



# Are instructional practices different between East and West? An analysis of Grade 8 TIMSS 2019 data

Christian Bokhove<sup>1</sup> 

## Abstract

After each round of International Large-scale Assessments such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA), excellent achievements by countries in East Asia are explained in terms of their instructional practices. However, instructional practices are not easily captured in self-reported scales. In addition, speculations on what approaches lead to the highest mathematics achievement are often combined and conflated with discussions on how mathematical content needs to be taught. This article reports on a secondary analysis of TIMSS 2019 Grade 8 data from 46 jurisdictions of teacher-reported answers on instructional practices. Using a data-driven exploratory factor analysis, three instructional scales are compiled and supplemented with variables on barriers to instruction and instructional time. Multilevel models of all countries are created for students and teachers, in which the relationship between these instructional practices and mathematics achievement is explored. Across countries, the results show that the three new scales for instructional practice work in limited and different ways, but that barriers toward instruction, and to a lesser extent, instructional time, predict mathematics achievement. However, there is no consistent pattern of instructional practices across all countries, and also not among East Asian countries.

## Keywords

instructional practices, Grade 8, TIMSS, international comparison, multilevel models

Date received: 26 April 2022; accepted: 9 June 2022

## 1. Introduction

One of the persistent discussions in secondary mathematics education is the one about which instructional practices in classrooms are the most effective to learn mathematics. Often such discussions

---

<sup>1</sup>University of Southampton, Southampton Education School, Southampton, UK

### Corresponding Author:

Christian Bokhove, University of Southampton, Southampton Education School, Highfield Campus, Building 32/2035, Southampton SO17 1BJ, UK.

Email: [c.bokhove@soton.ac.uk](mailto:c.bokhove@soton.ac.uk)

cover more than just that tension, with emphasis on the approaches that lead to the highest mathematics achievement (e.g., inquiry or teacher-directed), the nature of tasks that students are asked to do (e.g., more open or more closed tasks), discussions about conceptual and procedural knowledge, or whether there should be more practice of basic skills or often vaguely formulated “problem-solving tasks.” In these initial sentences, it can be hard to seem neutral in describing the quandary, because tensions can be so high that even calling basic skills “rote” or referring to “real” understanding, can be seen as biased, leading to a rather fruitless polarization of the discussion. It is no coincidence that earlier instantiations of the debate have been termed the “Math Wars” (Schoenfeld, 2004). The proliferation of International Large-Scale Assessments (ILSAs) such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) has contributed to the debate as well, as still in the most recent results, many East Asian countries score higher than Western countries in both content and cognitive domains (Mullis et al., 2020; OECD, 2019). This unsurprisingly then leads to speculations as to *why* they do better, with associated claims about East Asian mathematics teaching practices. Already some time ago, Biggs (1994, 1998) highlighted the “Paradox of the Asian learner” where—according to Western views—seemingly traditional Asian classrooms aimed at “low-level cognitive goals” (e.g., emphasizing memorization), still lead to better outcomes for reasoning and other “higher order” skills. More recently, Leung (2001, 2017) described how the features in both East Asian and Western mathematics education are based on deep-rooted cultural values and paradigms. Leung’s (2001) study indicates how multiple perceived dichotomies seem to come together in comparisons between the East and West.

Two things can be noted here. Firstly, these factors often do not, whether we like it or not, stand on their own, and that the discussion is multifaceted. Secondly, it remains to be seen that these distinctions, although useful for our thinking, *really* are dichotomies. As Leung (2001) points out, Western scholars might “consider their East Asian counterparts as old fashioned in sticking to a product view of mathematics, but East Asians believe that their Western counterparts may have gone too far towards the process extreme” (p. 39). Leung (2001) also highlights that seeing memorization as learning without understanding is “too simplistic a view” (p. 41). Despite such nuance, Leung (2017) also asserts that East Asian countries form a cultural cluster. So, on the one hand, research has shown us that a multifaceted view is necessary; on the other hand, we still see caricatural discussions on the “best” way to teach mathematics, not seldom pointing toward those countries in East Asia that score high on international assessments and a reductive view on their instructional practices. This article aims to reinvestigate the theme of instructional practices in East and West with recent data from Grade 8 (lower secondary) TIMSS 2019.

We focus here on the “East” as in “East Asia,” and in particular the East Asian jurisdictions included in TIMSS 2019: Japan, Hong Kong SAR, Chinese Taipei, Korea, Singapore, and Malaysia. Based on 13 teacher-reported questionnaire items on instructional practices and a data-driven exploratory factor analysis (EFA), new scales for instructional practices are created. The scales are complemented with variables on barriers toward instruction and instructional time. The article then explores whether these practices are associated with mathematics achievement, including 43 of 46 jurisdictions countries in the TIMSS 2019 dataset, by creating multilevel models. The aims are three-fold: firstly, to show that discussions about instructional practices depend on the operationalization of the key concepts, a data-driven approach is used to create our instructional scales. Secondly, this article provides a contemporary analysis with the 2019 data. Thirdly, the article shows that with these scales, there seems to be no consistent pattern in differences in instructional practices between “East” and “West.”

