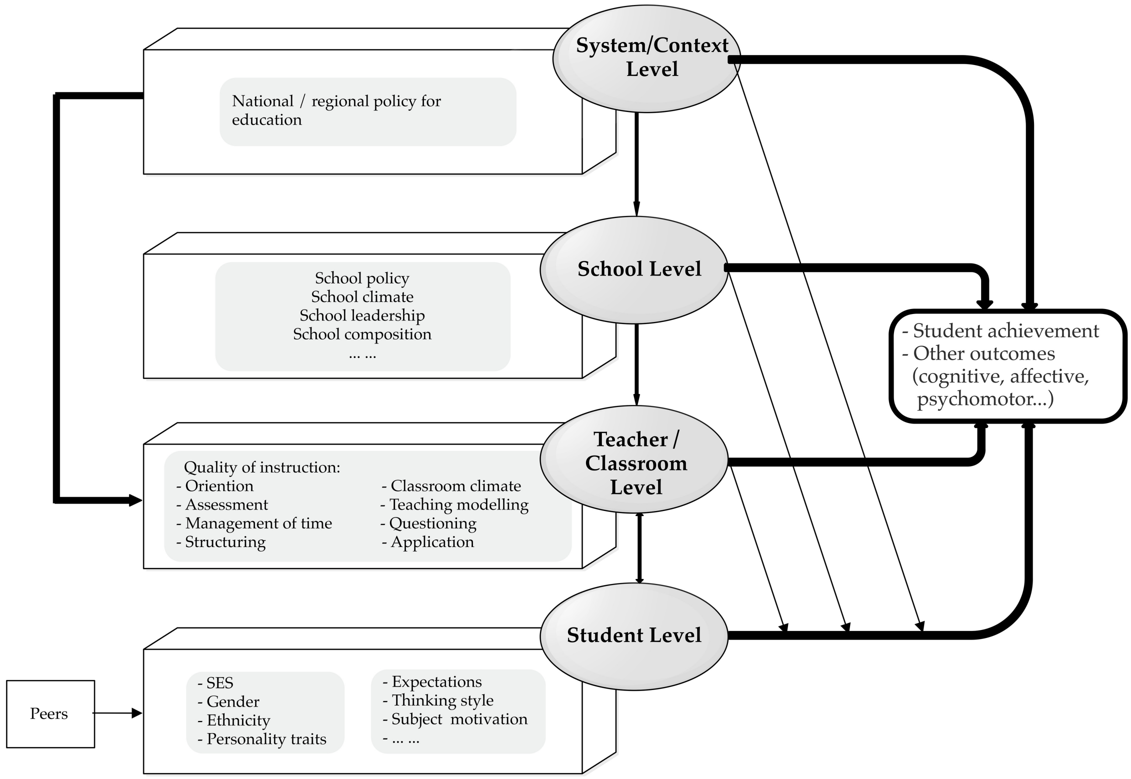

FIGURE 1
The Dynamic Model of Educational Effectiveness (adapted from Creemers and Kyriakides, 2007; Kyriakides et al., 2020).

across 38 countries using TIMSS 2007 and 2011 data. The main findings confirmed a positive and significant relationship between a positive school climate and mathematics outcomes. Meanwhile, teachers' attained education level and professional development were significantly and positively associated with mathematics achievement in grade eight. Other studies did build on PISA data, e.g., (Caro et al., 2016; Martínez-Abad et al., 2020; You et al., 2021). These studies revealed that student-level variables (e.g., socioeconomic status, motivation, enjoyment) and school factors (e.g., school type, school climate, school socioeconomic status) explain a significant proportion of the variation in student achievement.

Connecting mathematics teachers and mathematics performance applying teaching and learning international survey 2013 and programme for international student assessment 2012 linkage data

The PISA data provide insight into the backgrounds, beliefs, attitudes, motivations, mathematics achievement of students, and their perceptions of the learning environment but lack data collected from teachers in their classroom. In addition, the Teaching and Learning International Survey 2013 (TALIS), also set up by the OECD, collects data about the background, characteristics, beliefs, and teaching practices of teachers and their school principals (OECD, 2010). However, the absence of student data and their academic performance does not allow us to measure the association between teacher and teaching characteristics and student performance. This has been solved by the availability of the 2013 PISA-TALIS linkage database. Though a more recent PISA-TALIS linkage database from 2018 is available, the 2013 cycle is still the most recent one focusing on mathematics performance and instruction. Looking at the 2013 linkage database resulting from PISA and TALIS, the single anchor variable to accomplish a link is the school ID (variable "PISASCHOOLID"). This is the sole key variable shared in both TALIS and PISA. This implies that all analyses building on this database has to start from aggregated data at the school level in both TALIS and PISA. The linkage data helps in adding these distinctive teacher-level factors and perspectives (TALIS 2013 data) to the student mathematics performance data from PISA (PISA 2012 data). Moreover, comparisons between countries can center on differences in mathematics instruction, school environments, and education systems. This sounds promising, but much depends on the way we can link the two databases.

Several earlier studies already connected TALIS-PISA data by using the linkage database. These studies can be categorized into three types: (1) examining how school-level profiles of students impact teachers, e.g., (Austin et al., 2015; Sealy et al., 2016), (2) explaining student learning outcomes on the basis of the teacher or school variables at the school level, e.g., (Echazarra et al., 2016; Cordero and Gil-Izquierdo, 2018; Delprato and Chudgar, 2018; Mammadov and Cimen, 2019), and (3) statistical matching and guidelines for data fusion, e.g., (Kaplan and McCarty, 2013; Leunda Iztueta et al., 2017; Gil-Izquierdo and Cordero, 2018; Strietholt and Scherer, 2018). These studies provide – next to empirical evidence about theoretical assumptions – practical information on how to link available TALIS data and PISA data. For instance, a study was conducted by Cordero Ferrera and Gil-Izquierdo (2016). The researchers proposed guidelines for utilizing the original TALIS-PISA Link 2013 data and how this could be further linked to PISA 2012 data. They next studied the relationship between (general) teaching strategies and student mathematics performance in the Spanish context (Cordero and Gil-Izquierdo, 2018; Gil-Izquierdo and Cordero, 2018). Delprato and Chudgar (2018) utilized the linking database to link the variables competitive pressure, school autonomy, and teaching practices when looking at students performing in private and public schools, and this in the context of three countries. Huang et al. (2019) examined the relationships between variables of their school excellence model (e.g., school responsibility, distributed leadership, human resources, material resources) and student achievement in reading, mathematics, and science by applying data from Singapore. Also, three OECD working papers (Austin et al., 2015; Echazarra et al., 2016; Le Donné et al., 2016) focused on the link between student-level factors and teacher variables, between teaching strategies and student's learning strategies, and student PISA mathematics outcomes; in eight countries. An overview of the specific literature using TALIS and PISA linkage data is presented in **Table 1**.

Notwithstanding the availability of these earlier studies, the present study goes further. Firstly, the earlier studies did neglect that the teacher data did originate from different subject teachers. As such, they linked, e.g., data from language teachers to student mathematics outcomes. The PISA TALIS linkage dataset does not differentiate between mathematics and non-mathematics teachers. This raises the question about the adequacy of this choice: Is it possible to use a sample of teachers from other disciplines to convey "mathematical content knowledge" and "mathematical pedagogical content knowledge" to students during instruction? Is it plausible to use students perceived other subject teachers' instructional behaviors to represent their perceptions of "quality of mathematics instruction"? Is it reasonable to use the professional knowledge and instructional behaviors of teachers in other disciplines to explain "mathematical performance"?

Shulman (1986, 1987) highlighted three core categories of teachers' professional knowledge, namely, content knowledge (CK), general pedagogical knowledge (PK), and pedagogical content knowledge (PCK). CK is summarized as a teacher's deep and thorough understanding of the subject matter to be taught, such as the body of knowledge – facts, theories, principles, concepts, and ideas – they should master to be effective. PK refers to the knowledge about teaching and learning that transcends subject matter, such as general theories and principles of classroom behaviors and management, how students are learning, and how best to facilitate that learning in a variety of situations. PCK can be described as the knowledge of specific-subject instructional strategies, the knowledge of representations and explanations, and the knowledge of students' cognitions and (mis)conceptions (e.g., using appropriate strategies to describe ideas, understanding the particular needs of their particular students, providing explanations, making content accessible, setting up tasks to teach subject-matter knowledge). Of course, the three knowledge domains are interconnected. CK leads to teachers knowing what to teach (knowledge of subject matter). PK influences teachers knowing how to teach (general teaching knowledge). Moreover, PCK is the specialized expert kind of knowledge of how to transform subject matter representations "to make content comprehensible to students, combining an understanding of content and pedagogy specifically for instruction (Ball et al., 2005; Ma, 2010; Kleickmann et al., 2017).

In mathematics education, PCK features distinctive subject-specific characteristics. Shulman (1986) and several scholars expanded as such mathematical PCK. This refers to knowledge of the mathematics curriculum, knowledge of the aims of mathematics teaching, and knowledge of the construct of mathematics for teaching and learning (Grossman, 1990; Hill et al., 2004, 2005, 2008; Ball et al., 2008; Blömeke et al., 2012; Senk et al., 2012). Specifically, these components include, for example, conventional mathematical language, mathematical communication, worthwhile mathematical tasks, and making connections links between mathematical topics see (Hunter, 2005; Ainley et al., 2006; Anghileri, 2006; Watson and Mason, 2006; Chapin and O'Connor, 2007). In the case of mathematics teachers, holding a degree in mathematics is expected to ground their solid mathematical professional knowledge. However, also their pedagogical knowledge dimension is to be developed to guarantee that they adopt teaching behavior that leads to the effective delivery of the instructional content.

This critical stance toward the available linking data research in the literature explains the different approaches adopted in the present study. We prefer to interpret mathematics achievement and instructional quality by starting from the unique perspective of mathematics teachers. This implied a redesign of the available linkage dataset by focusing on "mathematics teachers." A second difference with earlier studies building on the PISA TALIS link is that we catered for the bias induced by the time gap between

TABLE 1 Overview of papers using the linkage data from teaching and learning international survey-programme for international student assessment (TALIS-PISA) Link 2013 and PISA 2012.

