Instructional Science
https://doi.org/10.1007/s11251-023-09652-6

ORIGINAL RESEARCH

Teacher versus student perspectives on instructional quality in mathematics education across countries

Xin Liu[1,2,5] · Kajsa Yang Hansen[3] · Jan De Neve[4] · Martin Valcke[1]

Received: 18 July 2022 / Accepted: 22 September 2023

Abstract

The present study examines the measurement property of instructional quality in mathematics education, building on data from teachers and students, by combing TALIS 2013 and PISA 2012 linkage data from seven countries. Confirmatory factor analysis was applied to examine the dimensionality of the construct instructional quality in mathematics instruction. Three dimensions were identified (i.e., classroom disciplinary climate, teacher support, and cognitive activation) when building on teacher data from TALIS. This three-dimensional model did not fit all countries. When analyzing PISA data, the same three dimensions could be identified, but two additional dimensions appeared: classroom management and student-orientated instruction. This five-dimensional factor structure reflected metric invariance across all countries. The findings imply that students and teachers seem to hold different perceptions about mathematics instructional quality reflect different dimensions. These differences seem to vary within and between countries. This implies that care should be taken when using the construct as an equivalent measure of instructional quality when studying school effectiveness in mathematics education across countries.

Keywords Mathematics instructional quality · Mathematics education · Measurement invariance · Confirmatory factor analysis · TALIS-PISA link

✉ Xin Liu
Xin.Liu@ils.UiO.No; Xin.Liu@UGent.be

[1] Department of Educational Studies, Ghent University, Ghent, Belgium

[2] Department of Teacher Education and School Research, University of Oslo, Oslo, Norway

[3] Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden

[4] Department of Data Analysis, Ghent University, Ghent, Belgium

[5] Departmant of Teacher Education and School Research, University of Oslo, Postbox 1099, Blindern, 0317 Oslo, Norway

Published online: 29 November 2023

Introduction

Instructional quality has received increasing attention in educational research and has become central in discussions about educational effectiveness in the classroom and in school contexts (Creemers & Kyriakides, 2008; Crossman & Harris, 2006; Klieme, 2013). Myriad studies underpin how instructional quality is significantly associated with learning outcomes (Hattie, 2008; Scherer et al., 2016; Wisniewski et al., 2020). The quality of instruction reflects the behavior of teachers applying their professional knowledge in teaching activities (Fauth et al., 2014; Good et al., 2009; Seidel & Shavelson, 2007). However, a large variation in measurement approaches and measurement properties of the concept "instructional quality" can be found, often leading to inconsistent estimations of its effect on learning outcomes (e.g., Bellens et al., 2019).

The literature reflects diversity in measuring instructional quality, such as collecting data based on teacher self-reports, student surveys, or peer observations (Boston, 2012; Brown & Kurzweil, 2017; Pianta & Hamre, 2009). Adopting a specific measurement instrument could reflect different conceptual rationales for instructional quality (e.g., Borko et al., 2005; Boston, 2012). This is explained by the fact that measuring latent constructs relies on responses to the test items or observations of behavior, meaning that measurements of the same structure may vary across studies and individuals (Allen & Yen, 2001; Bandalos, 2018; Shultz et al., 2020). Despite specific measurement perspectives that could result in an inconsistent understanding of instructional quality (De Jong & Westerhof, 2001; Maulana & Helms-Lorenz, 2016), distinct measurement approaches can complement each other (Kunter & Voss, 2013). Collecting data on the basis of a range of instruments helps in adopting multiple perspectives. A minority of studies embrace both teacher and student perspectives to evaluate the instructional quality (Aditomo & Köhler, 2020; Bellens et al., 2019; Nilsen & Gustafsson, 2016). Thus, multi-actor focus requires the availability of data from teachers and students from the same schools within the same educational system.

Studies point to the multi-dimensional nature of what is embraced with the concept of instructional quality. Researchers put forward two-, three-, four- or even more extensive multi-dimensional models to capture the richness and complexity of the instructional quality concept. For example, Klieme et al. (2009) identified a three-dimensional model of instructional quality: *classroom management*, *supportive climate,* and *cognitive activation*. Though many authors adopted this model (e.g., Bellens et al., 2019; Jentsch & Schlesinger, 2017; Scherer et al., 2016; Wagner et al., 2016; Wisniewski et al., 2020), others consider this framework to be too generic and especially ignoring domain-specific knowledge domain characteristics, e.g., in mathematics (Schlesinger & Jentsch, 2016; Schlesinger et al., 2018).

In mathematics education, teaching demands and teachers' professional knowledge features distinctive subject-specific characteristics. 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; 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 Anghileri, 2006; Chapin & O'Connor, 2007; Hunter, 2005; Watson & Mason, 2006).

Conceptualizing "mathematical instructional quality" across measurement instruments allows us to find common theoretical grounds in the overlap. It also helps in opening the mathematical instructional quality "black box" to capture new dimensions, contributing to a better understanding of mathematical instructional quality in the school context, especially when looking at mathematical teaching demands. The results provide a richer reflection on mathematical instructional quality in school mathematics education.

The above brings us to a key feature of the present article: applying data from teachers and students in the same school to reflect the nature of mathematical instructional quality. We utilized the new linkage database combining data from the Teaching and Learning International Survey 2013 (TALIS 2013) and the Program for International Student Assessment 2012 (PISA 2012). PISA 2012 investigated 15-year-students' literacy in mathematics and collected data about how they perceived their mathematics teachers' teaching activities. Nevertheless, PISA does not build on data collected from teachers (OECD, 2014a). TALIS 2013 involved a subgroup of mathematics teachers teaching the PISA 2012 eligible "15-year-old" students in the same school and studied explicitly mathematics teacher perceptions about their teaching work, but did not collect data from students, nor did it collect data about student learning outcomes (OECD, 2014b). TALIS 2013 and PISA 2012 focused on mathematics and shared one key anchor variable: PISASCHOOLID. This helps to link teacher and student data from the same schools and to measure mathematical instructional quality from both teacher and student perspectives in a school context. The two datasets adopted different questions and measurement items and adopted only partially convergent perceptions of mathematics instructional quality.

The study also focuses on checking the measurement invariance of the factor structure to facilitate a comparison of the construct across seven countries: Australia, Finland, Latvia, Portugal, Romania, Singapore, and Spain. The possibility of linking both databases provides a unique opportunity to use these multi-country studies to explore mathematical instructional quality from teacher and student perspectives. However, because PISA 2012 and TALIS 2013 were not originally designed to study mathematics instructional quality and its psychometric properties, the results of the current research should be interpreted with care. They can be viewed as a general screening of mathematics instructional quality with specific emphasis on student and teachers' perspectives and cross-country comparisons.

Theoretical framework

Multi-dimensional models to capture instructional quality

Instructional quality is a construct reflecting critical features of teacher behaviors in the learning environment (Fauth et al., 2014; Good et al., 2009; Seidel & Shavelson, 2007). Although researchers hold different conceptualizations of instructional quality, there is an agreement that it is a multidimensional construct (Baumert et al., 2010; Kyriakides et al., 2020; Nilsen & Gustafsson, 2016). During their development process, different frameworks and models have been used to conceptualize the nature of instructional quality. Charalambous and Praetorius (2018) proposed a continuum with different junctures from generic to more subject-specific approaches to capture instructional quality. Senden et al. (2022) adjusted this continuum by building on three categories: generic frameworks, subject-specific frameworks, and hybrid frameworks. We expanded our approach on this basis; see Fig. 1.

A generic framework emphasizes the general level aspects of instructional quality and is characterized by not considering subject-specific instructional performance or

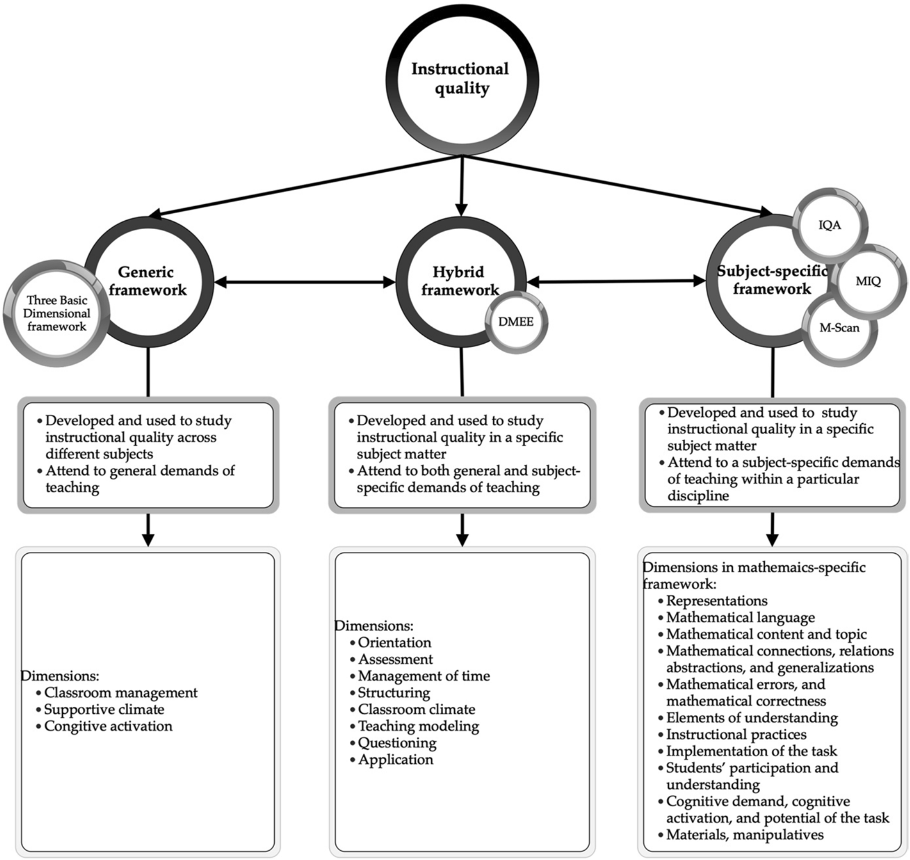
Instructional quality
Generic framework
Three Basic Dimensional framework
Hybrid framework
DMEE
Subject-specific framework
IQA
MIQ
M-Scan
• Developed and used to study instructional quality across different subjects
• Attend to general demands of teaching
• Developed and used to study instructional quality in a specific subject matter
• Attend to both general and subject-specific demands of teaching
• Developed and used to study instructional quality in a specific subject matter
• Attend to a subject-specific demands of teaching within a particular discipline
Dimensions:
• Classroom management
• Supportive climate
• Congitive activation
Dimensions:
• Orientation
• Assessment
• Management of time
• Structuring
• Classroom climate
• Teaching modeling
• Questioning
• Application
Dimensions in mathemaics-specific framework:
• Representations
• Mathematical language
• Mathematical content and topic
• Mathematical connections, relations abstractions, and generalizations
• Mathematical errors, and mathematical correctness
• Elements of understanding
• Instructional practices
• Implementation of the task
• Students' participation and understanding
• Cognitive demand, cognitive activation, and potential of the task
• Materials, manipulatives

Fig. 1 A continuum of classifying instructional quality framework (adapted from Charalambous & Praetorius, 2018; Senden et al., 2022). In the subject-specific framework, we apply mathematics as a specific domain to describe the primary frameworks and dimensions

the teaching demands of a particular subject (Charalambous & Praetorius, 2018). An example of such a framework is the Three Basic Dimensional framework (Klieme et al., 2009). The three basic dimensions consist of classroom management, cognitive activation, and supportive climate (also labeled teacher/student/learning support). This framework is called "BASIC" since all three dimensions are traceable in the other frameworks. *Classroom management* refers to establishing an orderly learning environment, preventing interrupting behaviors, and maximizing effective learning time-use (Pianta & Hamre, 2009; Rakoczy et al., 2007). *Cognitive activation* refers to invoking high-level thinking, knowledge integration, and cognitive engagement in problem-solving, decision-making, and knowledge application (Lipowsky et al., 2009; Pinger et al., 2017). *Supportive climate* relates to the instructional support that guarantees social interaction between students and teachers and invokes student motivation and interest (Pinger et al., 2017; Rakoczy et al., 2008). Plenty of studies adopted this model and set forth empirical evidence to ground the three dimensions (e.g., Baumert et al., 2010; Bellens et al., 2019; Fischer et al., 2019; Künsting et al., 2016; Praetorius et al., 2014).