## 2. Literature review

This study can be situated in many decades of discussions on the “best” way to teach mathematics in schools, which we here interpret as “leading to the highest mathematics achievement.” Although this

article does not aim to fully review the available literature, the impetus for the focus of this article lies in a prevailing quest to make generalizations about effective instructional practices in mathematics classrooms across the world. One of the influential publications on mathematics education in East Asia is the work by Leung (2001). In comparing West and East, Leung (2001) describes multiple dichotomies: product versus process; rote learning versus meaningful learning; studying hard versus pleasurable learning; extrinsic versus intrinsic motivations; whole-class teaching versus individualized learning; and competence of teachers: subject matter versus pedagogy. With such a range of variables, one question that arises is to what extent these approaches “occur together.” Are teachers who value process over product also inclined to prioritize conceptual knowledge? Do teachers who value pedagogy more than subject knowledge spend enough time on procedural knowledge? These are hard questions, especially in an international comparative context. We first review some of the more general literature on these dichotomies and then focus more on the literature about the distinction between East and West and studies that use ILSA data to do so. Finally, through some policy and media examples, we argue why our current study is relevant.

## 2.1 Components of mathematics teaching

Even without the international comparative element, there has been plenty of debate regarding the most effective ways in which to teach mathematics. The theme was central in what has been described as the “Math Wars” (Schoenfeld, 2004) in which an important issue concerned how students best acquire mathematical knowledge and skills: by practicing algorithms, or by focusing on reasoning and strategic problem-solving activities. In a synthesis of the available research on the concept of “mathematical proficiency,” Kilpatrick et al. (2001) synthesized five strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. Conceptual understanding was defined as “the comprehension of mathematical concepts, operations, and relations” and procedural fluency as the “skill in carrying out procedures flexibly, accurately, efficiently, and appropriately” (p. 116). The five strands were said to be interwoven and interdependent, with room for different representations and solution methods, so that students could “discuss the similarities and differences of the representations, the advantages of each, and how they must be connected if they are to yield the same answer” (p. 119). Here, we already see how teaching methods and types of knowledge also are interwoven. After all, a view that multiple solutions would need to be covered, requires a different teaching strategy than one where you use one algorithm and apply this one algorithm. The role of algorithms in education has combined numerous talking points (Fan & Bokhove, 2014). In numerous classrooms, “algorithms” have been seen as part of a broader discussion on “procedural fluency” and “conceptual understanding” (e.g., Brownell, 1945; Hiebert & Lefevre, 1986; Skemp, 1976; Star, 2005).

Rittle-Johnson et al. (2015) reviewed the literature on the longstanding debate about the relationship between conceptual and procedural knowledge, concluding that procedural knowledge supports conceptual knowledge, as well as vice versa, and that the relationship between the two is bidirectional. They also note that limited research has been done on specific sequencing of instruction. Of course, trying to provide more or less successful sequencing, does seem to imply that attempts to link types of knowledge to particular teaching methods and strategies, is what is being asked here. One form of learning and teaching that sometimes is mentioned in relation to more conceptual understanding is “generative learning” (see Brod, 2021, for a review). Generative (GEN) learning strategies are activities that prompt learners to *generate* meaningful information that goes beyond the instruction. In doing so, learners activate prior knowledge and link to the newly provided information, which aids the integration of the knowledge in existing knowledge structures. Two examples of generative strategies are elaborative interrogation and self-explanation, which, according to Brod (2021), show benefits for various student groups under varying learning conditions. In mathematics

education, Rittle-Johnson et al.'s (2017) meta-analysis of the literature on self-explanation prompts showed a moderate effect size for short-term improvement in conceptual knowledge. However, such effects were not always persistent and did not equate to time on task. Debates tend to mix different teaching practices. In their work on teaching practices, Pampaka and Williams (2016) explain how prior research has tried to relate student approaches to learning (strategic, surface/deep, etc.) to the teaching approaches literature. For example, an older review by Beattie et al. (1997) concluded that a fuller understanding of the “complex, composite and contingent nature of deep and surface learning and its interrelationship within the teaching-learning environment” (p. 1) is required. Beusaert et al. (2013) argued that a “teacher-centred approach predicts a surface approach to learning and a student-centered approach predicts a deep approach to learning” (p. 1). Finally, Chen and Li (2010) related Chinese teachers’ instructional practices and their possible impact on students’ learning, highlighting how teachers’ practices helped students build knowledge connections and coherence through lesson instruction. In sum, discussions about instructional practices seldom “stand alone” and are multifaceted.