Category	Author	Brief introduction
School context features impact teachers	Sealy et al., 2016	Examine the relationships between principal job satisfaction, school characteristics, roles of the principal, and student achievement in eight countries.
	Austin et al., 2015	Aggregate student data to the school level to examine how student factors in a school may influence teachers' work, their attitudes, and their perceived needs for support (multilevel regression models).
Teaching strategies and students' learning	Cordero and Gil-Izquierdo, 2018	Examine the different teaching strategies (teacher characteristics, satisfaction of teacher with profession, student management efficacy, school ownership, curriculum, and assessment) on student achievement in Spain. The research is based on an instrumental variable approach.
	Delprato and Chudgar, 2018	Focus on understanding how systemic differences between private and public educational institutions (namely competitive pressure, administrative autonomy, staffing practices, and accountability) can explain differences in students' performance (mathematics, reading, and science) in Australia, Portugal, and Spain.
	Echazarra et al., 2016	Examines how particular teaching and learning strategies are related to student performance on specific PISA test questions, particularly mathematics questions: four teaching strategies– teacher-directed, student orientation, formative assessment, and cognitive activation – and three approaches to learning mathematics – memorization, control, and elaboration strategies.
	Fernández-Díaz et al., 2016	Analyze the relationships between the results from PISA 2012 and those relating to the teaching practice of secondary TALIS 2013, trying to find out the consistencies and discrepancies between the results of both.
	Huang et al., 2019	Investigate the relationships between the key elements of school excellent model variables (e.g., school responsibility, distributed leadership, human resources, material resources) and student achievement in reading, mathematics, and science in Singapore.
	Le Donné et al., 2016	Explore the relationships between mathematics teachers' teaching strategies and student learning outcomes in eight countries: active learning, cognitive activation, and teacher-directed instruction (24 items) at teacher, class, and school levels.
Methodological perspective	Gil-Izquierdo and Cordero, 2018	Guideline of theoretical linkage of TALIS and PISA.
	Leunda Iztueta et al., 2017	Use R software for statistical matching to link the PISA and TALIS studies with Spain's data.

PISA 2012 and TALIS 2013. These time gaps affect the extent to which teachers were teaching in the actual schools sampled in 2013. Some earlier studies neglected teacher mobility and assumed that a one-year time gap did not result in differences in teacher presence at the school level within a country. This assumption might result in less reliable results, and uncontrolled bias. Hence, we added another additional selection criterion to the revised linkage database to ensure that mathematics teachers in our redesigned database did actually work in the schools when the PISA students were studied in 2012. This helped guarantee that the sample of teachers did actually teach PISA 2012 students in the same school, and how their "mathematics professional knowledge" could be associated with a proportion of the variation in "school mathematics performance.

Present study

The above helps to add focus to this study by connecting the topics "mathematics teachers" and "mathematics performance." This brings us to the main focus of the present paper – exploring how to link TALIS 2013 and PISA 2012 data to study the relations between multiple educational effectiveness factors and mathematics achievement as reflected in the dynamic model. The purpose of the linkage is to use school-level data from mathematics teachers' responses in TALIS 2013 to contextualize student performance in PISA 2012 and shed light on how teacher- and school variables explain student achievement. Linking the information from two databases can help identify and explain the relationships between student socioeconomic background, student motivation and attitudes, mathematics teacher background and characteristics, mathematics teaching practices (aggregated at the school level), school compositions, and other school factors (e.g., school leadership, school environment), and school-level profiles of student learning outcomes. This mirrors a multi-level model that might provide insight into what improves student's mathematics learning process and outcomes, how mathematics teachers effectively handle the classroom and motivate their teaching, and how school principals support their teachers and carry out policies in practice. The results of a linked database might additionally be informative for policymakers, school administrators, and teachers themselves (e.g., supporting resources, professional development, teaching quality). Additionally, the linkage allows comparing the results across countries and developing more effective educational policies to improve teaching and student learning.

The general aim of this study is to design a linkage dataset for providing valuable information about multiple mathematics educational factors that potentially infuse future research about PISA 2012 mathematics performance using a multilevel perspective, especially building on mathematics teacher-related factors. We propose the following research question: Is it feasible to exploit a revised dataset to reflect the school effectiveness in mathematics teaching using the linkage data from TALIS 2013 and PISA 2012?

The current paper is organized as follows. First, we describe the structure of the original TALIS and PISA database and related questionnaires. Secondly, the sample selection criteria are given database redesign, and linkage of the datasets is introduced. Thirdly, a multi-level case study is applied to demonstrate the potential of using this newly designed linked database. Lastly, we address the limitations of linking TALIS and PISA in this way when studying the dynamic model in the context of educational effectiveness research.

Original database: Teaching and learning international survey and programme for international student assessment

TALIS[1] aims to investigate teachers' and school principals' learning environment and working conditions in private and public schools, mainly at the lower secondary education level, by exploring teacher-related factors, examining the roles of school principals, and how they support their teachers (OECD, 2010). PISA involves samples of 15-year-olds from schools – independent of their grade – and focuses on mapping their reading, mathematics, and science literacy. The PISA cycle is repeated every three years and focuses on a different main literacy domain. The PISA measurement framework reflects a skill-orientated and helps to describe mastery of competencies to handle the real-world challenges at the end of – in most countries – the compulsory education cycle (OECD, 2013a, 2017, 2019a; Stacey, 2015).

When implementing the TALIS 2013 cycle, participating countries could apply TALIS to mathematics teachers in a subsample of teachers who participated in the PISA 2012 cycle. This particular option was labeled the TALIS-PISA Link (TPL). The TPL helped start a series of studies examining student mathematics achievement from a multi-level perspective.

The second cycle of TALIS 2013 included 34 countries and economies. Four additional countries and economies administered the survey in 2014, resulting in a total of 38 countries. TALIS 2013 provided data about teachers and school principals, mainly from lower secondary education (ISCED[2] Level 2). Three sampling options were offered: a representative sample of teachers and principals in *option 1* primary education (ISCED Level 1), *option 2* in upper secondary education (ISCED Level 3), and *option 3,* the representative teachers of 15-year-olds and their principals drawn from the schools that already participated in PISA 2012, the so-called TPL mentioned above (OECD, 2009, 2010, 2013a, 2014a, 2019b).

TALIS 2013 collected data based on three questionnaires (see **Figure 2**) filled out by teachers or school principals: the *Teacher General Questionnaire*, the *School General Questionnaire,* and *Mathematics Teacher Questionnaire*. The first covered the teacher background and characteristics (e.g., professional development, teacher self-efficacy, teacher cooperation) and teachers' perspectives about their working environment. The *School General Questionnaire* was filled out by the principals and collected data about the school background and composition, teacher induction and mentoring, formal teacher appraisal, school autonomy, school leadership, a principal's background and job satisfaction, and school climate (e.g., school delinquency and violence, mutual respect).

In countries that signed up for the third sampling option (TPL), after completing the *Teacher General Questionnaire*, all mathematics teachers were additionally asked to complete the *Mathematics Teacher Questionnaire*. This helped identify specific data about their mathematics classes and instructional school climate (OECD, 2013b, 2014c). Sampling *option 3* comprised next to all mathematics teachers of a school, 20 non-mathematics teachers and one school principal of each of the 150 schools in an *option 3*-country (OECD, 2014c). Eight countries opted for the TALIS-PISA Link approach: Australia (AUS), Finland (FIN), Latvia (LVA), Mexico (MEX), Portugal (PRT), Romania (ROU), Singapore (SGP), and Spain (ESP). The TALIS-PISA Link helped to center on teaching practices in the target class[3], mathematics teaching strategies, educational approaches, initial training/education for teaching mathematics, and self-efficacy in teaching mathematics. TALIS-PISA Link offers a school-level perspective on mathematics instructional quality from TALIS 2013 that can be linked to student-level data from PISA 2012.

1 Three cycles of TALIS had been conducted in 2008, 2013, and 2018. The first cycle was conducted in 2008 and involved 24 countries. The second cycle was in 2013 and involved 34 countries and economies. Another four countries and economies were administrated in 2014.The third cycle was in 2018 and involved 48 countries and economies.

2 Classification of levels of education is based on the International Standard Classification of Education 1997: pre-primary education (ISCED level 0), primary education or first basic education (ISCED level 1), lower secondary education or second stage of basic education (ISCED level 2), upper secondary education (ISCED level 3), post-secondary non-tertiary level of education (ISCED level 4), the first stage of tertiary education (ISCED level 5), thee second stage of tertiary education (ISCED level 6).

3 Target class: Considering the teaching practices in the class, TPL selected a necessary "target class" to finish the mathematics module about *Mathematics Teacher Questionnaire*. "Target class" was composited of the majority of PISA-eligible "15-year-old" students in the class and identified as the first-class attended by 15-year-old students teachers taught in the current school year in TPL.

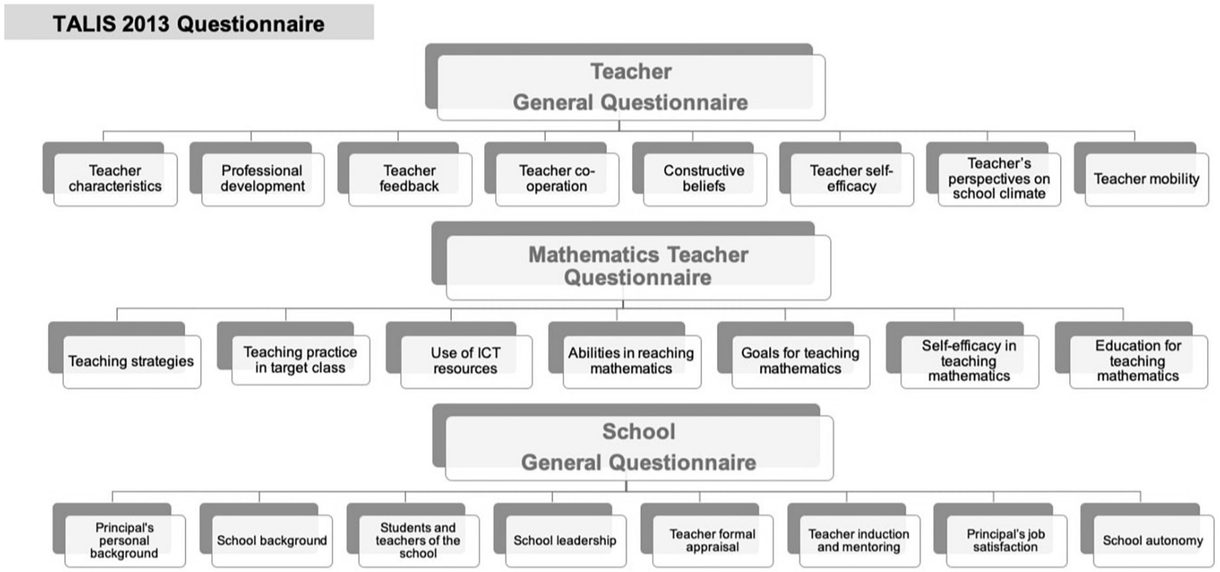

FIGURE 2
The main aspects of the teacher and school questionnaire in teaching and learning international survey 2013.