Despite the fact that the generic framework is widely used across disciplines, scholars have argued that its generality does not help in capturing the complexity introduced by subject-specific aspects of instructional quality, such as in the school subject mathematics (Charalambous & Praetorius, 2018; Cohen et al., 2018; Schlesinger et al., 2018). Researchers suggest that mathematical knowledge should be taken into account when evaluating the quality of instruction in mathematics education (Charalambous & Kyriakides, 2017; Charalambous & Praetorius, 2018; Schlesinger & Jentsch, 2016). The resulting subject-specific frameworks have largely been developed during the past decade. In mathematics education, several frameworks and measurement tools have been developed. These include, for instance, the Mathematical Quality of Instruction (MQI; Learning Mathematics for Teaching, 2011), the Mathematics-Scan (M-Scan; Walkowiak et al., 2014), and the Instructional Quality Assessment (IQA; Matsumura et al., 2008). Rather than capturing general teacher-student interactions in the classroom setting, existing mathematics-specific frameworks on instructional quality focus on these interactions through a content-related lens.

Another type of framework is situated between the general and subject-specific approach, named the hybrid framework (Charalambous & Praetorius, 2018). Hybrid frameworks have been developed by incorporating the features of both generic and subject-specific frameworks. For example, the Dynamic Model of Educational Effectiveness (DMEE) conceptualizes instructional quality by relying on eight dimensions that play a role in the learning environment (Creemers & Kyriakides, 2008; Kyriakides et al., 2020). Teachers play a central role in this model. They are seen as critical decision-makers when creating and fostering learning environments that promote instructional quality. Meanwhile, teachers' decisions and behaviors are also dependent on circumstances. Following the DMEE, eight key decisions can be identified to define instructional quality: management of time, classroom climate, orientation, questioning, structuring, teaching-modeling, application, and assessment (Creemers & Kyriakides, 2008; Kyriakides et al., 2020).

From a theoretical perspective, some dimensions in DMEE can be used as sub-dimensions of the Three Basic Dimensional framework. For instance, the dimension *management of time* is related to the basic dimension *classroom management*, referring to a teacher's ability to manage time on a task efficiently and to maximize student engagement in the learning process. The dimension *questioning*, *teaching modeling*, and *application* are related to the basic dimension *cognitive activation*, which means to teachers' efforts to active students' cognitive engagement and explore students' prior knowledge by completing challenging tasks in order to promote student's higher-order thinking and develop their problem-solving competencies.

Nevertheless, other dimensions extend the conceptualization of instructional quality in the Three Basic Dimensional framework. *Classroom climate* points to the orderly learning environment defined by the interactions between teacher/student and student, the way competition between students is being handled, the teacher's treatment of students, and the setting of classroom discipline. *Orientation* focuses on providing specific tasks or lessons to encourage the active participation of students. *Structuring* builds on a teaching procedure that starts from clarifying the learning objectives at the start, outlining the learning content, stressing connections between lessons parts, underlining the key points, to reviewing the lesson at the end of a session. *Assessment* includes strategies for collecting information about students' mastery of knowledge, skills, and attitudes and mapping beliefs related to student outcomes. It also includes the identification of students' needs and the evaluation of the teachers' instructional practices.

Researchers consider DMEE a general framework without subject-specific instruction (Charalambous & Praetorius, 2018; Senden et al., 2022). The dimensions of orientation and assessment seem subject-specific based on empirical evidence (Creemers & Kyriakides, 2008, p. 200). In mathematics education, student-oriented instruction reflects teachers' behaviors of providing mathematics-specific tasks to encourage students to participate actively and having students work in small groups (See Genç & Çolakoğlu, 2021; Yi & Lee, 2017). However, Through the TALIS 2013 and PISA 2012 data—used in this study—we can check whether the DMEE reflects such mathematics-specific characteristics.

The discussion about a comprehensive model and the identification of (alternative) dimensions can additionally be looked at from a different angle. Since many studies are conducted in a single country setting, the question that can be asked is whether a part of the variance in models depends on the national/regional context of an educational system. For instance, Nilsen et al. (2016) stressed the need for a better understanding of the factor structure of instructional quality in international studies, especially when looking at non-Western countries.

International comparison of instructional quality in mathematics education

When comparing instructional quality across different educational systems, measurement invariance has to be achieved. Otherwise, inferences and conclusions reached from the cross-country comparison might be weak or invalid for a specific context (Horn, 1991; Oliveri & von Davier, 2011; Vandenberg & Lance, 2000). Testing measurement invariance helps to evaluate whether the measurement model can be applied to multiple groups while studying between-group differences. In view of this analysis, three nested structured levels of measurement invariance are commonly considered: configural invariance (identical factor structure), metric invariance (equality of factor loadings), and scalar invariance (equality of factor loadings and measurement intercepts) (Cheung & Rensvold, 2002; Meitinger et al., 2020). The diversity in educational policies and practices suggests that measurement invariance across countries is hard to achieve, which challenges assumptions related to large-scale international studies (Rutkowski & Rutkowski, 2018). For example, Bellens et al. (2019) examined the factor structure of mathematics instructional quality based on TIMSS 2015 data from Belgium (Flanders), Germany, and Norway, and could not identify a common factor structure. Although the three basic instructional quality dimensions were identified in all three educational systems, the structure of both the *supporting climate* and *cognitive activation* dimensions differed.

It has been observed that linguistic similarity could interfere with the degree of measurement invariance. Scherer et al. (2016) measured students' perceptions of mathematics instructional quality, based on PISA 2012 data from Australia, Canada, and the USA, three predominantly English-speaking countries. The three-dimensional factor structure of instructional quality did fit the three countries, and scale invariance was achieved. Fischer et al. (2019) investigated measurement invariance for the three-dimension structure of instructional quality with PISA 2012 data from 15 countries, grouped along with five linguistic clusters (i.e., Chinese-, English-, French-, German- and Spanish-speaking groups). Even though the metric invariance was reached, the comparability of instructional quality in these countries was still limited partly due to language differences. The authors referred to the high similarity of linguistic background could enhance the degree of comparability in four clusters, excluding the Chinese-speaking countries.

The inconsistent picture emerging from the previous studies suggests that testing the measurement invariance of constructs should be continued. Un-identified variance might have implications for policy-making that is often directly based on empirical evidence derived from large-scale performance indicators studies such as PISA and TALIS.

Differences in teacher and student perspectives on instructional quality

The literature points to differences in the perceptions of teachers and students about instructional quality (Kunter & Baumert, 2006; Wagner et al., 2016). These differences are relatively stable over time (Wagner et al., 2016). Even when the teachers' and the students' ratings of instructional quality dimensions were based on the same or similar items, only a low level of agreement was observed, affecting the construct validity of the concept (Brok et al., 2006; Clausen, 2002; Kunter & Baumert, 2006). A reasonable explanation for the differences can be found in respondents' background characteristics, personal preferences, social identities, and social expectations (Fauth et al., 2014; Kunter & Baumert, 2006; Wagner et al., 2016).

According to measurement theory, the same latent construct may vary depending on the measurement items and the responses across individuals (Bandalos, 2018; Shultz et al., 2020). The combination of teacher and student perspectives can reflect the nature of instructional quality, but also puts forward questions about the theoretical grounding of the concept that respects these different perspectives (Urdan, 2004). Kunter and Baumert (2006) also add to the discussion about unreliable measurement approaches invoking low correlations between teacher and student perspectives. Building on the multi-level perspective on the dynamic model of educational effectiveness, one should be aware that teacher-perceived instructional quality is mutually influenced by factors at the student and school level.

Mapping instructional quality using TALIS data and PISA data

Table 1. gives an overview of the studies in which TALIS data or PISA data have been used to measure instructional quality, including studies in mathematics and other domains. Only few studies measured the quality of instruction on the base of these datasets. Two studies applied teacher self-reported data from TALIS to capture instructional quality across subjects: the TALIS 2018 technical report (OECD, 2019) and Chen et al. (2020). Other studies focused on subject-specific approaches; see, e.g., Aditomo and Köhler (2020), who identified five dimensions of instructional quality in science education using PISA 2015 data.

In the current study, we build on two datasets—TALIS 2013 and PISA 2012. This implies the related measurement instruments that specifically focus on mathematics. Neither of these two datasets measured the actual construct of instructional quality; see the technical reports (OECD, 2014a, 2014b). However, a picture of instructional quality can be "constructed" on the basis of these data. Four of the above studies already applied PISA 2012 data to explore the quality of mathematical instruction. The results of these studies support the three dimensions model, focusing on "classroom management", "cognitive activation", and "supportive climate". A new additional dimension, "student-oriented instruction" was identified by Yi and Lee (2017) and Genç and Çolakoğlu (2021). To the best of our knowledge, no studies have used TALIS 2013 to explore the construct instructional quality, let alone studies that address the quality of mathematical instruction. In addition, studies building on PISA 2012 reported inconsistent results, and it is, therefore, relevant to continue this field of study to check whether other new dimensions can be identified, especially when comparing the different educational systems and cultural settings.

Table 1 Summary of studies using TALIS and PISA data to capture instructional quality

	Classroom management	Cognitive activation	Supportive climate	Adaptive instruction	Clarity of structuring	Assessment	Classroom climate	nquiry-bases instruction	Student-oriented instruction	Teacher-oriented instruction	Mathematics	General subject/ Other subjects	PISA data	TALIS data	International comparative studies
1. Scherer et al. (2016)	X	X	X								X		X		X
2. Yi and Lee (2017)	X	X	X						X		X		X		X
3. Fischer et al. (2019)	X	X	X								X		X		X
4. OECD (2019)	X	X			X							X		X	X
5. Aditomo and Köhler (2020)	X		X	X		X		X		X			X		X
6. Chen et al. (2020)	X	X			X							X		X	
7. Genç and Çolakoğlu (2021)	X	X	X						X		X		X		

In sum, the review of substantive studies suggested that the dimensionality of the instructional quality in mathematics could be different between teacher vs. student perspectives and across different education systems (Fauth et al., 2014; Kunter & Baumert, 2006). Testing measurement invariance is therefore essential in comparing across perspectives and groups.

The present study aims to identify the factor structure of mathematics instructional quality from both a teacher (TALIS 2013) and a student (PISA 2012) perspective. This process allows us to understand whether the three basic dimensional mathematics instructional quality holds for different countries and perspectives. If not, we look at whether other determinants might help describe instructional quality when starting from the eight dimensions derived from the Dynamic Model of Educational Effectiveness. Additionally, we aim to check whether the countries studied can identify similar measurement properties.

Building on the available conceptual and empirical base, we pursue the general research aim through the following research questions:

Research question 1 How do teachers (TALIS 2013 data) and students (PISA 2012 data) from the same schools perceive mathematics instructional quality separately?

Research question 2 To what extent does the measurement model that is invariant across countries?

Methods

Data and sample

In TALIS 2013,[1] eight countries—also involved in PISA 2012—participated in the so-called TALIS-PISA Link study (TPL): Australia, Finland, Latvia, Mexico, Portugal, Romania, Singapore, and Spain. Mathematics teachers completed an additional TALIS 2013 Mathematics Teacher Questionnaire to collect data about their mathematics teaching practice and their beliefs about teaching mathematics (OECD, 2013, 2014b). The TPL database contains data about, among other things, teacher background and characteristics and their mathematics teaching practice. For each country participating in TPL, 150 schools, all eligible mathematics teachers from each school, and 20 non-mathematics teachers did participate (OECD, 2014b).

To explore the mathematics-related quality of instruction, only TPL data obtained from mathematics teachers were included in the link with the PISA 2012 student data. The linking procedure to produce a Redesigned TALIS-PISA Link (rTPL) database, can be found in Liu et al., (2020, 2022). Related information also has been provided in Appendix 1. The TPL dataset builds on data from 3473 teachers and 31,584 students in 1115 schools. The sample size for the individual countries is provided in Table 2.

[1] It might be argued that building on 2012 and 2013 data reflects using dated research data. We reply to this comment that our research aims and questions can be seen as independent of a specific data collection. Moreover, it is only for these two cycles that multi-country data are available that can be connected through a shared school ID in the two data sets. Additionally, we stress that the present study develops a new data set, based on the linking of both studies, while focusing on mathematics instructional quality.

Table 2 Overview of the Redesign TALIS-PISA Link database (Source from Liu et al., 2020; OECD, 2014a; OECD, 2014b)

	Australia	Finland	Latvia	Mexico	Portugal	Romania	Singapore	Spain	Pooled
Teacher	415	332	178	170	537	390	719	732	3 473
Student	2 251	4 010	2 013	2151	3 886	4 103	5 302	7 868	31 584
School	113	133	85	87	131	131	164	271	1 115

Variables

The teacher-perceived "instructional quality" is based on 12 TALIS items about teaching practices (TT2G42, TT2G43), classroom climate (TT2G41), and teaching mathematics approaches (TT2M13) in TALIS Questionnaires.[2] To capture the students' reported "instructional quality", 29 PISA questionnaire items about teacher support (ST77/ST83), learning environment (ST81, ST85), and learning process (ST79, ST80) were selected. Appendices 3 and 4 presented the descriptive statistics of the original items.

In the PISA Questionnaire, Question ST77 (i.e., How often do these things happen in your mathematics lessons) and ST83 (i.e., Thinking about the mathematics teacher who taught your last mathematics class) collect data about teacher support. ST77 captures the information on the general mathematics teaching procedure, and ST88 on the last mathematics class attended by the students. The item pairs ST77Q06 and ST83Q04, ST77Q04 and ST83Q03, ST77Q02, and ST83Q02, shared the wording of the items. Therefore, they are averaged into three indicator parcels M_TSMS1 (ST77Q06 and ST83Q04), M_TSMS2 (ST77Q04 and ST83Q03), and M_TSMS3 (ST77Q02 and ST83Q02). Parceling is a measurement practice using the sum or mean of two or more items. It strengthens factor loadings and correlations among indicators and reduces the number of indicators and model complexity (Little et al., 2002; Rigdon et al., 2019). We first calculated correlation coefficients before creating the variable parcels. Parceling was justified considering the substantial correlation. Table 3 presents the correlations between the variables in each of the three pairs.

Analytical methods and process

Confirmatory factor analysis

The analyses were carried out using confirmatory factor analysis (CFA) via Mplus 8.3 (Muthén & Muthén, 2017). CFA is used to examine the latent factor structure of a test instrument, investigate measurement properties of latent constructs, and test measurement invariance of the measurement properties across subgroups (Brown, 2015). Maximum likelihood estimation with robust standard errors (MLR) with the EM algorism (expectation–maximization) was used to handle missing and non-normal data. Since the Chi-square test (χ^2), along with its degrees of freedom (df), is sensitive to large

[2] Teaching practices, classroom climate and teaching approaches are the relevant questions in the TALIS Questionnaires. Teacher support, learning environment and learning process are the relevant questions in the PISA Questionnaires. We utilized to identify mathematics instructional quality, which not represent the theoretical dimensions.

Table 3 The correlation coefficients for parcels in pooled data

	Pooled	Australia	Finland	Latvia	Mexico	Portugal	Romania	Singapore	Spain
ST77Q02 WITH ST83Q02	0.5	0.6	0.6	0.5	0.4	0.5	0.3	0.5	0.5
ST77Q04 WITH ST83Q03	0.5	0.6	0.6	0.6	0.4	0.6	0.3	0.5	0.6
ST77Q06 WITH ST83Q04	0.6	0.6	0.7	0.6	0.4	0.6	0.3	0.6	0.6

sample sizes (Moshagen, 2012; Shi et al., 2018), we did not solely rely on χ^2 to test model fit. Other model fit indices also are used: Comparative Fit Index (CFI) with a cut-off value $\geq$ 0.95, Root Mean Square Error Approximation (RMSEA), a cut-off value $\leq$ 0.05, and Root Mean Square Residual (SRMR) with a cut-off value $\leq$ 0.08 (Hu & Bentler, 1999; Kline, 2015; Taasoobshirazi & Wang, 2016; Worthington & Whittaker, 2006).

Measurement invariance test

The model structure was established on the pooled data from all the countries. The factor structure was estimated for each country separately. Cross-country comparability was applied by the measurement invariance test. This is a statistical procedure to analyze the comparability of latent scales in the CFA framework across various groups (Cheung & Rensvold, 1999; Cheung & Rensvold, 2002; Davidov, 2008; Van de Vijver et al., 2019). Mplus helped to test measurement invariance in one model with the syntax "MODEL = CONFIGURAL METRIC SCALAR" in the ANALYSIS command.

As mentioned, measurement invariance might be observed at three hierarchically structured levels: the configural, the metric, or the scalar levels. Each level builds on the previous by requiring additional equality constraints to reach a higher degree of invariance. *Configural invariance* is the base level building on items and equal factor structure of the latent construct. If only configural invariance is achieved, no statistical comparison is applicable across groups (Cheung & Rensvold, 2002). *Metric invariance* requires the factor structure and factor loadings to be equal and is a prerequisite for meaningful cross-group comparison, such as used in regression analysis (Bollen, 1989; van de Schoot et al., 2012). *Scalar invariance* requires that the intercepts should be equal across groups in addition to equal factor structure and equal factor loadings. For scalar invariance, it is meaningful to compare the mean value of the constructs across groups (Byrne, 2008).

The criteria for testing measurement invariance—adopted in the present study–are from Rutkowski and Svetina (2014) for a large sample in groups (600 to 6,000 per group), the value changes in CFI (ΔCFI) are not less than − 0.020, and changes in RMSEA (ΔRMSEA) are less than 0.030 for metric invariance. The value changes in CFI (ΔCFI) are greater than − 0.010, and changes in RMSEA (ΔRMSEA) are less than 0.010 for scalar invariance. The recommended cut-off values are used to investigate the rejection rates for different degrees of invariance within each level and for various levels of invariance.

Analytical process

Firstly, we established the measurement model using pooled data from TALIS (teacher-perceived data) and PISA (student-perceived data) separately to explore the dimensions of the mathematical instructional quality construct. After obtaining the factor structure of the pooled data, the models were examined country by country. In the measurement process, we found that the measurement models of Mexico differ from all other countries, whether it is based on mathematics teacher data or student data. In the TALIS 2013 data, none of the dimensions of mathematics instructional quality can be identified in Mexico. In the PISA 2012 data, the factor structure for some dimension of mathematics instructional quality in Mexico is different from that of other countries. This implies that if Mexican data are included in the study, the measurement invariance among the eight countries can only reach the configural level. That is, these eight countries are not even fundamentally comparable to each other. We decide to remote the Mexico results into an Appendix 2. And in the result section below we will focus on the rest of the seven countries.

The next step consisted of testing the measurement invariance of the hypothesized model to see whether a common factorial structure and common measurement properties could be found across different countries in TALIS and PISA.

Results

Measurement property of mathematical instructional quality (Research question 1)

Factor structure across countries building on the teacher perspective

Building on the pooled data of the seven countries, three dimensions of mathematical instructional quality could be identified: *classroom disciplinary climate (DC-t)*, *teacher support (TS-t),* and *cognitive activation (CA-t)*. This model fits the data well, with CFI = 0.99, RMSEA = 0.02, SRMR = 0.03, and Chi-square = 90.86 with 51 df. Figure 2 displays the path diagram.

Each of the three factors is defined by four items, and all factor loadings exceed 0.30. The correlation between the two pairs of the latent variables is relatively low: 0.13 for *DC-t* and *TS*-t and 0.19 for *DC-t* and *CA-t*. The correlation between *TS-t* and *CA-t* is, however, substantially higher (0.65). However, when modeling the countries separately, the common three-dimensional oblique model does not hold for some countries. Table 4 presents the model-fit indices of the pooled teacher data and country profile.

Table 5 summarizes the instructional quality dimension, factor loadings, and factor correlation coefficient for pooled data and country profile. *Teacher support* could not be identified in Latvia and Romania, and only three items helped in constituting this dimension in Finland (the item TT2G43F, i.e., *I observe students when working on particular tasks and provide immediate feedback* was excluded). *Cognitive activation* in Latvia and Romania did differ from that in the other countries. In Australia and Spain, *teacher support* did not significantly correlate with *classroom disciplinary climate*, while in Finland and Portugal, the *disciplinary climate* was not significantly correlated with neither *teacher support* nor *cognitive activation*. In Romania, the correlation between *cognitive activation* and *disciplinary climate* was not significant.

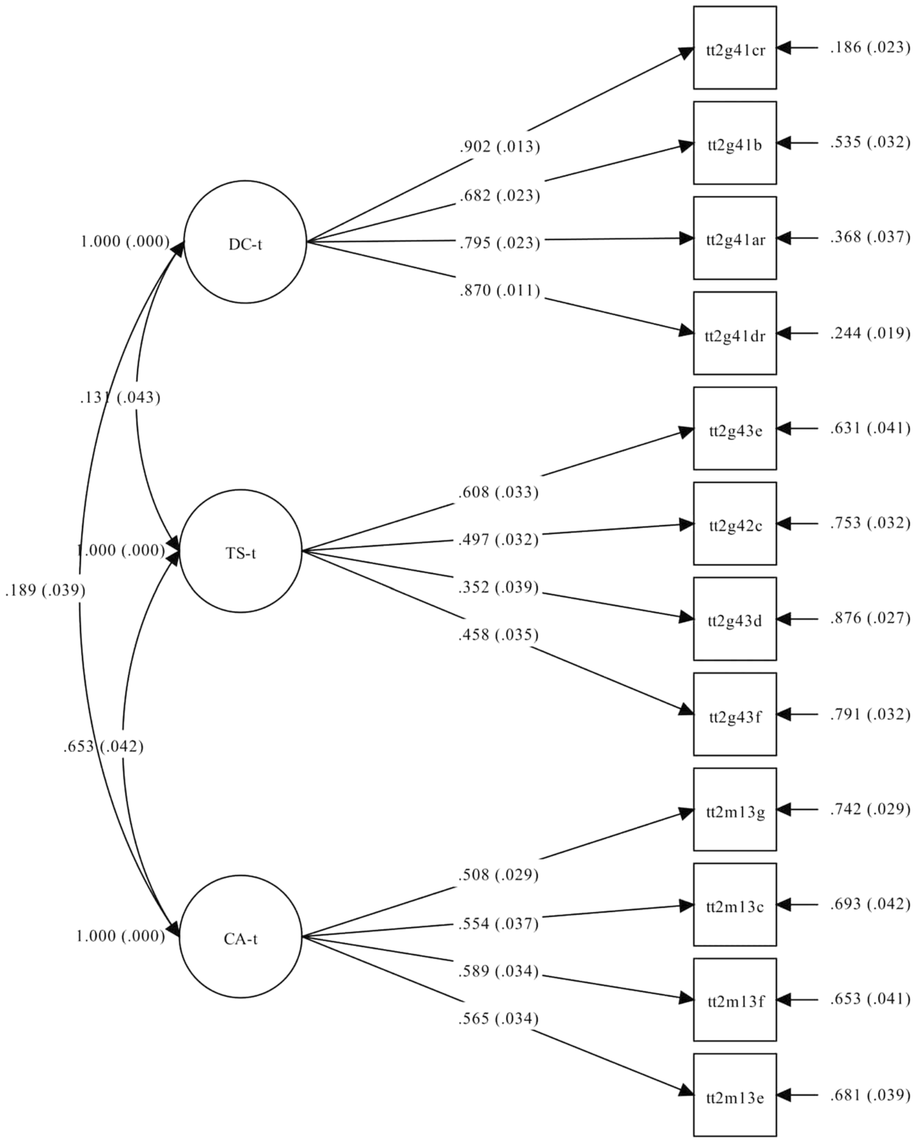

Fig. 2 The measurement model of the pooled teacher perspective data. *DC-t* Classroom disciplinary climate; *TS-t* Teacher support; *CA-t* Cognitive activation

Factor structure across countries building on the student perspectives

Five dimensions were identified in the pooled student responses when focusing on mathematics instructional quality: *teacher support (TS-p), cognitive activation (CA-p), classroom disciplinary climate (DC-p), classroom management (CM-p),* and *student-orientated instruction (SO-p)*. This model fits the data well (see Table 6), with CFI being 0.96, RMSEA being 0.03, SRMR being 0.03, and Chi-square being 5371.27 with 286 df.