PISA 2012 was the fifth cycle and covered reading, mathematics, science, problem-solving and financial literacy, with mathematics as the primary domain (OECD, 2013a). PISA 2012 data was collected with three questionnaires: the *Student Questionnaire,* the *School Questionnaire,* and the *Parent Questionnaire*.

The *Student Questionnaire* focused on student characteristics, family background, personal intrinsic factors, student perspectives on the learning environment, teaching practices and school climate. The *School Questionnaire* – filled out by school principals – looked into school background information, school climate, school leadership, school curriculum assessment, school mathematics policies, and instructional practices. In 11 countries, also the *Parent Questionnaire* was administered to collect data about parents' background, their attitudes toward school, parent support for learning in the home, mathematics in the job market, children's past academic performance and academic and professional expectations in the field of mathematics (OECD, 2013a). Around 510,000 students, aged 15 years three months to 16 years two months, from 65 countries participated in PISA 2012: 34 OECD countries and 31 partner countries and economies. The main aspects of the student questionnaire and school questionnaire in PISA 2012 are summarized in **Figure 3**.

In the PISA 2012 Questionnaires, only limited data about teachers are being collected, and therefore large parts of the EER dynamic model about teaching effectiveness cannot be studied directly. TALIS offers a rich database to study the dynamic model in full by focusing on original teacher self-reported information. But this requires linking both separate datasets. The linkage will be established at the school-level since the only anchor variable shared in both databases is the school ID – PISASCHOOLID.

Tables 2, 3 summarize the number of schools and teachers in TALIS-PISA Link and students sampled from schools for PISA 2012 of the eight participating countries.

Redesigning the teaching and learning international survey-programme for international student assessment link database

When considering the linkage of TALIS-PISA Link 2013 and PISA 2012, critical issues need to be addressed. Firstly, the key sampling variable differs in TALIS and PISA. The TALIS-*Teacher General Questionnaire* builds on "grades" (i.e., ISCED Level 1, ISCED Level 2, and ISCED Level 3). However, the PISA *Mathematics Teacher Questionnaire* starts with teachers teaching students "the age of 15 years" (OECD, 2013b, 2014b,c). Linking data from both TALIS and PISA requires focusing on students from the same age group.

Secondly, TALIS teacher data cannot directly be linked to PISA individual student data (OECD, 2013b, 2014b,c; Le Donné et al., 2016). In other words, it is not possible to link a student to her or his personal mathematics teacher. In both databases, there is only one single anchor variable that is shared: the ID of the school (variable "PISASCHOOLID"). In view of linking

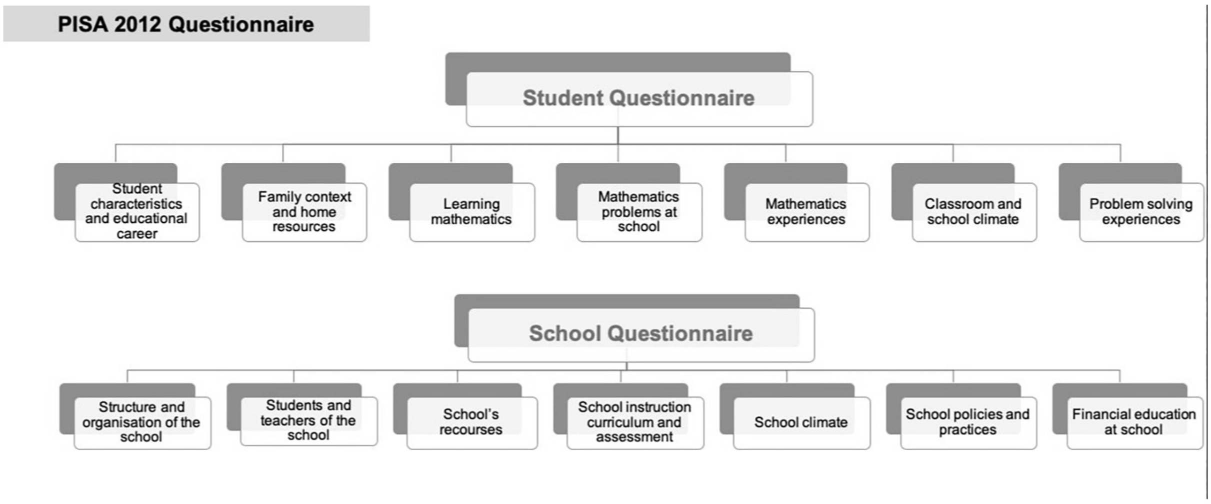

FIGURE 3
The main aspects of and student and school questionnaire in programme for international student assessment (PISA 2012).

TABLE 2 Overview of the original raw data of TALIS-PISA Link 2013 samples.

	AUS	FIN	LVA	MEX	PRT	ROU	SGP	ESP	Total
Number of schools for TALIS-PISA LINK	122	147	118	150	141	147	166	310	1 301
Respondent teachers in schools for TALIS-PISA Link	2 719	3 326	2 123	2 167	3 152	3 275	4 130	6 130	27 022

Source from OECD TALIS 2013 Database. AUS, Australia; FIN, Finland; LVA, Latvia; MEX, Mexico; PRT, Portugal; ROU, Romania; SGP, Singapore; ESP, Spain.

TABLE 3 Overview of the original raw data of and PISA 2012 samples.

	AUS	FIN	LVA	MEX	PRT	ROU	SGP	ESP	Total
Number of schools sampled for PISA 2012	775	311	211	1 471	195	178	172	902	4 215
Participating student sampled for PISA 2012	14 481	8 829	4 306	33 806	5 722	5 074	5 546	25 313	103 077

Source from OECD PISA 2012 Database. AUS, Australia; FIN, Finland; LVA, Latvia; MEX, Mexico; PRT, Portugal; ROU, Romania; SGP, Singapore; ESP, Spain.

the datasets, data have to be aggregated at the school level. This implies that no classroom-level information is available in the new dataset, but the average teacher and student factors in a school.

Thirdly, the administration of TALIS 2013 questionnaires occurred nearly one year after administering the PISA 2012 instruments. TALIS 2013 was conducted from September to December 2012 in Southern Hemisphere countries and from February to June 2012 in Northern Hemisphere countries. Whereas the Southern Hemisphere countries (AUS, SGP) developed PISA 2012 between May and August 2012, the Northern Hemisphere countries (FIN, LVA, MEX, PRT, ROU, ESP) were between March to May 2012 (OECD, 2014c; Echazarra et al., 2016). This resulted in a time gap that could create a misfit between teachers and students within the same school. To cater for this time gap, additional criteria were applied to refine teacher selection in view of a revised link dataset: a teacher should have at least one year of work experience in the Southern hemisphere and at least two years of work experience in the Northern hemisphere. In this way, we increased the probability to map the data from the actual teachers and students who participated in PISA 2012 with the data of teachers who participated in TALIS 2013.

Fourthly, since we focus on student mathematics achievement, the revised link database should solely center on data from mathematics teachers from the TALIS-PISA 2013 study. At the same time, we focused on mathematics literacy performance and related data from the PISA 2012 study.

Considering a linking procedure, Le Donné et al. (2016) proposed two approaches: either (A) PISA student data are aggregated at the school level and next merged with TALIS data; or (B) TALIS teacher data are aggregated at the school level and next merged with PISA data. Also, Gil-Izquierdo and Cordero (2018) see the two databases as potentially different "donor" or "recipient" datasets, and how this reflects a different merging approach: (a) TALIS as the recipient dataset and merging PISA data into TALIS based on the same PISASCHOOLID; (b) TALIS as the donor dataset and PISA as a recipient dataset that are merging TALIS data into PISA based on the same PISASCHOOLID.

One could state that (A) and (a) is equivalent to examining teacher outcomes (e.g., professional development, beliefs about teaching, self-efficacy) in the learning environment by measuring some constructs based on student's self-reported (e.g., learning motivation, attitudes toward school, teacher and student relation) in PISA (Austin et al., 2015). On the other hand, (B) and (b) can be seen as equivalent to evaluating student achievement depending on teachers' characteristics, teaching practice, and educational approach in the classroom. Making a choice for either approach depends on the nature of the research question being addressed.

Since we aim to use the redesigned linking dataset to analyze student-level data (mathematics literacy) by considering the teacher and school-level data, we opted for the second approach with PISA as a recipient dataset and merge the TALIS donor-data into PISA. The teacher information was aggregated at the school level before being merged into the student dataset. The resulting dataset structure fits the multi-level perspectives as reflected in the Dynamic Model of Educational Effectiveness (Creemers and Kyriakides, 2007). The resulting redesigned dataset consists of data organized at the individual student and school level from PISA 2012, the school profile of teacher factors from TALIS 2013, and school factors from both PISA 2012 and TALIS 2013.

Building on the above rationale, a Redesigned TALIS-PISA Link database (rTPL) was created to link "mathematics teacher" data to "student mathematics achievement" data resulting from the two original data sets. This rTPL reflects the following teacher sampling criteria (see **Figure 4**):

- Teacher data are from teachers with at least one year of work experience in the Southern hemisphere and at least two years of work experience in the Northern hemisphere.
- Teachers did teach mathematics to 15-year-old students in the test administration school year.
- The selected teachers did teach mathematics in the target class: the "target class" contains potential PISA pupils. In this way, teacher factors can be linked to pupils and their math performance.
- The teachers did fill out the *Mathematics Teacher Questionnaire*.
- The teacher was, as such, also a PISA mathematics teacher.

The "redesigned TALIS-PISA Link database" (rTPL) consisted of data from 3473 valid teachers from 1115 valid schools and representing 31,548 students from schools with matching PISASCHOOL ID in the TALIS-PISA Link and PISA 2012 (see **Table 4**).

Feasibility of using the redesigned teaching and learning international survey-programme for international student assessment link database

Next to the design of the rTPL database, the present article explores the feasibility of using the rTPL to test complex EER related multi-level models. Such a model builds on the theoretical assumptions from the Dynamic Model of Educational Effectiveness and the Opportunity-Propensity framework. The Opportunity-Propensity (O-P) framework has been put forward to explain associations with student performance; see **Figure 5**. Three main categorical predictors are presented in the model. These include antecedent factors, opportunity factors, and propensity factors (Byrnes, 2003, 2020; Byrnes and Miller, 2007; Byrnes and Wasik, 2009; Byrnes and Miller-Cotto, 2016). The antecedent factors are related to aspects of a students' home environment and socio-cultural demographics, including socioeconomic status, gender, race, ethnicity, and parental expectations for their children's academic achievement. The opportunity factors comprise aspects of the learning context (i.e., at home and school) that promote learning and development, such as content exposure, teaching strategies, and overall instructional quality. The propensity factors are related to a student's ability and willingness to learn in a particular context (e.g., prior knowledge, academic motivation, cognitive level).