Table 4 The model-fit indices in the pooled teacher perspective data and broken down by country

	Pooled	Australia	Finland	Latvia	Portugal	Romania	Singapore	Spain
CFI	0.99	0.97	0.98	0.98	0.97	0.96	0.98	0.99
RMSEA	0.02	0.04	0.03	0.05	0.04	0.05	0.03	0.01
SRMR	0.03	0.05	0.04	0.05	0.04	0.06	0.04	0.03
Chi-square	90.86	84.24	54.51	19.21	86.61	36.35	93.77	53.42
df	51	51	41	13	51	18	51	51

Table 5 The factor structure of instructional quality for seven countries is based on teacher perspectives

	Disciplinary climate (DC-t)		Teacher support (TS-t)		Cognitive activation (CA-t)		Correlation coefficient	
	Variable	FL	Variable	FL	Variable	FL		
Pooled	TT2G41A†	0.80	TT2G42C	0.50	TT2M13C	0.55	DC-t with TS-t	0.13
	TT2G41B	0.68	TT2G43D	0.35	TT2M13E	0.57	DC-t with CA-t	0.19
	TT2G41C†	0.90	TT2G43E	0.61	TT2M13F	0.59	TS-t with CA-t	0.65
	TT2G41D†	0.87	TT2G43F	0.46	TT2M13G	0.51		
Australia	TT2G41A†	0.78	TT2G42C	0.40	TT2M13C	0.55	DC-t with TS-t	*0.13*
	TT2G41B	0.69	TT2G43D	0.45	TT2M13E	0.56	DC-t with CA-t	0.19
	TT2G41C†	0.95	TT2G43E	0.58	TT2M13F	0.67	TS-t with CA-t	0.60
	TT2G41D†	0.84	TT2G43F	0.36	TT2M13G	0.47		
Finland	TT2G41A†	0.79	TT2G42C	0.35	TT2M13C	0.53	DC-t with TS-t	*0.11*
	TT2G41B	0.57	TT2G43D	0.51	TT2M13E	0.46	DC-t with CA-t	*0.14*
	TT2G41C†	0.91	TT2G43E	0.68	TT2M13F	0.62	TS-t with CA-t	0.41
	TT2G41D†	0.92			TT2M13G	0.36		
Latvia	TT2G41A†	0.76			TT2M13C	0.77	DC-t with CA-t	0.35
	TT2G41B	0.67			TT2M13F	0.48		
	TT2G41C†	0.82			TT2M13G	0.64		
	TT2G41D†	0.82						
Portugal	TT2G41A†	0.81	TT2G42C	0.47	TT2M13C	0.62	DC-t with TS-t	*0.05*
	TT2G41B	0.75	TT2G43D	0.52	TT2M13E	0.43	DC-t with CA-t	*0.04*
	TT2G41C†	0.91	TT2G43E	0.61	TT2M13F	0.65	TS-t with CA-t	0.52
	TT2G41D†	0.81	TT2G43F	0.57	TT2M13G	0.71		
Romania	TT2G41A†	0.73			TT2M13E	0.57	DC-t with CA-t	*0.10*
	TT2G41B	0.47			TT2M13F	0.80		
	TT2G41C†	0.95			TT2M13G	0.47		
	TT2G41D†	0.86			TT2G43F	0.51		
Singapore	TT2G41A†	0.79	TT2G42C	0.44	TT2M13C	0.66	DC-t with TS-t	0.16
	TT2G41B	0.64	TT2G43D	0.51	TT2M13E	0.54	DC-t with CA-t	0.28
	TT2G41C†	0.88	TT2G43E	0.64	TT2M13F	0.58	TS-t with CA-t	0.59
	TT2G41D†	0.89	TT2G43F	0.50	TT2M13G	0.45		
Spain	TT2G41A†	0.78	TT2G42C	0.58	TT2M13C	0.49	DC-t with TS-t	*0.08*
	TT2G41B	0.70	TT2G43D	0.41	TT2M13E	0.58	DC-t with CA-t	0.24
	TT2G41C†	0.86	TT2G43E	0.59	TT2M13F	0.54	TS-t with CA-t	0.59
	TT2G41D†	0.90	TT2G43F	0.47	TT2M13G	0.47		

FL Factor loading; Italic indicates the non-significant estimate ($P > 0.05$). *TT2G41* Classroom climate; *TT2G43, TT2G42* Teaching practice; *TT2M13* Teaching approaches in the mathematics class

† Item was reverse coded

Table 6 The model-fit indices in the pooled data and broken by country from student perspectives

	Pooled	Australia	Finland	Latvia	Portugal	Romania	Singapore	Spain
CFI	0.96	0.96	0.95	0.95	0.97	0.95	0.95	0.96
RMSEA	0.03	0.04	0.04	0.04	0.04	0.03	0.04	0.03
SRMR	0.03	0.04	0.06	0.04	0.04	0.04	0.04	0.04
Chi-square	5177.64	854.47	1455.27	765.04	1212.52	1145.8	2133.57	1322.99
df	286	286	286	286	286	286	286	286

Figure 3 depicts the model factor structure. *Teacher support (TS-p)* was measured using seven items (ST77Q01, ST77Q05, M_TSMS1, M_TSMS2, M_TSMS3, ST80Q08), *cognitive activation (CA-p)* using eight items (ST80Q01, ST80Q04, ST80Q05, ST80Q07, ST80Q08, ST80Q09, ST80Q10, ST80Q11), *classroom disciplinary climate (DC-p)* using six items (ST81Q01, ST81Q02, ST81Q03, ST81Q04, ST81Q05, ST85Q04), *classroom management (CM-p)* using three items (ST85Q01, ST85Q02, ST85Q03), and *student-orientated instruction (SO-p)* using four items (ST79Q03, ST79Q04, ST79Q07, ST79Q10). The residual correlation of items M_TSMS3 and M_TSMS2, and of items ST81Q02 and ST81Q01 were added to the model to account for the common variance of the variable residuals and to improve the model fit.

The factor structure of instructional quality in the pooled data was applicable to each country. Table 7 provides the dimensions of instructional quality, factor loadings, and factor correlation coefficients for the pooled data and each country separately.

Though the five-factor structure of instructional quality could be tracked in each country, the strength of factor loadings, and the correlations among the five factors, varied extensively. In Australia, the latent variable correlation coefficients between *teacher support (TS-p)* and *classroom management (CM-p)*, *teacher support,* and *cognitive activation (CA-p)* were both 0.72. In Finland and Latvia, *classroom disciplinary climate (DC-p)* was not correlated with *student-orientated instruction (SO-p)*. Meanwhile, the *teacher support* and *cognitive activation* dimensions shared 55% of the variance in Latvia. Two pairs of dimensions had no correlations with each other in Romania, namely, *cognitive activation (CA-p)—classroom disciplinary climate (DC-p)* and *classroom management (CM-p)—student-oriented instruction (SO-p)*.

Measurement invariance between countries (Research question 2)

Limited possibility to compare models based on teacher data

Our second research question concerns the invariance of the measurement model across seven countries. Based on the TALIS teacher data, different factor structures of mathematical instructional quality were identified for each country. Since configural invariance requires the factor structure to be equal across groups, these analysis results imply that the configural invariance level was not achieved. Therefore, model comparison from teacher perceived mathematical instructional quality between countries was hardly possible.

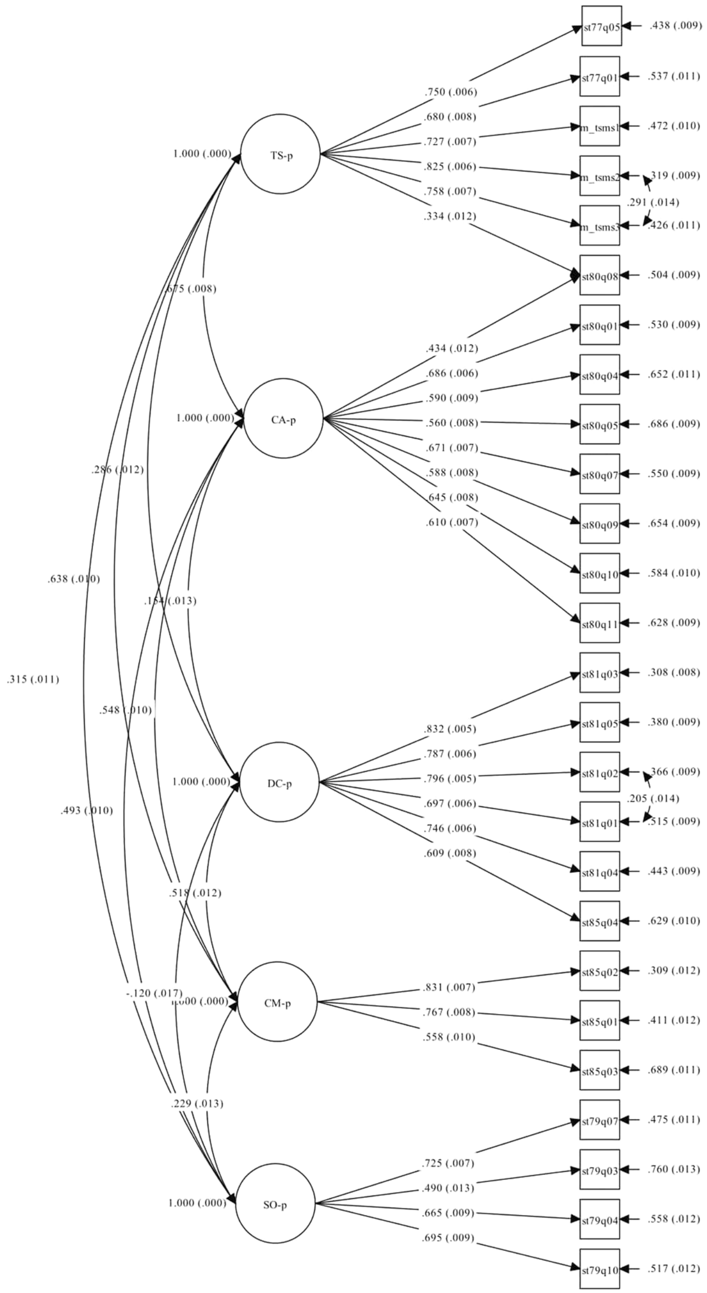

Fig. 3 The measurement model on the basis of the pooled student perspective data. *DC-p* classroom Disciplinary climate; *TS-p* Teacher support; *CA-p* Cognitive activation; *CM-p* Classroom management; *SO-p* Student-oriented instruction

Table 7 The factor structure of instructional quality for seven countries from student perspectives

	Variable	Pooled	Australia	Finland	Latvia	Portugal	Romania	Singapore	Spain
TS-p	ST77Q01†	0.68	0.75	0.76	0.54	0.79	0.50	0.70	0.77
	ST77Q05†	0.75	0.79	0.79	0.71	0.81	0.65	0.75	0.75
	M_TSMS1†	0.73	0.78	0.70	0.71	0.79	0.67	0.75	0.71
	M_TSMS2†	0.83	0.88	0.85	0.84	0.88	0.75	0.85	0.88
	M_TSMS3†	0.76	0.83	0.80	0.81	0.84	0.67	0.80	0.80
	ST80Q08†	0.33	0.37	0.41	0.31	0.32	0.20	0.30	0.38
CA-p	ST80Q01†	0.69	0.72	0.70	0.65	0.77	0.60	0.72	0.71
	ST80Q04†	0.59	0.69	0.60	0.37	0.75	0.45	0.68	0.56
	ST80Q05†	0.56	0.60	0.61	0.39	0.66	0.52	0.57	0.56
	ST80Q07†	0.67	0.71	0.69	0.63	0.74	0.58	0.70	0.62
	ST80Q08†	0.43	0.43	0.40	0.39	0.52	0.49	0.47	0.38
	ST80Q09†	0.59	0.66	0.57	0.58	0.70	0.54	0.64	0.55
	ST80Q10†	0.65	0.70	0.65	0.62	0.72	0.54	0.71	0.57
	ST80Q11†	0.61	0.66	0.63	0.56	0.68	0.51	0.67	0.54
DC-p	ST81Q01	0.72	0.78	0.73	0.69	0.69	0.61	0.75	0.68
	ST81Q02	0.80	0.79	0.79	0.84	0.83	0.74	0.81	0.82
	ST81Q03	0.83	0.84	0.85	0.83	0.86	0.78	0.85	0.82
	ST81Q04	0.75	0.85	0.75	0.72	0.77	0.67	0.79	0.79
	ST81Q05	0.79	0.82	0.82	0.71	0.80	0.75	0.80	0.79
	ST85Q04	0.61	0.67	0.68	0.60	0.62	0.52	0.67	0.54
CM-p	ST85Q01†	0.77	0.81	0.84	0.68	0.86	0.66	0.72	0.78
	ST85Q02†	0.83	0.88	0.88	0.84	0.89	0.68	0.84	0.82
	ST85Q03†	0.56	0.65	0.45	0.57	0.54	0.63	0.63	0.52
SO-p	ST79Q03†	0.49	0.51	0.33	0.46	0.68	0.61	0.60	0.50
	ST79Q04†	0.67	0.51	0.76	0.67	0.74	0.70	0.61	0.56
	ST79Q07†	0.73	0.72	0.80	0.66	0.72	0.71	0.70	0.64
	ST79Q10†	0.70	0.75	0.59	0.59	0.69	0.68	0.70	0.69
Correlation coefficient									
TS-p with CA-p		0.68	0.72	0.61	0.74	0.68	0.66	0.69	0.68
TS-p with DC-p		0.29	0.47	0.33	0.28	0.30	0.19	0.29	0.24
TS-p with CM-p		0.64	0.72	0.68	0.69	0.65	0.60	0.65	0.58
TS-p with SO-p		0.32	0.36	0.24	0.41	0.35	0.29	0.30	0.30
CA-p with DC-p		0.15	0.34	0.19	0.22	0.19	*– 0.06*	0.17	0.14
CA-p with CM-p		0.55	0.62	0.55	0.63	0.53	0.48	0.57	0.51
CA-p with SO-p		0.49	0.47	0.40	0.47	0.44	0.49	0.51	0.52
DC-p with CM-p		0.52	0.62	0.61	0.55	0.59	0.34	0.46	0.51
DC-p with SO-p		– 0.12	0.11	*0.00*	*– 0.01*	– 0.08	– 0.44	– 0.14	– 0.11
CM-p with SO-p		0.23	0.22	0.21	0.30	0.26	*0.05*	0.23	0.52