According to the O-P framework, antecedent factors operate earlier and already lead to variations in opportunity factors and propensity factors. For instance, students from high-socioeconomic families are financially able to relocate to neighborhoods with schools that employ more qualified and effective teachers, receiving high-quality instruction (opportunity factor) while being able to mobilize high-level prerequisite knowledge. Hence, the academic achievement and development outcomes vary between students.

The empirical evidence is, for instance, abundant with studies linking teacher characteristics to student achievement. For example, teacher self-efficacy is an essential teacher characteristic and has been found to be strongly associated with the quality of instruction (Holzberger et al., 2013). In turn, effective teaching is a vital characteristic of high-performing schools, mirroring high student achievement and other educational outcomes (Muijs and Reynolds, 2002; Caprara et al., 2006). Meta-analysis studies from Hattie (2008) and others, e.g., (Desimone et al., 2002; Snow-Renner and Lauer, 2005) reiterate consistently that teachers' professional development may have the strongest impact on teachers' learning. Effective professional development seems to increase teacher self-efficacy and their instructional beliefs (Robardey et al., 1994; Rimm-Kaufman et al., 2006; Tschannen-Moran and McMaster, 2009) and with

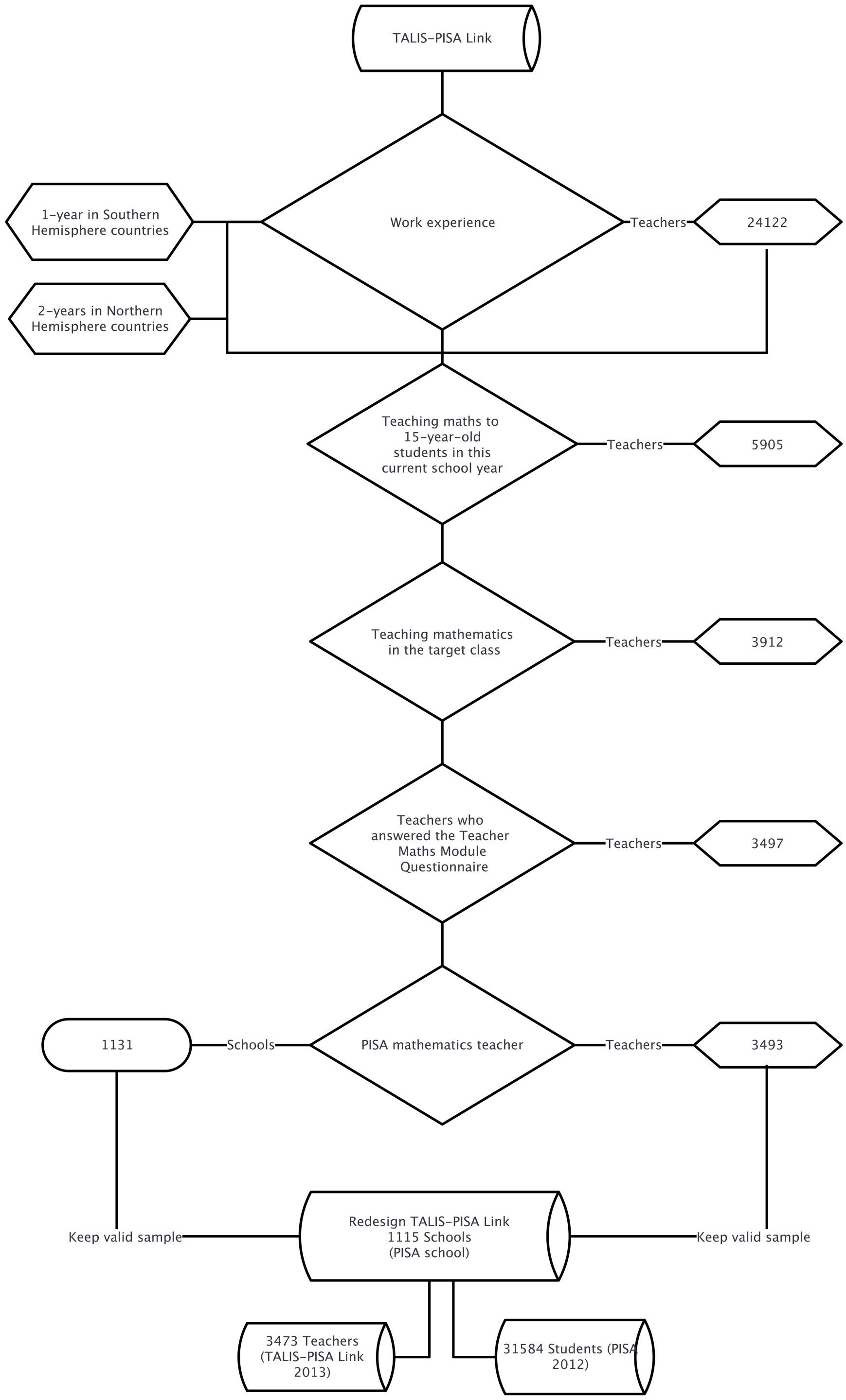

FIGURE 4
The teacher sample selection criteria in the redesigned TALIS-PISA Link database (rTPL).

the identification of strong effects on student achievement (Borko and Putnam, 1995; Timperley et al., 2008). Teacher cooperation seems to be a powerful form of professional development and is regarded as a vital facet of teacher professional practices in the school environment (Goddard et al., 2007; Timperley et al., 2008; Desimone, 2009). In the process of professional communicating and sharing among teachers, improvement-oriented changes seem to develop

TABLE 4 Overview of redesign TALIS-PISA Link database (rTPL).

	AUS	FIN	LVA	MEX	PRT	ROU	SGP	ESP	Total
1-year (S) / 2-year work (N) at same school	2686	2873	1980	1723	2670	2977	4066	5147	24122
Teaching maths to 15-year-old students in this current school year	853	856	315	405	616	516	1163	1181	5905
Teaching maths in the target class	528	370	225	212	569	420	776	812	3912
Teachers who answered the Maths Module Questionnaire	419	332	191	175	537	392	719	732	3497
Pisa Maths Teacher	415	332	191	175	537	392	719	732	3493
Number of teachers in XX Schools	415 in 113	332 in 133	191 in 94	175 in 92	537in 131	392 in 133	719 in 164	732in 271	3 497 in 1 131
Valid number of teachers in the valid same PISASCHOOL	415 in 113	332 in 133	178 in 85	170 in 87	537 in 131	390 in 131	719 in 164	732 in 271	3 473 in 1 115
Number of students in the valid same PISASCHOOL	2 251	4 010	2 013	2 151	3 886	4 103	5 302	7 868	31 584

Source from OECD TALIS 2013 Database and PISA 2012 Database. AUS, Australia; FIN, Finland; LVA, Latvia; MEX, Mexico; PRT, Portugal; ROU, Romania; SGP, Singapore; ESP, Spain.

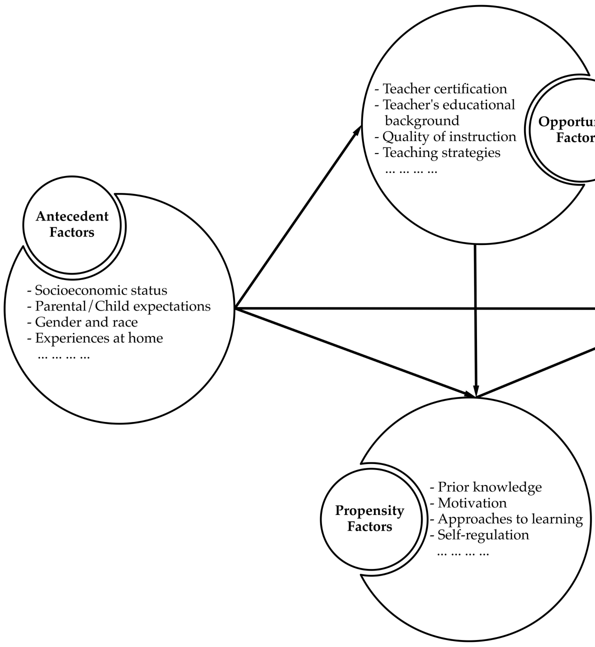

FIGURE 5
The original and the latest combined version of the of Opportunity-Propensity framework (adapted from Byrnes and Miller, 2007; Byrnes and Wasik, 2009; Byrnes and Miller-Cotto, 2016; Byrnes, 2020).

from an evolving knowledge base, professional development, and teacher self-efficacy (Garet et al., 2001; Erickson et al., 2005).

In addition to teacher variables, adding student variables helps look at a more complex way to EER. Considering the socioeconomic status (SES), a meta-analysis study from Sirin (2005) integrated 58 studies published between 1990 and 2000, underpinning the association between SES and academic achievement. A longitudinal study – based on a ten-year window – by Yang Hansen et al. (2011) examined the relations between SES and reading achievement at the individual and school level in Sweden. They found that school differences were highly related to SES differences in 2001, and SES differences did explain more than half of the average reading attainment variation at the school level in 2001, compared to about 30% in 1991. Muijs and Reynolds (2003) analyzed the relationship of SES, classroom social context, classroom organization, teacher behavior and mathematics achievement. Teacher behavior was the strongest performance predictor and was significantly related to student achievement, explaining over 5.6% of the total variance, while individual student background variables explained 3% of the variance in student academic performance. Opdenakker et al. (2002) applied multi-level analyses to examine the associations between SES, gender, average class SES, learning environment, and mathematics attainment at different levels. They concluded that learning environment factors mediated the relationship between individual variables and mathematics attainment. When researching educational effectiveness, plenty of studies suggest that higher-level factors should be considered, such as at the classroom-, teacher- or school-level (Hattie, 2002; Van Ewijk and Sleegers, 2010; Kelly, 2012; Creemers and Kyriakides, 2015; Hornstra et al., 2015; Verhaeghe et al., 2018).

As explained earlier, the present study tests the feasibility of the rTPL against this background by focusing on identifying the relationships between school-level profiles of teacher characteristics (e.g., self-efficacy, beliefs, cooperation) and how each contributes to student mathematics achievement. Additionally, variables mapping socioeconomic status, teacher qualifications (i.e., years of experience, educational background), and school climate (i.e., school size, mutual respect) are used as control variables at the school-level in the analytical procedure. To illustrate the potential of the rTPL database in testing such a model, we selected the Singaporean rTPL data as a case study.