ST77, M_TSMS Mathematics teacher support; *ST79, ST80* Learning process; *ST81, ST85* Learning environment. *DC-p* Classroom disciplinary climate; *TS-p* TEACHER support; *CA-p* Cognitive activation; *CM-p* Classroom management; *SO-p* Student-oriented instruction

†Item was reverse coded; Italic indicates the non-significant estimate ($P > 0.05$)

Table 8 The model-fit indices for three levels of measurement invariance model on the base of student perspective data

	CFI	RMSEA	SRMR	Chi-square	df	ΔCFI	ΔRMSEA	ΔSRMR
Configural	0.955	0.035	0.043	8861.28	2002			
Metric	0.950	0.036	0.050	9773.38	2134	0.005	− 0.001	− 0.007
Scalar	0.906	0.048	0.068	16,525.55	2260	0.044	− 0.012	− 0.018

Metric level of comparison in student survey data

The factor structure of instructional quality resulting from student data did fit well in all seven countries. Measurement invariance was therefore tested to ensure comparability between countries. Table 8 summarizes the results. The differences in fit indices between the configural and the metric invariance models were consistently within the cut-off values as defined by Rutkowski and Svetina (2014), with ΔCFI equals 0.005 (≥ − 0.020) and ΔRMSEA being − 0.001(≤ 0.030).

The differences between the metric and the scalar invariance models exceeded predefined cut-off values. Therefore, we have to conclude that metric invariance holds. Therefore, comparisons can be made as to the association between mathematical instructional quality and, for example, mathematics outcomes of students across countries. However, a comparison of the mean value of the latent variable should be avoided.

In sum, three dimensions of instructional quality, namely *classroom disciplinary climate*, *teacher support,* and *cognitive activation*, emerge based on teacher data. However, a five-dimension model—*classroom disciplinary climate, teacher support, cognitive activation, classroom management, and student-orientated instruction*—was identified by students' perception of mathematics instructional quality. The five-factor model has been replicated successfully on the base of data from each country, and the metric invariance was achieved across the seven countries. In contrast, no comparability could be reached for the measurement model of teacher perceived mathematical instructional quality since the factor structure differed across educational systems.

Correlations between the instructional quality dimensions

The *classroom disciplinary climate* dimension was identified in data from both TALIS and PISA. Surprisingly, based on the student data, we did not observe a significant correlation between *classroom disciplinary climate* and *student-oriented instruction* in Finland and Latvia, and no significant correlation with *cognitive activation* in Romania. Furthermore, *classroom disciplinary climate* was not significantly related to *teacher support* in Australia, Finland, Portugal, and Spain, and showed no significant correlation with *cognitive activation* in Finland, Portugal, and Romania when building on teacher data.

Building on the non-significant or small correlations between dimensions, we further tested a higher-order model in which instructional quality was seen as a general factor, identified by three dimensions in the teacher data and five dimensions in the student data. Higher-order factor analysis allows for a multi-dimensional test structure and builds on repeated steps of factor analysis-oblique rotations (Eid & Koch, 2014). Testing the higher-order model also helps to ensure whether dimensions map the construct instructional quality in an empirical way.

Looking at the oblique model, based on the PISA student data, a significant higher-order general instructional quality factor could be identified, wrapping the five lower-order dimensions. This result did hold for all seven countries, as can be derived from the model-fit indices and factor loadings presented in Table 9. However, the oblique model did not hold for the TALIS data (teacher-perceived data). This might imply that teachers prefer to focus on specific aspects of their teaching during their daily practice, such as classroom discipline, cognitive activation, and teacher support. These aspects are correlated but could also play a role as stand-alone factors that do not necessarily define general instructional quality.

Discussion

A 'dialogue' between teacher and student: mirroring school mathematical instructional quality using the different measurement instruments

Regarding the debate over who is the "expert" in judging the quality of instruction, previous research has pointed out that students seem to be apt at evaluating the different modes of teaching. Nevertheless, teachers are rather apt at evaluating the variety in instructional approaches, teaching standards, and catering for subject domain features (see Clausen, 2002; De Jong & Westerhof, 2001; Lanahan et al., 2005; Porter, 2002). Students seem to focus on the nature of the learning environment, and the way their teachers support their learning matters. But teachers rather emphasize the nature of the instructional tasks and methods. The question is, as such, not who is the best "expert" in judging instructional quality, but rather how a "dialogic format" can be adopted to align the perspectives of students and teachers within schools in a certain country.

We repeat that we conceptualized mathematics instructional quality on the base of two measurement instruments (TALIS 2013 and PISA 2012) that reflect either teacher or student perspectives. Our results show that the Three Basic Dimensional framework does not hold for different countries when looking at mathematics instructional quality, neither based on the teacher nor student self-reported data. The factor structure of the mathematical instructional quality – as reflected in the data from teachers and students – was different. According to measurement theory, the same latent construct may vary depending on the measurement items and the responses across individuals (Bandalos, 2018; Shultz et al., 2020). That is to say, although we applied the responses from teachers and students on different items in TALIS and PISA, all identified the nature of mathematics instructional quality.

Building on student perceptions, the evidence from our study results reflects the same three factors as found in teacher perceptions, but two additional non-cognitive-oriented dimensions arise from the analysis: *classroom management* and *student-oriented instruction*. In particular, teachers seem to stress the teaching content, the actual instructional process, and how to teach based on the perceptions of their profession. This resulted in a three-dimensional approach to instructional quality: *cognitive activation, teacher support,* and *classroom disciplinary climate*. However, students seem to focus on grasping the knowledge and improving their performance while observing teacher behaviors.

Looking at teacher self-reported data, the classroom management and student-oriented instruction dimensions—put forward by the students—were not apparent, but these dimensions could be seen as an integral part of teachers' three core dimensions of instructional quality. Teachers could consider these "extra" dimensions as part of their holistic perspective toward instructional quality. However, the teaching content and process are often

Table 9 The general instructional quality higher-order model fitted to the student data

	Model Fit				Teacher support	Cognitive activation	Disciplinary climate	Classroom management	Student-oriented instruction
	CFI	RMSEA	SRMR	X^2/df					
Australia	0.96	0.04	0.05	912/290	0.99	0.81	0.48	0.77	0.41
Finland	0.95	0.04	0.06	1481/289	0.86	0.71	0.35	0.79	0.28
Latvia	0.95	0.04	0.05	789/290	0.89	0.84	0.28	0.76	0.48
Portugal	0.96	0.04	0.05	1335/290	0.87	0.79	0.28	0.72	0.45
Romania	0.93	0.04	0.06	1422/287	0.71	0.80	−0.27	0.27	0.61
Singapore	0.95	0.04	0.05	2278/289	0.88	0.78	0.29	0.74	0.33
Spain	0.95	0.03	0.05	1509/290	0.81	0.86	0.20	0.64	0.46

influenced by factors that are unique in a local (national/regional) setting (e.g., course syllabus, time allocation, evaluation difference). These differences might help to explain the partially different constructs when building on teacher data. In addition, instructional quality can be influenced by differences in teacher professional knowledge, induction training, professional development, or educational policies that help shape educational systems. We can also not forget that teachers are influenced by peers in subject teams, and their school policies. It is, therefore, not surprising that differences are being observed between countries in the construct of instructional quality.

It is interesting to note that few studies have identified *student-oriented instruction* and *classroom disciplinary climate* as key dimensions of mathematical instructional quality (Scherer et al., 2016; Yi & Lee, 2017). *Student-orientated instruction* refers to the teacher's behaviors of providing specific tasks or lessons to encourage students to participate actively and having students work in small groups. *Classroom disciplinary climate* focuses on the creation of conditions that result in orderly learning and teaching environment, unlike *classroom management*,[3] which clusters teachers' actions to achieve learning goals and make effective use of lesson time (Martin et al., 2016; van Tartwijk & Hammerness, 2011). These two core dimensions can clearly be established on the base of the current study. Student-oriented instruction—in the context of the present study–describes a number of mathematics-specific instructional strategies that can be labeled as pedagogical content knowledge (PCK) as part of teachers' professional knowledge (e.g., the knowledge of specific-subject instructional strategies; the knowledge of students' cognitions and conceptions or misconceptions; Shulman, 1986). *Classroom disciplinary climate* reflects as such pedagogical knowledge (e.g., general theories and principles of classroom behaviors, how best to facilitate that learning in a variety of situations; Shulman, 1986). These two dimensions can be mirrored in parts of the Dynamic Model of Educational Effectiveness (DMEE). Though many authors argue the DMEE is a generic framework (e.g., Charalambous & Praetorius, 2018; Senden et al., 2022), our findings also stress that the DMEE can be seen as a hybrid framework that considers subject domain-specific features. Our results help as such identify shared and different theoretical features between generic and subject-specific approaches toward instructional quality.

[3] *Classroom management* is a more general concept that refers to all teacher decisions that help attaining the learning objectives at stake. This refers to strategies that provide structured tasks, time management, clear indicators of quality, setting priorities, and giving students responsibilities (Martin, et al., 2016). *Classroom disciplinary climate* points at the orderly learning environment defined by the interactions between teacher/student and student, the way competition between students is being handled, the teacher's treatment of students, and the setting of classroom discipline (Creemers & Kyriakides, 2008).

The items we applied to define *classroom management* focus on describing specific teacher actions that respond to student behavior (i.e., My teacher gets students to listen to him or her; My teacher keeps the class orderly; My teacher starts lessons on time). It is important to stress that the items operationalizing classroom disciplinary climate mirror the entire learning environment. In PISA, such items include "Students don't listen to what the teacher says; There is noise and disorder; The teacher has to wait a long time for students to < quiet down > ; Students cannot work well; students don't start working for a long time after the lesson begins." In TALIS 2013, the items are related to "I lose quite a lot of time because of students interrupting the lesson; When the lesson begins, I have to wait quite a long time for students to quiet down; Students in this class take care to create a pleasant learning atmosphere; There is much disruptive noise in this classroom.".

When using the CFA to test the construct instructional quality, two different dimensions (classroom management & classroom disciplinary climate) were identified in the model. Thus, in terms of theoretical conceptualizations, item wordings, and statistical models, we confirm that these are substantially different aspects of mathematical instructional quality.

Both dimensions reflect different aspects of instructional quality and might also result from a stronger focus on teaching disciplinary knowledge. As such, our findings can be aligned with other studies suggesting the three-dimensional framework is rather generic in nature and is less suitable for mapping specific knowledge domain-related quality features (Schlesinger & Jentsch, 2016; Schlesinger et al., 2018). The study results enrich the conceptual foundations of instructional quality models in mathematics education and add an international comparative dimension to the ongoing discussion.

Comparability of mathematical instructional quality across countries

An international cross-country comparison is an effective approach to uncovering less observable ubiquitous behaviors, thus providing policymakers and practitioners opportunities to examine current practices and beliefs about teaching (Riley et al., 2012). This approach is seen as an investigation of similarities and differences, enhancing national understanding of mathematics education by considering what constitutes "good classroom instruction" in different cultural contexts (Stigler et al., 2000). Specifically, cross-country comparisons enable researchers to gain an explicit understanding of the ways in which mathematics is taught in local contexts and the ways in which children learn mathematics within these contexts, as well as the ways in which mathematics is taught in schools in other countries.