Variables in testing the use of redesigned teaching and learning international survey-programme for international student assessment link

According to the DMEE and the O-P framework, and the literature supports, we selected the variables of student socioeconomic status (PISA 2021 data) and teacher and school characteristics (TALIS 2013). As explained in the former section, TALIS 2013 indicators were aggregated at the school level: *teacher educational background, teacher work experience, teacher self-efficacy, self-efficacy in teaching mathematics, teacher cooperation, effective professional development*, and *constructivist beliefs*. Since no variation was found in the variables of *school composition (i.e., public or private school systems, school location)*, or the variables of *teacher gender and age* between schools in Singapore, they were excluded from the study.

Teacher self-efficacy (TSELEFFS) was defined on the base of three subscales with efficacy in classroom management, efficacy in instruction, and efficacy in student engagement. All scales are built on four-point Likert items, with response categories ranging from "not at all" to "a lot."

Self-efficacy in teaching mathematics (TMSELEFFS) was derived from the TPL instruments presented to mathematics teachers. This is different from the indicator *teacher self-efficacy* and built on statements about teachers' ability to teach mathematics. The four scale items were based on a four-point Likert scale, with response categories ranging from "strongly disagree" to "disagree," "agree" and "strongly agree."

The composite scale, *teacher cooperation (TCOOPS)* consisted of two subscales that centered on exchange and coordination in view of teaching and professional collaboration. Eight six-point Likert scale items were presented with response options ranging from "never" to "once a year or less," "2-4 times a year," "5-10 times a year," "1-3 times a month" and "once a week or more."

Teacher effective professional development (TEFFPROS) focused on the opportunities for active learning and collaborative learning activities or research with other teachers. The four four-point items response options ranged from "not in any activities" to "yes, in all activities."

Constructivist beliefs (TCONSBS) were mapped with four four-point scale items, with response categories ranging from "strongly disagree" to "strongly agree." This index concerned teacher personal beliefs on teaching and learning.

The indicator of *mutual respect (PSCMUTRS)* consisted of four items: school staff have an open discussion about difficulties, mutual respect for colleagues' ideas, a culture of sharing success and the relationships between teacher and student. Items required a response on the base of a four-point scale with response categories ranging from "strongly disagree" to "strongly agree."

The PISA 2012 index of *student economic, social and cultural status (ESCS)* was defined at the student and school level and consisted of three subscales: *the highest parental occupation (HISEI), the highest parental education expressed as years of schooling (PARED),* and *the home possessions (HOMEPOS)*. The HISEI index was coded on the base of ISCO-08 and next mapped onto the international socioeconomic index of occupational status (ISEI) (Ganzeboom, 2010), students' responses to PARED were classified using ISCED (United Nations Educational Scientific and Cultural Organization [UNESCO], 2003).

Other variables included *school size (SCHSIZE)* and the first plausible value for *student mathematics achievement (PIVMATH)* in PISA. PISA 2012 datasets include five plausible values (PV1MATH, PV2MATH, PV3MATH, PV4MATH, PV5MATH) in relation to mathematics literacy, computed by administering 34 mathematics items. It is essential to understand that plausible values are not actual test scores. "They are random numbers that were taken from the distribution of scores that could be reasonably assigned to each individual. Plausible values contain random error variance components and are not as optimal as scores to be used as an indicator of individual student performance. Plausible values are rather suited to describe the performance of the population" (OECD, 2014a). The PISA 2012 plausible values were equated to the PISA scale by utilizing common item equating. In our analytical procedure, the five-combined plausible values and the first plausible value have initially been used and compared. When combined values were used, the separate results of the model parameters across the five datasets were combined using the command TYPE = IMPUTATION in Mplus. The results showed that there was no substantial difference in using multiple plausible values or the first value, either at the individual or the school level. To facilitate the operation of the analysis process in MPlus and its subsequent interpretation, we have used only the first plausible value. Therefore, our further analyses were based on the first plausible value – PV1MATH – as the indicator of individual students' mathematics achievement.

For more detailed information about each scale, see the PISA 2012 technical report (OECD, 2014a) and TALIS 2013 Technical Report (OECD, 2014c).

Analytical methods

Multi-level Path Analysis was applied using Mplus 8.4 (Muthén and Muthén, 2017). The Maximum Likelihood Estimator with robust standard errors (MLR) was used to handle missing and non-normal data. Chi-Square statistics with the degree of freedom and other goodness-of-fit indices (e.g., RMSEA, CFI and SRMR)[4] were used to evaluate whether the model fits the data. When the cut-off value for CFI is greater or equal to 0.95, for RMSEA being less than 0.06, and for SRMR being less than 0.08, the model can be regarded as an acceptable fitting model (Hu and Bentler, 1999).

The interaction correlation coefficient (ICC) is a key tool to check whether the model structure impacts the outcome variable by grouping clusters in multi-level modeling. It also represents the correlation between randomly selected individuals in the same group (Hox et al., 2017). An ICC value exceeding 0.05 indicates that a multi-level structure is needed to model the data (Dyer et al., 2005). R-square represents the proportion of the variance in the dependent variable that is explained by the independent variable and therefore reflects the capability of the model and the predictors to explain a proportion of the variance in the outcome of interest (Finch and Bolin, 2017).

Analytical process

Analysis of variance (ANOVA) model 1 helped estimate the variance within the individuals ($\sigma^2{}_w$) and between the clusters ($\sigma^2{}_B$). These values are used to estimate ICC (ρ), as in Equation (1),

$$\rho = \sigma^2 B/(\sigma^2 w + \sigma^2 B) \tag{1}$$

4 RMSEA is an absolute measure of model fit, which stands for Root Mean Square Error of Approximation. CFI is short for Comparative Fit Index. Both RMSEA and CFI pay the penalty for model complexity. SRMR (Standardized Root Mean Square Residual) measures the absolute model fit.

In the Random Intercept with Level-1 Predictor model (model 2), the subscript *i* refers to the individual in the *j* school-cluster; ε_{ij} and μ_{oj} are error terms at Level-1 and Level-2; β_{oj} is the intercept of achievement for each school; γ_{00} represents an average intercept value across schools. The predictors of student variables in PISA, *student economic, social and cultural status (ESCS)* and *mathematics achievement (PV1MATH)* were added at the student-level. We estimated the values for the two fixed effects of *level-2 PV1MATH* (γ_{00}) and *ESCS* (γ_{10}) as well as for the residual variance of *PVIMATH* (μ_{oj}) and other predictors (ε_{ij}). The equation for model 2 is given by:

$$PV1MATH_{ij} = \gamma_{00} + \gamma_{10} ESCS_{ij} + \mu_{oj} + \varepsilon_{ij} \tag{2}$$

In model 3, we added school-level variables to ascertain how much variation in *PVIMATH* was present across schools. Specific TALIS and PISA data were entered in this model. Student *ESCS* (PISA 2012) was used as a predictor at the student-level. School size (*SCHSIZE*) and school *ESCS* in PISA 2012, teacher characteristics (e.g., *TSELFEFFS, TCOOPS, TEFFPROS, TMSELEFFS, TCONSBS*) and mutual respect (*PSCMUTRS*) in TALIS 2013 were entered as predictors at the school-level.

Results

The ANOVA model results help estimate the variance of student mathematics achievement. This is 0.407 and 0.708 at the individual- and between-level, respectively; thus, the value for ICC is estimated as 0.37 based on Equation 1. The value indicates that the correlation of the mathematics achievement among students within the same schools is 0.37, and about 37% of the variability of student mathematics achievement can be explained by schools' diversity in Singapore.

The goodness-of-fit indices of model 2 are satisfactory: CFI = 0.95, RMSEA = 0.03, and SRMR = 0.04. The estimated slope for *ESCS* is 0.20 and is significantly associated with *PV1MATH,* indicating that as ESCS score increased by 1 point, the mathematics achievement shows an associated increase by an estimated 0.20 points.

Figure 6 presents the results for model 3, with the indices CFI being 1.00, RMSEA = 0.01, within-level SRMR = 0.00, and between-level SRMR = 0.02. At the individual student-level, the indicator of *student economic, social and cultural status (ESCS)* reflects significant and positive associations with mathematics achievement considering an estimated slope of 0.21. The R-square of outcome variable PVIMATH is about 0.05 at the individual student-level, stating that around 4% variation of student mathematics achievement can be explained within the schools, accounting for 1.9% (ICC * R^2 = 0.37*0.04) of the total variance (ICC = 0.37).

At the school level, as shown in **Figure 6**, four indicators do positively and directly contribute to student achievement:

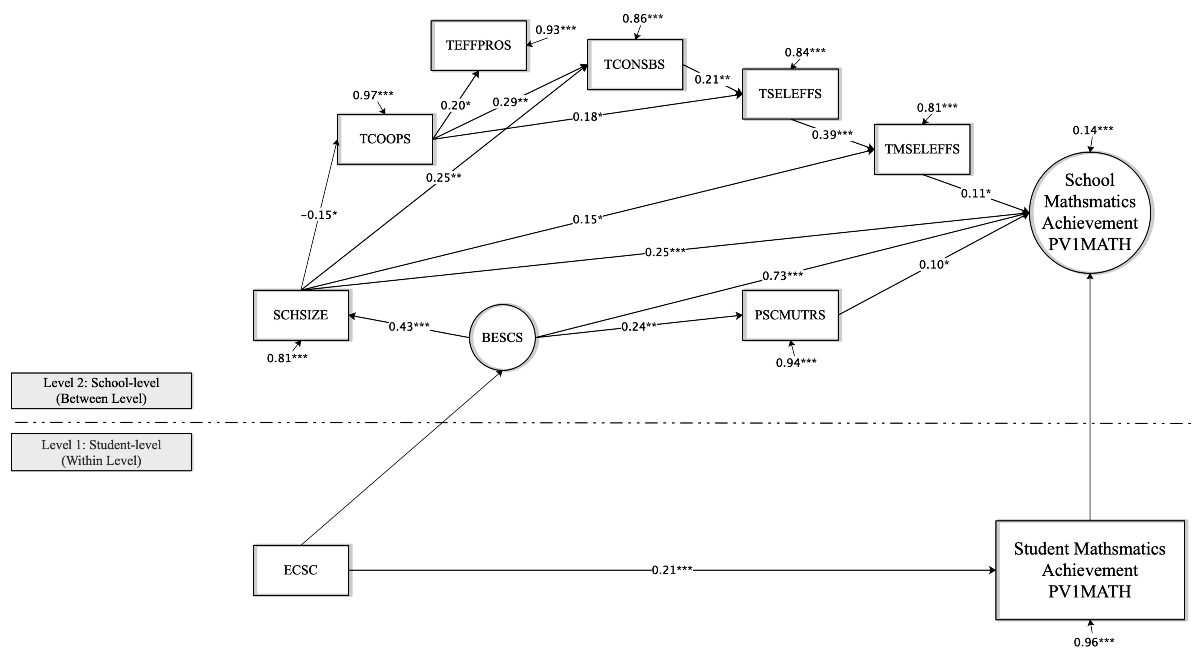

FIGURE 6
The standardized model results of the relationships between teacher characteristics and student mathematics achievement at the student- and school-level. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$.

school SES *economic, social and cultural status (BESCS, 0.73), self-efficacy in teaching mathematics (TMSELEFFS, 0.11), school size (SCHSIZE, 0.25)* and *mutual respect (PSCMUTRS, 0.10).* It is interesting to observe that a higher mutual respect working environment *(i.e., school staff have an open discussion about difficulties, mutual respect for colleagues' ideas, a culture of sharing success, and the relationships between teacher and student)* is associated with higher academic performance. The predictors explain 86% of the variation in mathematics achievement between schools, accounting for about 32% (ICC * R^2 = 0.37*0.32) of the total variance (ICC = 0.37).

In Singapore, teachers working in relatively small schools (-0.15) prefer cooperating with other colleagues. In turn, *teacher cooperation (TCOOPS)* is positively correlated with *teacher self-efficacy (TSELEFFS, 0.18), effective professional development (TEFFPRO, 0.20S)* and *constructivist beliefs (TCONSBS, 0.29). Teacher self-efficacy (TSELEFFS)* helps predict *constructivist beliefs (TCONSBS),* with a standardized coefficient of 0.21. As a result, *teacher self-efficacy (TSELEFFS) is* directly associated with *self-efficacy in teaching mathematics (TMSELEFFS, 0.39).*

Also, *school size (SCHSIZE)* seems significantly and positively correlated with *constructivist beliefs (TCONSBS, 0.25)* and *self-efficacy in teaching mathematics (TMSELEFFS, 0,15). School economic, social and cultural status (BESCS)* positively contributes to *school size (SCHSIZE, 0.43)* and *mutual respect (PSCMUTRS, 0.24).*

In summary, the results in **Figure 6** show the direct and indirect factors that are significantly related to mathematics achievement in Singapore. Student mathematics achievement vary according to students with different socioeconomic status. School socioeconomic status, school size, school collective mathematics teachers' teaching self-efficacy, and mathematics teacher mutual respect are positively and significantly related to the school's mathematics performance.

Discussion

The present study aimed to study the educational effectiveness from a multi-level perspective by building on a newly designed linkage database, connecting student and teacher data collected *via* TALIS and PISA. Applying the linkage dataset was expected to help unravel the interconnections between students' mathematics performance while considering individual learner characteristics, mathematics teacher variables in the teaching/classroom environment, and the school-level variables. As stated earlier, this requires new and adequate teacher sampling procedures.

Building on the rTPL based analysis results, our findings help operationalize specific school variables, teaching style elements, and culture-related constructs that play a significant role. These – exemplary – findings could become ingredients to inspire instructional policies to foster quality measures at the different levels in the model. This could also, on the one hand, promote a school mathematics culture and related instructional approaches in view of improving mathematics teaching effectiveness and student learning outcomes. On the other hand, this could also foster between-country comparison to identify explanatory variables building on differences in mathematics curricula, school context, and educational systems.

The case study analysis results demonstrate the feasibility and potential for linking TALIS and PISA. The findings suggest that, in Singapore, schools with highly educated teachers and higher self-efficacy teachers in teaching mathematics contributed to improving schools' mathematics performance. The results of the case study will not be discussed in-depth. Still, nevertheless, some aspects are noteworthy since they complement previous research and further illustrate the potential of the rTPL database. This is tackled in the next paragraphs.

Available research considers teacher background and teacher characteristics as critical differences between teachers in classrooms (Fraser, 2013; Creemers and Kyriakides, 2015). However, studies rarely focus on looking at the effects of these differences on student learning outcomes. Even the Dynamic Model of Educational Effectiveness primarily concentrates on teaching activities (e.g., classroom management of time, classroom climate, teaching-modeling, assessment) to study student learning outcomes (Kyriakides et al., 2020). Teacher background and teacher characteristics are mostly approached as teacher-level input variables when studying teaching effectiveness/instructional quality, e.g., (Scheerens, 2007; Creemers and Kyriakides, 2015).

International large-scale assessments have the potential to boost multi-level analysis studies that fit state-of-the-art educational effectiveness models. Nevertheless – as tackled in the present paper – this potential is often flawed by methodological constraints in the data available for the studies. The present paper stated solutions, procedures, and strategies to develop overarching databases that link datasets from earlier studies; more specifically, TALIS 2013 and PISA 2012. Using the redesigned TALIS-PISA Link (rTPL) dataset to evaluate student mathematics achievement in Singapore, we provided the feasibility of developing a multi-level perspective from the teacher self-reported data and the student survey. Compared to available studies linking TALIS and PISA data, we extended this new avenue of research by developing a linked database that is geared to the specific mathematics teaching and learning domain. The rTPL considered specific inclusion and exclusion criteria to construct a better fitting database to reflect the school mathematics educational environment.

The further potential of the rTPL is to center between-country comparisons when explaining differences in learning performance. International large-scale assessments studies suggest that the theoretical constructs are "universal" and apply to all countries. This introduces the question of whether relationships put forward in specific national contexts do hold in other countries. International comparison studies might help identify factors associated with differences in mathematics achievement in each country and test measurement invariance to check the comparability in the eight national contexts that are contained in the rTPL dataset. Meanwhile, the three-level model can be conducted to examine which country-level profile of teacher and school factors appear to play a role in predicting or explaining student mathematics achievement. We found that about 22% variation of achievement varies across schools, and around 19% vary across eight participating countries.

Although the present study offers valuable insights into linking TALIS and PISA, the rTPL dataset reflects some apparent limitations. In TALIS, we miss student-level data, and in PISA, we miss specific teacher-level data that can be related to the unique student data. This was tackled by aggregating data at the school level. Therefore, it is not possible to look at the impact of unique characteristics situated within and between classroom settings in a school. Specific teaching style approaches and unique classroom composition effects cannot be identified. Several statistical and conceptual challenges should be taken into account when using the rTPL dataset: the original number of schools, teachers, and pupils participating in TALIS 2013 and PISA 2012 is far larger than the number in the rTPL dataset, and this affects the weights to be used when looking at values in the database.

The next thing to consider is how to solve the time gap in the TALIS 2013 and PISA 2012 administration and the way we selected teachers with at least one or two years of experience, depending on the hemisphere. This resulted in a smaller sample of schools, teachers, and pupils; but could also have harmed the representativeness of the final sample. For instance, the teacher sample of Mexico and Latvia was reduced to less than 200, while in other countries, more teachers could be retained in the rTPL sample. This smaller sample size could result in a loss of statistical power. Since the rTPL dataset will contain only data from eight countries, this also affects the extent to which we can generalize findings.

Additionally, the current study focuses on exploring the linking and possible use of the two databases. Regarding the case study, we emphasize using TALIS data to explain the achievement at the school level but less considering individual factors at the student level. In the subsequent studies, the student-related indicators, such as mathematics self-efficacy, and mathematics anxiety, could be considered.

Conclusion

The current study aimed to develop a linked database geared to mathematics teaching and learning to reflect the school mathematics educational environment. Taking into the subject-specific characteristics of mathematics education, we extend a recent new avenue of research by (re)developing a linked database geared to the specific mathematics teaching and learning domain to reflect the school mathematics educational environment. The redesigned linkage dataset connects student and teacher data collected *via* TALIS 2013 and PISA 2012. It explores how to link TALIS and PISA data to study the dynamic relations between multiple educational effectiveness factors and student achievement as reflected in the Dynamic Model of Educational Effectiveness and the Opportunity-Propensity framework. Data from seven educational systems are used in this linkage process, and four critical issues related to five selection criteria are considered to address the specific sample of mathematics teachers. A case study, using Singapore linkage data through Multilevel Path Analysis demonstrated the feasibility and potential of exploring school effectiveness on the base of this new data set. Meanwhile, we pointed out that the Redesigned TALIS 2013 and PISA 2012 data presented challenges in terms of identifying a linkage variable, the aggregation of variables, and a sample selection procedure to identify the relevant mathematics teachers.

Student learning outcomes are the product of teachers and teaching, schools, educational systems, and students' diverse background characteristics, e.g., (Kyriakides and Luyten, 2009; Kyriakides et al., 2020). The current study provided new perspectives to understand this complex relationship while using a newly designed database of TALIS 2013 and PISA 2012. The design of the rTPL presented challenges in terms of identifying a linkage variable, the aggregation of variables, and a sample selection procedure to identify the relevant mathematics teachers.

Taken together, the study approach potentially stimulates future research about multi-level perspectives on PISA students' mathematics learning outcomes in various national contexts building on the EER dynamic model. A next avenue was suggested to focus on the international comparison of the relationships in the EER model. Also, the study could inspire future attempts linking data from TALIS 2018 and PISA 2018, with a focus on reading literacy as the primary domain. Nine countries participated in the TALIS-PISA Link 2018. Since both studies were administered in the same year, some drawbacks of the current linking approach do not apply. A collaboration with other researchers in view of this new endeavor is welcomed to tackle the methodological challenges and study the richness of the Dynamic Model of Educational Effectiveness and Opportunity-Propensity framework.

Data availability statement

The original contributions presented in this study are publicly available. The datasets for this study can be found at https://www.oecd.org/education/talis/talis-2013-data.htm (TALIS 2013 data) and https://www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm (PISA 2012 data).

Ethics statement

The Organization for Economic Co-operation and Development (OECD) reviewed and approved the studies involving human participants. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

XL was responsible for the theoretical framework and a major contributor in writing the manuscript. MV, KYH, and JDN supervised the research project together, ensured the article's coherence, and offers guidance on misunderstanding text. MV set out the objectives of the project and gave feedback during all phases. KYH had been responsible for the accuracy and logic of any part of the work and provides comments to the overall text. All authors put joint effort for this article, read, and approved the final manuscript.

Funding

This research was fully funded by the China Scholarship Council (CSC), grant number: CSC201807930019 and partially funded by the Fonds Wetenschappelijk Onderzoek (FWO), grant number: V412020N.