In Latvia and Romania, the dimensionality of mathematics instructional quality from teacher perspectives are different from other countries. Not only was the factor structure of cognitive activation different in these two countries, but the dimension of teacher support was not identified. The insignificant variance of the latent variable could indicate the latent variable does not exist in a particular context or country. It can also mean that the latent variable cannot be identified with the available items due to the low construct validity of the items (simply the indicators are not good measures of the latent variables). It may be that this latent variable could be identified when using other items and resulting in different data. It is also important to repeat that the measurement structure of the latent variable instructional quality was derived from the pooled data, and that next we tested the factor structure while using data from the different individual countries (TALIS 2013). Applying the above to our findings in relation to Latvia and Romania, this could imply the available items are insufficient to identify the dimension of teacher support in these countries. Next to a potential lack of variance in these country data, it could also mean that the latent variable teacher support is connected to other variables when talking about mathematics instructional quality. This questions whether the variable can be determined as a single latent variable and suggests that teacher perceptions about instructional quality vary between countries in different ways than expected. When looking at the educational system of Latvia, the Ministry of Education and Science remains the primary policy decision-making body, responsible for drafting policy and legislation, setting academic standards, evaluating teachers' scientific and pedagogical performance, as well as organizing and coordinating its implementation (Andersone & Černova, 2007; Dirba, 2013). According to the Population Census 2011 (we applied TALIS 201 data), a large number of minority students have diverse cultural and multilingual backgrounds in Latvia (2011). The educational system in Romania is highly centralized. Assessing cross-group invariance with a confirmatory factor analysis approach is a robust procedure for testing (non)equivalence in multiple group data. The approach helps to overcome or address measurement bias and helps to interpret differences or similarities in a meaningful way between groups. In this way, we contributed to the literature by using international data to compare instructional quality from either teacher or student perspectives. Thus, in making comparisons across countries by applying international large-scale data, it is worthwhile to evaluate measurement invariance in this study.

Regarding the construct "instructional quality", neither the original TALIS 2013 nor PISA 2012 studies applied measurement invariance testing. Although earlier studies applied PISA 2012 data to test measurement invariance in the quality of mathematics teaching, they were mostly limited to focusing on the three-dimensional model. Moreover, the invariance measure for the quality of mathematics instruction in the TALIS 2013 data has thus far not yet been explored. We contribute as such to the literature by presenting new empirical evidence.

When comparing the factor structure based on teacher data, the configural level could not be reached. We were, therefore, not able to compare teacher perceptions of mathematical instructional quality between the seven countries. However, the factor structure based on student data was comparable between countries, considering the metric invariance of the measurement. This implies that a valid and meaningful comparison of mathematical instructional quality is possible when looking at the educational effectiveness in the seven countries.

Other considerations

The items we selected from TALIS were related to teaching practice, the teaching approach, and the learning environment. In the analyses, all 12 TALIS items were used to explore the factor structure of instructional quality. However, when applying the factor structure to each country, not all items seemed related to the instructional quality dimensions. For example, item TT2G43F, did not load on *teacher support* in Finland, and *teacher support* was not identified as a dimension of instructional quality in Romania and Latvia. In contrast to students' perceptions of their teachers' behaviors, which are more influenced by family background (e.g., SES; Wenger et al., 2020) and peers from group composition (Manski, 1993; Van Ewijk & Sleegers, 2010), international differences in teacher perceived instructional quality can again be attributed to the diversity in teachers' professional knowledge and resulting teaching behaviors, the difference in school contexts and conditions, or differences in educational system policies. 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 effective teaching. This implies that teachers have to adjust and apply effective teaching practices based on the characteristics of students or classroom compositions to adapt teaching to their needs. School factors influence teaching and learning by implementing and developing a school policy and creating a fitting school learning environment. Nonetheless, students, teachers, and schools are all part of a system or context that is influenced 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 learning environment, school leadership, or teaching style varies, pending the impact of these supra-school levels. It is plausible to explain that the factor structure of the teacher data (TALIS 2013) is not cross-country comparable.

In TALIS 2013, eight countries participated in the TALIS-PISA Link survey, but the data from only seven countries were analyzed in the present study. Initially, the data from Mexico were also part of the pooled data. However, analysis of the Mexico data resulted in a model that could not be compared in either TALIS 2013 or PISA 2012. This resulted in the Mexico data being removed in the current study. Future research could use Mexico as a case study to analyze the educational features of the system in terms of instructional quality.

Implications and conclusions

Worldwide, instructional quality is accepted as a determining factor for students' learning outcomes (see Atlay et al., 2019; Blömeke et al., 2016; Hattie, 2008; Nilsen & Gustafsson, 2016; Scherer et al., 2016; Yi & Lee, 2017). The quality of instruction reflects the behavior of teachers applying their professional knowledge in teaching activities. In mathematics education, this professional knowledge reflects disciplinary characteristics and imposes as such specific teaching demands. These are related to mathematical language and communication, mathematics connections and tasks, and mathematical reasoning and sense-making (Anghileri, 2006; Blömeke & Delaney, 2012; Chapin & O'Connor, 2007; Hunter, 2005).

In the literature, next to a diversity in conceptual frameworks to address instructional quality, also a diversity in measurement instruments (TALIS and PISA) is being presented. This leads to an inconsistent understanding of instructional quality. Even when measurement instruments seem alike in nature, they might vary in the type of information that can be extracted from the data. For instance, self-reported data can be collected from either students or teachers. Differences in the way teachers and students experience the same reality lead to differences in view on instructional quality. This also has implications for the theoretical grounding of the concept (Holzberger et al., 2013; Wagner et al., 2016).

As a multidimensional construct, scholars have argued that the nature of instructional quality can be conceptualized through (1) a generic or subject-specific multi-dimensional framework, (2) the adoption of different measurement instruments, (3) by building on data from teachers and students, and (4) by building on data from across educational systems with diverse cultural backgrounds (Bellens et al., 2019; Charalambous & Litke, 2018; Fischer et al., 2019; Wagner et al., 2016).

Considering the above, the current study examined the measurement properties of the concept "mathematics instructional quality" from teacher versus student perspectives in diverse educational systems to mirror school effectiveness in mathematics education across countries. We applied the linkage dataset—TALIS 2013 and PISA 2012—building on different international assessments. The redesign of the linkage database helped in focusing on additional variables in both OECD studies that contributed to the mapping of school mathematics culture and related instructional approaches. The following provides an extensive discussion of the findings and centers on: (1) the diversity in conceptualizations of mathematics-specific instructional quality using different measurement instruments, (2) the diversity in perceived mathematical instructional quality when building on either teacher or student perspectives, and (3) the diversity in mapping the nature of mathematical instructional quality when comparing countries.

The results suggest that different instruments are needed when measuring mathematics instructional quality that reflects unique teacher or student perspectives as well as additional dimensions when measuring mathematical instructional quality. Building on the available results, we can also start thinking about the way perceptions about instructional quality—in TALIS and PISA—can be linked to mathematics performance at the student- and school-level.

The findings confirm that mathematical instructional quality requires a specific factor structure, and that the three-dimensional framework does not hold for either teacher or student self-reported data. The study results also put forward empirical evidence about the shared nature and the potential differences in instructional quality when looking at key players in schools (teachers and students) and when comparing countries. We could confirm how teachers and students hold different perspectives, and how these perceptions reflect different dimensions of instructional quality. This adds to theoretical perspectives stressing the need to start from lived classroom experiences when looking and/or comparing instructional quality.

In addition, the cross-country nature of the data helped in developing a fresh insight into our understanding of education in diverse cultures and nations. The core feature of international large-scale assessments is that they are designed to establish comparability between diverse educational systems. The term 'comparability' refers to both the population that is being investigated as well as the validity of the instruments used across cultural boundaries. 'Validity' is the property of measurement and is the foundation for exploring the nature of latent constructs such as instructional quality. Johansson (2016) emphasizes the importance of (re) considering the concept of'validity' when looking at international comparative data. Rutkowski and Rutkowski (2018) point as such to meaningful cross-country comparisons that depend on item parameter equivalence in relation to the constructs being used and measured. Hence, 'measurement invariance' tests were considered in the current study. This methodology helped in ensuring the cross-cultural validity of the instruments and measurement procedures.

Although the present research reflects strengths, its limitations and directions for future research should not be forgotten. Because our newly designed database links TALIS and PISA data to conduct analyses at the school level, it is not possible to link individual teachers to specific groups of students in the classroom context. To some extent, this limit studying the construct "instructional quality". Our results did now uncover an "aggregated school level perspective" about instructional quality in mathematics education. This could neglect within-school variations. Next, we applied the multigroup CFA to test measurement invariance. Subsequent studies might consider using a more sophisticated method to check measurement invariance (e.g., Bayesian invariance testing, Alignment approach).

Taking this into account, future research could examine the relationship between the school levels–instructional quality and student achievement to study school effectiveness of mathematics education across countries. Wagner et al. (2016) point out how instructional quality can be linked to mathematics achievement–from both a teacher and student perspective. As such, our new linked TALIS-PISA database could be used to study in an international context (1) how and to what extent teacher and student perceptions are linked to mathematics performance and (2) whether differences in models of instructional quality—across countries—are also reflected in differences in performance. This will strengthen the comparative angle in studies about school effectiveness in mathematics education and respect the unique features of each educational system and how its multilevel structure interacts with key player behavior in the mathematics classroom.

Appendix 1

Applying four teacher sampling selection criteria to connect the teachers who participated in TALIS 2013 to the students who participated in PISA 2012 within schools.

The teachers: (1) 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 (time gap for administrating TALIS 2013 and PISA 2012 in Southern and Northern Hemisphere countries); (2) need to teach mathematics to 15-year-old students in the test administration school year; (3) should be teaching mathematics in the target class (the “target class” contains potential PISA pupils); (4) should have answered the Teacher Mathematics Module Questionnaire; teacher is as such a PISA mathematics teacher.

By using the anchor variable PISASCHOOLID, teachers from the specific school in TALIS-PISA Link linked the students and related variables in PISA 2012, the new linkage dataset is the Redesigned TALIS-PISA Link (rTPL). In other words, not yet possible to link the unique characteristics between teachers and students to reflect the instructional quality at classroom-level, the aggregating process is only tackled at school-level.

Appendix 2

The factor structures of instructional quality from teacher data (TALIS 2013) and student data (PISA 2012) in Mexcio are presented in the Tables 10, 11.

Table 10 The factor structure of instructional quality for Mexico is based on teacher perspectives

	Disciplinary Climate (DC-t)	
	Variable	Factor Loading
Mexico	TT2G41A†	0.83
	TT2G41B	0.64
	TT2G41C†	0.73
	TT2G41D†	0.80

† Item was reverse coded

Table 11 The factor structure of instructional quality for Mexico is based on student perspectives

	TS-p		CA-p		DC-p		CM-p	
	Variable	FL	Variable	FL	Variable	FL	Variable	FL
Mexico	ST77Q01†	0.71	ST80Q01†	0.68	ST81Q01	0.60	ST85Q01†	0.74
	ST77Q05†	0.69	ST80Q04†	0.41	ST81Q02	0.73	ST85Q02†	0.81
	M_TSMS1†	0.72	ST80Q05†	0.49	ST81Q03	0.75	ST85Q03†	0.72
	M_TSMS2†	0.84	ST80Q07†	0.62	ST81Q04	0.70		
	M_TSMS3†	0.78	ST80Q08†	0.68	ST81Q05	0.68		
			ST80Q09†	0.62	ST85Q04	0.39		
			ST80Q10†	0.63				
			ST80Q11†	0.57				
Correlation coefficient								
TS-p with CA-p	0.60							
TS-p with DC-p	0.20							
TS-p with CM-p	0.62							
CA-p with DC-p	*0.04*							
CA-p with CM-p	0.47							
DC-p with CM-p	0.40							

† Item was reverse coded; Italic indicates the non-significant estimate ($P > 0.05$). ST77, *M_TSMS* Mathematics teacher support; *ST79, ST80* Learning process; *ST81, ST85* Learning environment. *DC-p* classroom disciplinary climate; *TS-p* Teacher support; *CA-p* Cognitive activation; *CM-p* Classroom management; *FL* Factor loading

Compared to other countries, the dimensions of teacher support and cognitive activation could not be identified from teacher perspectives; that is, only the dimension of disciplinary climate can be validated in Mexico. As a separate dimension, it seems inappropriate to define it as instructional quality. From student perspectives, instructional quality can be defined by four dimensions, excluding student-oriented instruction. Meanwhile, in terms of teacher support, indicator ST80Q08 had a factor loading of approximately 0.10 and was removed from this dimension.