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

Ainley, J., Pratt, D., and Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *Br. Educ. Res. J.* 32, 23–38. doi: 10.1080/01411920500401971

Anghileri, J. (2006). Scaffolding practices that enhance mathematics learning. *J. Math. Teach. Educ.* 9, 33–52. doi: 10.1007/s10857-006-9005-9

Antoniou, P., and Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teach. Teach. Educ.* 29, 1–12. doi: 10.1016/j.tate.2012.08.001

Austin, B., Adesope, O. O., French, B. F., Gotch, C., Bélanger, J., and Kubacka, K. (2015). *Examining School Context and its Influence on Teachers: Linking TALIS 2013 with PISA 2012 Student Data. Education Working Paper No. 115*. Paris: OECD Publishing.

Ball, D. L., Hill, H. C., and Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *Am. Educ.* 29, 14–17; 20–22; 43–46.

Ball, D. L., Thames, M. H., and Phelps, G. (2008). Content knowledge for teaching: What makes it special? *J. Teach. Educ.* 59, 389–407. doi: 10.1177/0022487108324554

Blömeke, S., Suhl, U., Kaiser, G., and Döhrmann, M. (2012). Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teach. Teach. Educ.* 28, 44–55. doi: 10.1016/j.tate.2011.08.006

Borko, H., and Putnam, R. T. (1995). "Expanding a teacher's knowledge base: a cognitive psychological perspective on professional development," in *Professional Development in Education: New Paradigms and Practices*, eds T. R. Guskey and M. Huberman (New York, NY: Teachers College Press), 35–65.

Byrnes, J. P. (2003). Factors predictive of mathematics achievement in white, black, and Hispanic 12th graders. *J. Educ. Psychol.* 95, 316–326. doi: 10.1037/0022-0663.95.2.316

Byrnes, J. P. (2020). The potential utility of an opportunity-propensity framework for understanding individual and group differences in developmental outcomes: A retrospective progress report. *Dev. Rev.* 56:100911. doi: 10.1016/j.dr.2020.100911

Byrnes, J. P., and Miller, D. C. (2007). The relative importance of predictors of math and science achievement: An opportunity-propensity analysis. *Contemp. Educ. Psychol.* 32, 599–629. doi: 10.1016/j.cedpsych.2006.09.002

Byrnes, J. P., and Miller-Cotto, D. (2016). The growth of mathematics and reading skills in segregated and diverse schools: An opportunity-propensity analysis of a national database. *Contemp. Educ. Psychol.* 46, 34–51. doi: 10.1016/j.cedpsych.2016.04.002

Byrnes, J. P., and Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity-propensity analysis. *Contemp. Educ. Psychol.* 34, 167–183. doi: 10.1016/j.cedpsych.2009.01.002

Caprara, G. V., Barbaranelli, C., Steca, P., and Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *J. Schl Psychol.* 44, 473–490. doi: 10.1016/j.jsp.2006.09.001

Caro, D. H., Lenkeit, J., and Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Stud. Educ. Eval.* 49, 30–41. doi: 10.1016/j.stueduc.2016.03.005

Chapin, S. H., and O'Connor, C. (2007). "Academically productive talk: Supporting students' learning in mathematics," in *The Learning of Mathematics*, Vol. 69, eds W. G. Martin, M. E. Strutchens, and P. C. Elliott (Reston, VA: National Council of Teachers of Mathematics), 113–128.

Chapman, C., Muijs, D., Reynolds, D., Sammons, P., and Teddlie, C. (2015). *The Routledge International Handbook of Educational Effectiveness and Improvement: Research, Policy, and Practice*. London: Routledge. doi: 10.4324/9781315679488

Cordero Ferrera, J. M., and Gil-Izquierdo, M. (eds) (2016). "TALIS-PISA link: guidelines for a robust quantitative analysis," in *Proceedings of the International Conference on Qualitative and Quantitative Economics Research (QQE)*, (Singapore: Global Science and Technology Forum). doi: 10.5176/2251-2012_QQE16.19

Cordero, J. M., and Gil-Izquierdo, M. (2018). The effect of teaching strategies on student achievement: An analysis using TALIS-PISA-link. *J. Policy Model.* 40, 1313–1331. doi: 10.1016/j.jpolmod.2018.04.003

Creemers, B. P. M., and Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge.

Creemers, B. P., and Scheerens, J. (1994). Developments in the educational effectiveness research programme. *Int. J. Educ. Res.* 21, 125–140. doi: 10.1016/0883-0355(94)90028-0

Creemers, B., and Kyriakides, L. (2007). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge. doi: 10.4324/9780203939185

Creemers, B., and Kyriakides, L. (2015). Developing, testing, and using theoretical models for promoting quality in education. *Schl Effect. Schl Improv.* 26, 102–119. doi: 10.1080/09243453.2013.869233

Delprato, M., and Chudgar, A. (2018). Factors associated with private-public school performance: Analysis of TALIS-PISA link data. *Int. J. Educ. Dev.* 61, 155–172. doi: 10.1016/j.ijedudev.2018.01.002

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educ. Res.* 38, 181–199. doi: 10.3102/0013189X08331140

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., and Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educ. Eval. Policy Anal.* 24, 81–112. doi: 10.3102/01623737024002081

Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Stud. Educ. Eval.* 28, 347–368. doi: 10.1016/S0191-491X(02)00043-3

Dyer, N. G., Hanges, P. J., and Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *Leadersh. Q.* 16, 149–167. doi: 10.1016/j.leaqua.2004.09.009

Echazarra, A., Salinas, D., Méndez, I., Denis, V., and Rech, G. (2016). *How Teachers Teach and Students Learn: Successful Strategies for School. OECD Education Working Papers, No 130*. Paris: OECD Publishing.

Erickson, G., Brandes, G. M., Mitchell, I., and Mitchell, J. (2005). Collaborative teacher learning: Findings from two professional development projects. *Teach. Teach. Educ.* 21, 787–798. doi: 10.1016/j.tate.2005.05.018

Fernández-Díaz, M. J., Rodríguez-Mantilla, J. M., and Martínez-Zarzuelo, A. (2016). PISA y TALIS ¿congruencia o discrepancia? *RELIEVE - Rev. Electr. Investig. Eval. Educ.* 22:9. doi: 10.7203/relieve.22.1.8247

Finch, H., and Bolin, J. (2017). *Multilevel Modeling using Mplus*. Boca Raton, FL: CRC Press. doi: 10.1201/9781315165882

Fraser, B. J. (2013). "Classroom learning environments," in *Handbook of Research on Science Education*, (London: Routledge), 117–138.

Ganzeboom, H. B. (ed.) (2010). "A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007," in *Paper Presented at the Annual Conference of International Social Survey Programme*, (Lisbon).

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., and Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *Am. Educ. Res. J.* 38, 915–945. doi: 10.3102/00028312038004915

Gil-Izquierdo, M., and Cordero, J. M. (2018). Guidelines for data fusion with international large scale assessments: Insights from the TALIS-PISA link database. *Stud. Educ. Eval.* 59, 10–18. doi: 10.1016/j.stueduc.2018.02.002

Goddard, Y. L., Goddard, R. D., and Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teach. Coll. Record.* 109, 877–896. doi: 10.1177/016146810710900401

Grossman, P. L. (1990). *The Making of a Teacher: Teacher Knowledge and Teacher Education*. New York, NY: Teachers College Press.

Hattie, J. (2002). Classroom composition and peer effects. *Int. J. Educ. Res.* 37, 449–481. doi: 10.1016/S0883-0355(03)00015-6

Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Abingdon: Taylor & Francis. doi: 10.4324/9780203887332

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cogn. Instruc.* 26, 430–511. doi: 10.1080/07370000802177235

Hill, H. C., Rowan, B., and Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *Am. Educ. Res. J.* 42, 371–406. doi: 10.3102/00028312042002371

Hill, H. C., Schilling, S. G., and Ball, D. L. (2004). Developing of teachers' measures mathematics knowledge for teaching. *Elem. Schl J.* 105, 11–30. doi: 10.1086/428763

Holzberger, D., Philipp, A., and Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *J. Educ. Psychol.* 105, 774–786. doi: 10.1037/a0032198

Hornstra, L., van der Veen, I., Peetsma, T., and Volman, M. (2015). Does classroom composition make a difference: Effects on developments in motivation, sense of classroom belonging, and achievement in upper primary school. *Schl Effect. Schl Improv.* 26, 125–152. doi: 10.1080/09243453.2014.887024

Hox, J. J., Moerbeek, M., and Van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications*. London: Routledge. doi: 10.4324/9781315650982

Hu, L. T., 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

Huang, J. L., Tang, Y. P., He, W. J., and Li, Q. (2019). Singapore's school excellence model and student learning: Evidence from PISA 2012 and TALIS 2013. *Asia Pacif. J. Educ.* 39, 96–112. doi: 10.1080/02188791.2019.1575185

Hunter, R. (2005). "Reforming communication in the classroom: One teacher's journey of change," in *Building Connections: Research, Theory and Practice (Proceedings of the Annual Conference of the Mathematics Education Research Group of Australasia)*, eds P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, et al. (Melbourne: MERGA), 451–458.

Kaplan, D., and McCarty, A. T. (2013). Data fusion with international large scale assessments: A case study using the OECD PISA and TALIS surveys. *Large Scale Assess. Educ.* 1:6. doi: 10.1186/2196-0739-1-6

Kelly, A. (2012). Measuring equity' and equitability' in school effectiveness research. *Br. Educ. Res. J.* 38, 977–1002. doi: 10.1080/01411926.2011.605874

Kleickmann, T., Tröbst, S., Heinze, A., Bernholt, A., Rink, R., and Kunter, M. (2017). "Teacher knowledge experiment: conditions of the development of pedagogical content knowledge," in *Competence Assessment in Education*, eds D. Leutner, J. Fleischer, J. Grünkorn, and E. Klieme (Cham: Springer), 111–129. doi: 10.1007/978-3-319-50030-0_8

Klieme, E. (2013). "The role of large-scale assessments in research on educational effectiveness and school development," in *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*, eds M. von Davier, E. Gonzalez, I. Kirsch, and K. Yamamoto (Dordrecht: Springer), 115–147. doi: 10.1007/978-94-007-4629-9_7

Kyriakides, L., and Luyten, H. (2009). The contribution of schooling to the cognitive development of secondary education students in Cyprus: An application of regression discontinuity with multiple cut-off points. *Schl Effect. Schl Improv.* 20, 167–186. doi: 10.1080/09243450902883870

Kyriakides, L., Christoforou, C., and Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teach. Teach. Educ.* 36, 143–152. doi: 10.1016/j.tate.2013.07.010

Kyriakides, L., Creemers, B. P. M., Muijs, D., Rekers-Mombarg, L., Papastylianou, D., Van Petegem, P., et al. (2014). Using the dynamic model of educational effectiveness to design strategies and actions to face bullying. *Schl Effect. Schl Improv.* 25, 83–104. doi: 10.1080/09243453.2013.771686