Appendix 3

See Table 12, 13.

Table 12 Instructional quality of teacher-related items in rTPL (TALIS 2013)

Variable	Item wording	Response category
Disciplinary climate		
TT2G41A[†]	I lose quite a lot of time because of students interrupting the lesson	1 Strongly disagree 2 Disagree 3 Agree 4 Strongly agree
TT2G41B	When the lesson begins, I have to wait quite a long time for students to quiet down	
TT2G41C[†]	Students in this class take care to create a pleasant learning atmosphere	
TT2G41D[†]	There is much disruptive noise in this classroom	
Support climate		
TT2G42C	I give different work to the students who have difficulties learning and/or to those who can advance faster	1 Never or almost never 2 Occasionally 3 Frequently 4 In all or nearly all lessons
TT2G43D	I provide written feedback on student work in addition to a < mark, i.e., numeric score or letter grade >	
TT2G43E	I let students evaluate their own progress	
TT2G43F	I observe students when working on particular tasks and provide immediate feedback	
Cognitive activation		
TT2M13C	I expect students to explain their thinking on complex problems	1 Never or almost never 2 Occasionally 3 Frequently 4 In all or nearly all lessons
TT2M13E	I connect mathematics concepts I teach to uses of those concepts outside of school	
TT2M13F	I encourage students to solve problems in more than one way	
TT2M13G	I require students to provide written explanations of how they solve problems	

[†] Item was reverse coded

Table 13 Instructional quality of student-related items in rTPL (PISA 2012)

Variable	Item wording	Response category
Teacher support		
ST77Q01†	The teacher shows an interest in every student's learning	0 Never or hardly ever 1 Some lessons 2 Most lessons 3 Every lesson
ST77Q02†	The teacher gives extra help when students need it	
ST77Q04†	The teacher helps students with their learning	
ST77Q05†	The teacher continues teaching until the students understand	
ST77Q06†	The teacher gives students an opportunity to express opinions	
ST83Q02†	My teacher provides extra help when needed	
ST83Q03†	My teacher helps students with their learning	
ST83Q04†	My teacher gives students the opportunity to express opinions	
Learning process: student-oriented instruction		
ST79Q03†	The teacher gives different work to classmates who have difficulties learning and/or to those who can advance faster	0 Never or hardly ever 1 Some lessons 2 Most lessons 3 Every lesson
ST79Q04†	The teacher assigns projects that require at least one week to complete	
ST79Q07†	The teacher has us work in small groups to come up with joint solutions to a problem or task	
ST79Q10†	The teacher asks us to help plan classroom activities or topics	

Table 13 (continued)

Variable	Item wording	Response category
Learning process: cognitive activation		
ST80Q01[†]	The teacher asks questions that make us reflect on the problem	0 Never or rarely 1 Sometimes 2 Often 3 Always or almost always
ST80Q04[†]	The teacher gives problems that require us to think for an extended time	
ST80Q05[†]	The teacher asks us to decide on our own procedures for solving complex problems	
ST80Q07[†]	The teacher presents problems in different contexts so that students know whether they have understood the concepts	
ST80Q08[†]	The teacher helps us to learn from mistakes we have made	
ST80Q09[†]	The teacher asks us to explain how we have solved a problem	
ST80Q10[†]	The teacher presents problems that require students to apply what they have learned to new contexts	
ST80Q11[†]	The teacher gives problems that can be solved in several different ways	
Learning environment: classroom disciplinary climate		
ST81Q01	Students don't listen to what the teacher says	0 Every lesson 1 Most lessons 2 Some lessons 3 Never or hardly ever
ST81Q02	There is noise and disorder	
ST81Q03	The teacher has to wait a long time for students to < quiet down >	
ST81Q04	Students cannot work well	
ST81Q05	Students don't start working for a long time after the lesson begins	
ST85Q04	The teacher has to wait a long time for students to < quiet down >	
Learning environment: classroom management		
ST85Q01[†]	My teacher gets students to listen to him or her	0 Strongly disagree 1 Disagree 2 Agree 3 Strongly agree
ST85Q02[†]	My teacher keeps the class orderly	
ST85Q03[†]	My teacher starts lessons on time	

[†] Item was reverse coded

Appendix 4

See Table 14.

Table 14 Descriptive statistics of the variables broken by the countries analysed from teacher perspectives in rTPL (TALIS 2013—TALIS-PISA Link)

Variable	Australia			Finland			Latvia			Portugal			Romania			Singapore			Spain		
	M	V	% Missing	M	V	% Missing	M	V	% Missing	M	V	% Missing	M	V	% Missing	M	V	% Missing	M	V	% Missing
TT2G41A[†]	2.94	0.65	0.24	2.62	0.69	0.00	2.96	0.46	0.56	2.62	0.63	0.37	3.23	0.41	0.51	2.86	0.54	0.00	2.70	0.73	0.14
TT2G41B	2.70	0.65	0.24	2.50	0.62	0.00	2.67	0.41	1.12	2.65	0.60	0.37	2.94	0.30	0.26	2.72	0.45	0.14	2.57	0.57	0.14
TT2G41C[†]	2.87	0.83	0.48	2.67	0.69	0.00	3.06	0.46	0.56	2.60	0.62	0.37	3.29	0.45	0.51	2.88	0.56	0.14	2.65	0.74	0.14
TT2G41D[†]	2.94	0.72	0.48	2.67	0.62	0.00	2.83	0.55	0.56	2.86	0.66	0.00	3.30	0.42	0.51	2.93	0.55	0.00	2.76	0.69	0.14
TT2G42C	2.45	0.71	0.24	2.69	0.62	0.00	–	–	–	2.45	0.55	0.37	–	–	–	2.10	0.44	0.00	2.13	0.66	0.68
TT2G43D	2.60	0.47	0.48	1.86	0.42	0.60	–	–	–	2.72	0.52	0.37	–	–	–	2.86	0.48	0.28	2.62	0.61	0.14
TT2G43E	2.10	0.58	0.72	1.92	0.38	0.60	–	–	–	2.46	0.46	0.37	–	–	–	2.22	0.47	0.70	1.78	0.49	0.27
TT2G43F	3.23	0.58	0.48	–	–	–	–	–	–	3.13	0.45	0.37	3.13	0.46	0.00	2.98	0.45	0.14	3.14	0.47	0.14
TT2M13C	2.98	0.59	0.24	2.92	0.55	0.90	2.85	0.46	0.56	3.40	0.35	0.37	–	–	–	2.70	0.49	0.28	2.63	0.55	0.27
TT2M13E	2.95	0.52	0.24	2.51	0.43	0.60	–	–	–	3.13	0.36	0.37	2.92	0.36	0.26	2.60	0.45	0.14	2.85	0.43	0.14
TT2M13F	2.90	0.53	0.24	2.48	0.58	0.90	2.88	0.45	0.56	3.20	0.37	0.37	3.15	0.36	0.00	2.92	0.38	0.28	3.06	0.42	0.14
TT2M13G	2.75	0.95	0.72	3.06	1.04	0.60	2.79	0.70	0.56	3.17	0.47	0.37	2.84	0.77	0.00	2.45	0.94	0.28	2.61	0.87	0.14

M Mean; *V* Variance

[†] Item was reverse coded

Acknowledgements Not applicable.

Author contributions XL: Conceptualization, Methodology, Formal analysis, Data Curation, Visualization, Writing–Original Draft; KYH: Conceptualization, Methodology, Validation, Formal analysis, Data Curation, Writing–Review & Editing, Supervision; JDN: Writing–Review & Editing, Supervision; MV: Conceptualization, Writing–Review & Editing, Supervision, Project administration.

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital). This work was supported by the China Scholarship Council (CSC) under Grant number CSC201807930019; the Fundamenteel Wetenschappelijk Onderzoek (FWO, Research Foundation—Flanders) under Grant number V412020N.

Availability of data and materials The study represents the secondary public data provided by OECD. The data and materials were retrieved from the TALIS website in 2013 from https://stats.oecd.org/index.aspx?datasetcode=talis_2013%20 and PISA website in 2012 from https://www.oecd.org/pisa/pisaproducts/pisa2012database-downloadabledata.htm

Declarations

Conflict of interest None.

Consent for publication All authors read and approved the final manuscript.

References

Aditomo, A., & Köhler, C. (2020). Do student ratings provide reliable and valid information about teaching quality at the school level? Evaluating measures of science teaching in pisa 2015. *Educational Assessment, Evaluation and Accountability, 32*(3), 275–310. https://doi.org/10.1007/s11092-020-09328-6

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.

Andersone, R., & Černova, E. (2007). Latvia. *The education systems of Europe* (pp. 418–437). Springer.

Anghileri, J. (2006). Scaffolding practices that enhance mathematics learning. *Journal of Mathematics Teacher Education, 9*(1), 33–52.

Atlay, C., Tieben, N., Hillmert, S., & Fauth, B. (2019). Instructional quality and achievement inequality: How effective is teaching in closing the social achievement gap? *Learning and Instruction, 63*, 101211. https://doi.org/10.1016/j.learninstruc.2019.05.008

Austin, B., Adesope, O. O., French, B. F., Gotch, C., Bélanger, J., & Kubacka, K. (2015). Examining school context and its influence on teachers: Linking TALIS 2013 with PISA 2012 student data. https://doi.org/10.1787/5js3f5fgkns4-en

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Bellens, K., Van Damme, J., Van Den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: Catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel sem analyses of TIMSS 2015 data in Flanders (Belgium), germany, and norway. *Large-Scale Assessments in Education, 7*(1), 1–27. https://doi.org/10.1186/s40536-019-0069-2

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM-Mathematics Education, 44*(3), 223–247.

Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 21–50). Springer Nature.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303–316. https://doi.org/10.1177/0049124189017003004

Borko, H., Stecher, B. M., Alonzo, A. C., Moncure, S., & McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment, 10*(2), 73–104.

Boston, M. (2012). Assessing instructional quality in mathematics. *Elementary School Journal, 113*(1), 76–104. https://doi.org/10.1086/666387

Brok, P. D., Bergen, T., & Brekelmans, M. (2006). Convergence and divergence between students' and teachers' perceptions of instructional behaviour in dutch secondary education. *Contemporary approaches to research on learning environments* (pp. 125–160). World Scientific. https://doi.org/10.1142/9789812774651_0006

Brown, J., & Kurzweil, M. (2017). *Instructional quality, student outcomes, and institutional finances*. American Council on Education.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. London: Guilford Publications.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*(4), 872–882.

Chapin, S. H., & O'Connor, C. (2007). Academically productive talk: Supporting students' learning in mathematics. In W. G. Martin, M. E. Strutchens, & P. C. Elliott (Eds.), *The learning of mathematics* (Vol. 69, pp. 113–128). National Council of Teachers of English.

Charalambous, C. Y., & Kyriakides, E. (2017). Working at the nexus of generic and content-specific teaching practices: An exploratory study based on timss secondary analyses. *The Elementary School Journal, 117*(3), 423–454. https://doi.org/10.1086/690221

Charalambous, C. Y., & Litke, E. (2018). Studying instructional quality by using a content-specific lens: The case of the mathematical quality of instruction framework. *ZDM-Mathematics Education, 50*(3), 445–460. https://doi.org/10.1007/s11858-018-0913-9

Charalambous, C. Y., & Praetorius, A. K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM-Mathematics Education, 50*(3), 355–366. https://doi.org/10.1007/s11858-018-0914-8

Chen, R. J. C., Lin, H. C., Hsueh, Y. L., & Hsieh, C. C. (2020). Which is more influential on teaching practice, classroom management efficacy or instruction efficacy? Evidence from talis 2018. *Asia Pacific Education Review, 21*(4), 589–599. https://doi.org/10.1007/s12564-020-09656-8

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*(1), 1–27. https://doi.org/10.1177/014920639902500101

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5

Clausen, M. (2002). *Unterrichtsqualität: Eine frage der perspektive?* Waxmann Verlag.

Cohen, J., Ruzek, E., & Sandilos, L. (2018). Does teaching quality cross subjects? Exploring consistency in elementary teacher practice across subjects. *Aera Open, 4*(3), 2332858418794492. https://doi.org/10.1177/2332858418794492

Commission, E. (2017). *Romania-Eurydice*. Retrieved 19 May 2022, from https://eacea.ec.europa.eu/national-policies/eurydice/content/romania_en.