Kyriakides, L., Creemers, B. P. M., Panayiotou, A., and Charalambous, E. (2020). *Quality and Equity in Education: Revisiting Theory and Research on Educational Effectiveness and Improvement*. London: Routledge. doi: 10.4324/9780203732250

Kyriakides, L., Creemers, B., Antoniou, P., and Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *Br. Educ. Res. J.* 36, 807–830. doi: 10.1080/01411920903165603

Kyriakides, L., Georgiou, M. P., Creemers, B. P. M., Panayiotou, A., and Reynolds, D. (2017). The impact of national educational policies on student achievement: A European study. *Schl Effect. Schl Improv.* 29, 171–203. doi: 10.1080/09243453.2017.1398761

Le Donné, N., Fraser, P., and Bousquet, G. (2016). *Teaching Strategies for Instructional Quality: Insights from the TALIS-PISA Link Data. OECD Education Working Papers No. 148.* Paris: OECD Publishing, doi: 10.1787/5jln1hlsr0lr-en

Leunda Iztueta, I., Garmendia Navarro, I., and Etxeberria Murgiondo, J. (2017). Statistical matching in practice – An application to the evaluation of the education system from PISA and TALIS. *Rev. Investig. Educ.* 35, 371–388. doi: 10.6018/rie.35.2.262171

Ma, L. (2010). *Knowing and Teaching Elementary Mathematics: Teachers' Understanding of Fundamental Mathematics in China and the United States.* London: Routledge. doi: 10.4324/9780203856345

Mammadov, R., and Cimen, I. (2019). Optimizing teacher quality based on student performance: A data envelopment analysis on PISA and TALIS. *Int. J. Instruc.* 12, 767–788. doi: 10.29333/iji.2019.12449a

Martínez-Abad, F., Gamazo, A., and Rodríguez-Conde, M.-J. (2020). Educational data mining: Identification of factors associated with school effectiveness in PISA assessment. *Stud. Educ. Eval.* 66:100875. doi: 10.1016/j.stueduc.2020.100875

Muijs, D., and Reynolds, D. (2002). Teachers' beliefs and behaviors: What really matters? *J. Classroom Interact.* 37, 3–15.

Muijs, D., and Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics: A longitudinal study. *Educ. Res. Eval.* 9, 289–314. doi: 10.1076/edre.9.3.289.15571

Muijs, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H., and Earl, L. (2014). Teaching effectiveness. A state of the art review. *Schl Effect. Schl Improv.* 24, 231–256. doi: 10.1080/09243453.2014.885451

Muthén, L. K., and Muthén, B. (2017). *Mplus User's Guide: Statistical Analysis with Latent Variables*. New York, NY: Wiley.

Nilsen, T., and Gustafsson, J.-E. (2016). *Teacher Quality, Instructional Quality and Student Outcome. Relationships Across Countries, Cohorts and Time*. Berlin: Springer. doi: 10.1007/978-3-319-41252-8

OECD (2009). *Creating Effective Teaching and Learning Environments First Results from TALIS*. Paris: OECD Publishing. doi: 10.1787/9789264068780-en

OECD (2010). *TALIS 2008 Technical Report*. Paris: OECD Publishing. doi: 10.1787/9789264079861-en

OECD (2013a). *PISA 2012 Assessment and Analytical Framework*. Paris: OECD Publishing.

OECD (2013b). *TALIS 2013: Conceptual Framework*. Paris: OECD Publishing.

OECD (2014a). *PISA 2012 Technical Report*. Paris: OECD Publishing.

OECD (2014b). *TALIS 2013 Results: An International Perspective on Teaching and Learning*. Paris: OECD Publishing.

OECD (2014c). *TALIS 2013 Technical Report*. Paris: OECD Publishing.

OECD (2017). *PISA 2015 Assessment and Analytical framework*. Paris: OECD Publishing. doi: 10.1787/9789264281820-en

OECD (2019a). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing.

OECD (2019b). *TALIS 2018 Technical Report*. Paris: OECD Publishing.

Opdenakker, M. C., and Van Damme, J. (2000). The importance of identifying levels in multilevel analysis: An illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. *Schl Effect. Schl Improv.* 11, 103–130. doi: 10.1076/0924-3453(200003)11:1;1-#;FT103

Opdenakker, M. C., and Van Damme, J. (2006a). Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools. *Schl Effect. Schl Improv.* 17, 87–117. doi: 10.1080/09243450500264457

Opdenakker, M. C., and Van Damme, J. (2006b). Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teach. Teach. Educ.* 22, 1–21. doi: 10.1016/j.tate.2005.07.008

Opdenakker, M. C., and Van Damme, J. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *Br. Educ. Res. J.* 33, 179–206. doi: 10.1080/01411920701208233

Opdenakker, M. C., Van Damme, J., De Fraine, B., Van Landeghem, G., and Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *Schl Effect. Schl Improv.* 13, 399–427. doi: 10.1076/sesi.13.4.399.10283

Panayiotou, A., Kyriakides, L., and Creemers, B. P. (2016). Testing the validity of the dynamic model at school level: A European study. *Schl Leadersh. Manag.* 36, 1–20. doi: 10.1080/13632434.2015.1107537

Panayiotou, A., Kyriakides, L., Creemers, B. P., McMahon, L., Vanlaar, G., Pfeifer, M., et al. (2014). Teacher behavior and student outcomes: Results of a European study. *Educ. Assess. Eval. Acc.* 26, 73–93. doi: 10.1007/s11092-013-9182-x

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., et al. (2014). Educational effectiveness research (EER): A state-of-the-art review. *Schl Effect. Schl Improv.* 25, 197–230. doi: 10.1080/09243453.2014.885450

Rimm-Kaufman, S. E., Storrn, M. D., Sawyer, B. E., Pianta, R. C., and LaParo, K. M. (2006). The teacher belief Q-sort: A measure of teachers' priorities in relation to disciplinary practices, teaching practices, and beliefs about children. *J. Schl Psychol.* 44, 141–165. doi: 10.1016/j.jsp.2006.01.003

Robardey, C. P., Allard, D. W., and Brown, D. M. (1994). An assessment of the effectiveness of full option science system training for third-through sixth-grade teachers. *J. Elem. Sci. Educ.* 6, 17–29. doi: 10.1007/BF03170647

Rutkowski, L., von Davier, M., and Rutkowski, D. (2013). *Handbook of International Large-Scale Assessment*. New York, NY: Chapman and Hall. doi: 10.1201/b16061

Scheerens, J. (2007). *Conceptual Framework for the PISA 2009 Context Questionnaires and Thematic Reports*. Oslo: PISA Governing Board.

Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *Schl Effect. Schl Improv.* 24, 1–38. doi: 10.1080/09243453.2012.691100

Scheerens, J., and Bosker, R. (1997). *The Foundations of Educational Effectiveness*. Oxford: Pergamon.

Sealy, K. M., Perry, S. M., and DeNicola, T. (2016). *Relationships and Predictors of Principal Job Satisfaction Across Multiple Countries: A Study using TALIS 2013 and PISA 2012*. Rochester, NY: SSRN. doi: 10.2139/ssrn.2717056

Senk, S. L., Tatto, M. T., Reckase, M., Rowley, G., Peck, R., and Bankov, K. (2012). Knowledge of future primary teachers for teaching mathematics: An international comparative study. *ZDM Math. Educ.* 44, 307–324. doi: 10.1007/s11858-012-0400-7

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harv. Educ. Rev.* 57, 1–23. doi: 10.17763/haer.57.1.j463w79r56455411

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educ. Res.* 15, 4–14. doi: 10.3102/0013189X015002004

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Rev. Educ. Res.* 75, 417–453. doi: 10.3102/00346543075003417

Snow-Renner, R., and Lauer, P. A. (2005). *Professional Development Analysis. McREL Insights*. Aurora, CO: Mid-continent Research for Education and Learning.

Stacey, K. (ed.) (2015). "The international assessment of mathematical literacy: PISA 2012 framework and items," in *Proceedings of the 12th International Congress on Mathematical Education*, (Seoul). doi: 10.1007/978-3-319-17187-6_43

Strietholt, R., and Scherer, R. (2018). The contribution of international large-scale assessments to educational research: Combining individual and institutional data sources. *Scand. J. Educ. Res.* 62, 368–385. doi: 10.1080/00313831.2016.1258729

Timperley, H., Wilson, A., Barrar, H., and Fung, I. (2008). *Teacher Professional Learning and Development*. Wellington, NZ: Ministry of Education.

Tschannen-Moran, M., and McMaster, P. (2009). Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy. *Elem. Schl J.* 110, 228–245. doi: 10.1086/605771

United Nations Educational Scientific and Cultural Organization [UNESCO] (2003). *International Standard Classification of Education, ISCED 1997*. Paris: UNESCO.

Van Damme, J., Liu, H., Vanhee, L., and Pustjens, H. J. E. E. (2010). Longitudinal studies at the country level as a new approach to educational effectiveness: Explaining change in reading achievement (PIRLS) by change in age, socio-economic status and class size. *Effect. Educ.* 2, 53–84. doi: 10.1080/19415531003616888

Van Ewijk, R., and Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educ. Res. Rev.* 5, 134–150. doi: 10.1016/j.edurev.2010.02.001

Verhaeghe, J. P., Vanlaar, G., Knipprath, H., De Fraine, B., and Van Damme, J. (2018). Can group composition effects explain socioeconomic and ethnic achievement gaps in primary education? *Stud. Educ. Eval.* 57, 6–15. doi: 10.1016/j.stueduc.2017.07.006

Wagemaker, H. (2014). "International large-scale assessments: From research to policy," in *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*, eds L. Rutkowski, M. von Davier, and D. Rutkowski (London: CRC Press), 11–36.

Wagemaker, H. (2020). *Reliability and Validity of International Large-Scale Assessment: Understanding IEA's Comparative Studies of Student Achievement*. Cham: Springer. doi: 10.1007/978-3-030-53081-5

Watson, A., and Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Math. Think. Learn.* 8, 91–111. doi: 10.1207/s15327833mtl0802_1

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychol. Bull.* 91, 461–481. doi: 10.1037/0033-2909.91.3.461

Yang Hansen, K., Rosén, M., and Gustafsson, J. E. (2011). Changes in the multi-level effects of socio-economic status on reading achievement in Sweden in 1991 and 2001. *Scand. J. Educ. Res.* 55, 197–211. doi: 10.1080/00313831.2011.554700

You, H. S., Park, S., and Delgado, C. (2021). A closer look at US schools: What characteristics are associated with scientific literacy? A multivariate multilevel analysis using PISA 2015. *Sci. Educ.* 105, 406–437. doi: 10.1002/sce.21609