Cordero, J. M., & Gil-Izquierdo, M. (2018). The effect of teaching strategies on student achievement: An analysis using talis-pisa-link. *Journal of Policy Modeling, 40*(6), 1313–1331. https://doi.org/10.1016/j.jpolmod.2018.04.003

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

Crossman, A., & Harris, P. (2006). Job satisfaction of secondary school teachers. *Educational Management Administration & Leadership, 34*(1), 29–46.

Davidov, E. (2008). A cross-country and cross-time comparison of the human values measurements with the second round of the European social survey. *European Survey Research Association, 2*, 33–46.

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*(1), 51–85.

Dirba, M. (2013). Education in Latvia. *Exploring childhood in a comparative context: An introductory guide for students*, 106.

Eid, M., & Koch, T. (2014). The meaning of higher-order factors in reflective-measurement models. Measurement: Interdisciplinary. *Research & Perspectives, 12*(3), 96–101. https://doi.org/10.1080/15366367.2014.943591

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Buttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fischer, J., Praetorius, A. K., & Klieme, E. (2019). The impact of linguistic similarity on cross-cultural comparability of students' perceptions of teaching quality. *Educational Assessment Evaluation and Accountability, 31*(2), 201–220. https://doi.org/10.1007/s11092-019-09295-7

Genç, M., & Çolakoğlu, Ö. M. (2021). Modeling the effects of instructional quality on mathematical literacy performance from the students' perspective: PISA 2012 turkey sample. *Egitim ve Bilim, 46*(206).

Gil-Izquierdo, M., & Cordero, J. M. (2018). Guidelines for data fusion with international large scale assessments: Insights from the talis-pisa link database. *Studies in Educational Evaluation, 59*, 10–18. https://doi.org/10.1016/j.stueduc.2018.02.002

Good, T. L., Wiley, C. R., & Florez, I. R. (2009). Effective teaching: An emerging synthesis. *International handbook of research on teachers and teaching* (pp. 803–816). Springer.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press.

Hamberger, A. (2010). Immigration and the integration of immigrants in Romania. *Migrationonline. cz, 7*.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Taylor & Francis.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. https://doi.org/10.3102/00028312042002371

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing of teachers' measures mathematics knowledge for teaching. *Elementary School Journal, 105*(1), 11–30.

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774–786. https://doi.org/10.1037/a0032198

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. *Woodcock-Johnson Technical Manual*, 197–232.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hunter, R. (2005). Reforming communication in the classroom: One teacher's journey of change. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: Research, theory and practice (proceedings of the annual conference of the mathematics education research group of australasia)* (pp. 451–458). MERGA.

Jentsch, A., & Schlesinger, L. (2017). *Measuring instructional quality in mathematics education* CERME 10, Dublin, Ireland. https://hal.archives-ouvertes.fr/hal-01949106.

Johansson, S. (2016). International large-scale assessments: What uses, what consequences? *Educational Research, 58*(2), 139–148.

Klieme, E. (2013). The role of large-scale assessments in research on educational effectiveness and school development. *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 115–147). Springer.

Klieme, E., Pauli, C., & Reusser, K. (2009). The pythagoras study: Investigating effects of teaching and learning in swiss and german mathematics classrooms. In J. Tomáš & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Künsting, J., Neuber, V., & Lipowsky, F. (2016). Teacher self-efficacy as a long-term predictor of instructional quality in the classroom. *European Journal of Psychology of Education, 31*(3), 299–322. https://doi.org/10.1007/s10212-015-0272-7

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*(3), 231–251.

Kunter, M., & Voss, T. (2013). The model of instructional quality in coactiv: A multicriteria analysis. *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 97–124). Springer. https://doi.org/10.1007/978-1-4614-5149-5_6

Kyriakides, L., Creemers, B. P. M., Panayiotou, A., & Charalambous, E. (2020). *Quality and equity in education: Revisiting theory and research on educational effectiveness and improvement*. Routledge.

Kyriakides, L., Georgiou, M. P., Creemers, B. P. M., Panayiotou, A., & Reynolds, D. (2017). The impact of national educational policies on student achievement: A European study. *School Effectiveness and School Improvement, 29*(2), 171–203.

Lanahan, L., McGrath, D. J., McLaughlin, M., Burian-Fitzgerald, M., & Salganik, L. (2005). *Fundamental problems in the measurement of instructional processes: Estimating reasonable effect sizes and conceptualizing what is important to measure.* Washington, DC: American Institutes for Research.

Latvia, National Statistical System of Latvia. (2011). *National statistical database.* Retrieved 19 May 2022, from https://data.stat.gov.lv/pxweb/en/OSP_PUB/START__POP__IR/.

Le Donné, N., Fraser, P., & Bousquet, G. (2016). *Teaching strategies for instructional quality: Insights from the talis-pisa link data*. OECD Publishing. https://doi.org/10.1787/5jln1hlsr0lr-en

Learning Mathematics for Teaching, P. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education, 14*(1), 25–47.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the pythagorean theorem. *Learning and Instruction, 19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 151–173. https://doi.org/10.1207/s15328007sem0902_1

Liu, X., Valcke, M., Yang Hansen, K., & De Neve, J. (2020). A technical report on linking talis 2013 and pisa 2012: Developing a multi-level perspective on the student, teacher and school-level variables related to mathematics achievement. In *12th international conference on education and new learning technologies. EDULEARN20 Proceedings*, pp. 7147–7153.

Liu, X., Valcke, M., Yang Hansen, K., & 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. *Frontiers in Education, 7*, 912837. https://doi.org/10.3389/feduc.2022.912837

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531–542.

Martin, N. K., Schafer, N. J., McClowry, S., Emmer, E. T., Brekelmans, M., Mainhard, T., & Wubbels, T. (2016). Expanding the definition of classroom management: Recurring themes and new conceptualizations. *Journal of Classroom Interaction, 51*(1), 31–41.

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale." *Educational Assessment, 13*(4), 267–300.

Maulana, R., & Helms-Lorenz, M. (2016). Observations and student perceptions of the quality of preservice teachers' teaching behaviour: Construct representation and predictive quality. *Learning Environments Research, 19*(3), 335–357.

Meitinger, K., Davidov, E., Schmidt, P., & Braun, M. (2020). Measurement invariance: Testing for it and explaining why it is absent. *Survey Research Methods, 14*(4), 345–349. https://doi.org/10.5167/uzh-192239

Moshagen, M. (2012). The model size effect in sem: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(1), 86–98. https://doi.org/10.1080/10705511.2012.634724

Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide: Statistical analysis with latent variables*. Wiley.

Nilsen, T., & Gustafsson, J.-E. (2016). *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time*. Springer Nature.

Nilsen, T., Gustafsson, J.-E., & Blömeke, S. (2016). Conceptual framework and methodology of this report. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality, and stu- dent outcomes: Relationships across countries, cohorts and time* (pp. 1–19). Springer. London. https://doi.org/10.1007/978-3-319-41252-8

OECD. (2013). *Talis 2013: Conceptual framework*. OECD Publishing. Retrieved 19 May 2022, from https://www.oecd.org/education/school/TALIS%20Conceptual%20Framework_FINAL.pdf.

OECD. (2014a). *Pisa 2012 technical report*. OECD Publishing. Retrieved 19 May 2022, from https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf.

OECD. (2014b). *Talis 2013 technical report*. OECD Publishing. Retrieved 19 May 2022, from http://www.oecd.org/education/school/TALIS-technical-report-2013.pdf.

OECD. (2019). *Talis 2018 technical report*. OECD Publishing. Retrieved 19 May 2022, from https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf.

Oliveri, M. E., & von Davier, M. (2011). Investigation of model fit and score scale comparability in international assessments. *Psychological Test and Assessment Modeling, 53*(3), 315.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. https://doi.org/10.3102/0013189x09332374

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2017). Interplay of formative assessment and instructional quality—interactive effects on students' mathematics achievement. *Learning Environments Research, 21*(1), 61–79. https://doi.org/10.1007/s10984-017-9240-2

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational researcher, 31*(7), 3–14.

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12. https://doi.org/10.1016/j.learninstruc.2013.12.002

Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C., & Reusser, K. (2007). Structure as a quality feature in mathematics instruction: Cognitive and motivational effects of a structured organisation of the learning environment vs. A structured presentation of learning content. *Studies on the educational quality of schools. The final report on the DFG Priority Programme*, pp. 101–120.

Rakoczy, K., Klieme, E., Burgermeister, A., & Harks, B. (2008). The interplay between student evaluation and instruction–grading and feedback in mathematics classrooms. *Zeitschrift Fur Psychologie-Journal of Psychology, 216*(2), 111–124. https://doi.org/10.1027/0044-3409.216.2.111

Rigdon, E. E., Becker, J. M., & Sarstedt, M. (2019). Parceling cannot reduce factor indeterminacy in factor analysis: A research note. *Psychometrika, 84*(3), 772–780. https://doi.org/10.1007/s11336-019-09677-2

Riley, P., Lewis, R., & Wang, B. (2012). Investigating teachers' explanations for aggressive classroom discipline strategies in China and Australia. *Educational Psychology, 32*(3), 389–403.

Rutkowski, L., & Rutkowski, D. (2018). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research, 62*(3), 354–367. https://doi.org/10.1080/00313831.2016.1261044

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7*, 1–16. https://doi.org/10.3389/fpsyg.2016.00110

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM-Mathematics Education, 48*(1–2), 29–40. https://doi.org/10.1007/s11858-016-0765-0

Schlesinger, L., Jentsch, A., Kaiser, G., Konig, J., & Blomeke, S. (2018). Subject-specific characteristics of instructional quality in mathematics education. *ZDM-Mathematics Education, 50*(3), 475–490. https://doi.org/10.1007/s11858-018-0917-5

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Senden, B., Nilsen, T., & Blömeke, S. (2022). Instructional quality: A review of conceptualizations, measurement approaches, and research findings. In M. Blikstad-Balas, K. Klette, & M. Tengberg (Eds.), *Ways of analyzing teaching quality: Potentials and pitfalls* (pp. 140–172). Scandinavian University Press.

Senk, S. L., Tatto, M. T., Reckase, M., Rowley, G., Peck, R., & Bankov, K. (2012). Knowledge of future primary teachers for teaching mathematics: An international comparative study. *ZDM Mathematics Education, 44*(3), 307–324.

Shi, D. X., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(1), 21–40. https://doi.org/10.1080/10705511.2017.1369088

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.

Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2020). *Measurement theory in action: Case studies and exercises*. Routledge.

Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist, 35*(2), 87–100.

Taasoobshirazi, G., & Wang, S. (2016). The performance of the srmr, rmsea, cfi, and tli: An examination of sample size, path size, and degrees of freedom. *Journal of Applied Quantitative Methods, 11*(3), 31–39.

Urdan, T. (2004). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology, 96*(2), 251–264. https://doi.org/10.1037/0022-0663.96.2.251

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Van Ewijk, R., & Sleegers, P. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement, 21*(3), 237–265.

van Tartwijk, J., & Hammerness, K. (2011). The neglected role of classroom management in teacher education. *Teaching Education, 22*(2), 109–112. https://doi.org/10.1080/10476210.2011.567836

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Van de Vijver, F. J. R., Avvisati, F., Davidov, E., Eid, M., Fox, J.-P., Le Donné, N., Lek, K., Meuleman, B., Paccagnella, M., & van de Schoot, R. (2019). Invariance analyses in large-scale studies. Oecd education working papers, no. 201. *OECD Publishing*.

Wagner, W., Gollner, R., Werth, S., Voss, T., & Schmitz, B. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108*(5), 705–721. https://doi.org/10.1037/edu0000075

Walkowiak, T. A., Berry, R. Q., Meyer, J. P., Rimm-Kaufman, S. E., & Ottmar, E. R. (2014). Introducing an observational measure of standards-based mathematics teaching practices: Evidence of validity and score reliability. *Educational Studies in Mathematics, 85*(1), 109–128.

Watson, A., & Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Mathematical Thinking and Learning, 8*(2), 91–111.

Wenger, M., Gärtner, H., & Brunner, M. (2020). To what extent are characteristics of a school's student body, instructional quality, school quality, and school achievement interrelated? *School Effectiveness and School Improvement, 31*(4), 548–575. https://doi.org/10.1080/09243453.2020.1754243

Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction, 66*, 101303.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research–a content analysis and recommendations for best practices. *Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Yi, H. S., & Lee, Y. (2017). A latent profile analysis and structural equation modeling of the instructional quality of mathematics classrooms based on the pisa 2012 results of korea and singapore. *Asia Pacific Education Review, 18*(1), 23–39. https://doi.org/10.1007/s12564-016-9455-4

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.