## 2.2 Mathematics education in East and West

Numerous scholars have tried to describe the differences between West and East, covering multiple themes of Eastern and Western cultures (e.g., in addition to previous references, Bryan et al. 2007; Kaiser & Blömeke, 2013; Shimizu & Williams, 2012). In their analyses of mathematics education in the East and West, Leung (2006) and Leung et al. (2015) reported how culture mattered, especially a Confucian Heritage Culture, which included an examination culture, belief in effort, stress on memorization and practice, stress on reflection, discussion, and the Chinese language. However, Leung (2006) has also pointed out how even within the cluster of East Asian countries there are differences. For example, in comparing the results from a comparative study of primary school mathematics teachers in China and the United States by Ma (1999) and a small-scale replication of that study in Hong Kong SAR and Korea (Leung & Park, 2002), the majority of the teaching strategies reported by Hong Kong SAR and Korean teachers were procedurally rather than conceptually directed. This, however, then immediately introduces more discussion about what we mean by procedurally and conceptually; this is not so much a paradox, but a combination of both, for example in variation (Marton et al., 1996). Rather than seeing “memorization” and “understanding” as two mutually exclusive entities, in the context of East Asian countries, the starting point is that both work together to produce higher quality outcomes, not just in terms of test outcomes, but also in terms of metrics for mathematical understanding (Fan & Bokhove, 2014; Hess & Azuma, 1991; Marton et al., 1996; Rittle-Johnson et al., 2015).

In trying to decipher the concrete differences between West and East, Dahlin and Watkins (2000) argue that the linking pin between both is the role of repetition. Meaningful repetition can create “a deep impression”, which leads to memorization, and it can also lead to “discovering new meaning” (Dahlin & Watkins, 2000, p. 66), which in turn leads to understanding (Li, 1999). Such literature again gives the impression that many of the concepts associated with conceptual and procedural knowledge, also are associated with teaching methods. Cheng (2014) explicitly linked teaching practices to “procedural” and “conceptual,” positing a priori that teaching quality might be related to certain types of instructional practices. However, using TIMSS 2011 data, no combinations of the components of conceptual and procedural mathematics teaching practices were apparent across five high-performing Asian education systems. Despite this, scholars such as Leung et al. (2015) argued that there are “distinctive features in the classroom teaching and teacher education and development in East Asian countries which are markedly different from the corresponding practices in Western countries” (p. 137). Leung (2017) highlights three studies (Guo, 2014; Lie, 2014; Wu, 2009) that show such a “cultural cluster.”

### 2.3 ILSAs and instructional practices

Research with ILSAs data has also tried to unpick instructional practices in the mathematics classroom. In the seminal TIMSS 1999 Video Study, Hiebert et al. (2005) reported on results from seven countries. The countries participating in the mathematics portion of the TIMSS 1999 Video Study were Australia, the Czech Republic, Hong Kong SAR, Japan, the Netherlands, Switzerland, and the United States. Among the striking differences were that for applications of mathematics, the United States was the lowest with 34% problems per lesson, and Japan high with 74%. However, Hong Kong SAR was at the low end with 40%. On repeating procedures, Hong Kong SAR (81%), scored high just like the United States and European countries, but Japan was low. Both Hong Kong SAR and Japan were relatively low on reviewing and high on introducing and practicing new content. These observational results indicate instructional differences, also between East Asian jurisdictions.

With a focus on PISA 2012 mathematics results, Caro et al. (2016) looked at subject-specific teaching strategies related to mathematics performance of students across education systems whilst considering curvilinear associations and interactions with the socioeconomic and instructional context. They found that cognitive activation and teacher-directed instruction were positively related to mathematics performance, but that too high levels of the latter were negative for student performance. Associations of student-oriented strategies with mathematics performance were inconsistent. Wu et al. (2020) also used PISA 2012 data to examine the use of learning strategies in mathematics among East Asian students in East Asian educational systems. Their latent class analysis of the data found four classes of learning strategy types. The majority of the students in all seven East Asian educational systems utilized metacognitive strategies in combination with memorization. The authors concluded that “the cognitive processes employed by students of East Asian backgrounds are more complex and nuanced than the previous perception that they relied heavily on memorization” (Wu et al., 2020, p. 643).

Again using PISA 2012, Echazarra et al. (2016) studied the predictive nature of four instructional scales: an index of teacher-direct instruction, an index of student-oriented instruction, an index of cognitive activation, and an index of formative-assessment instruction. Theoretically, the authors seem to include assumptions for particular practices, for example, that “teacher-directed instruction risks leaving students in a passive and disengaged role, and performing monotonous tasks” (Echazarra et al., 2016, p. 34), student-centered teaching methods giving “students a more active role in classroom processes” (Echazarra et al., 2016, p. 34), and “[i]n order to foster conceptual understanding and students’ motivation to learn, teachers have to use content and stimuli that are challenging for students and demand the use of higher-order skills” (Echazarra et al., 2016, p. 34). The authors also assert that “teacher-directed and student-oriented strategies represent somewhat opposite ideas about instruction” (p. 48). However, there are substantial differences across all countries and East Asian countries in particular. For teacher-directed instruction, Japan, Shanghai–China, and Korea were at the higher end, Chinese Taipei, Singapore, and Hong Kong SAR more in the middle, and Malaysia at the lower end in the relative use of teacher-direct instruction. For student-oriented instruction, Singapore, Shanghai–China, and Hong Kong SAR were at the lower end, while Korea, Malaysia, and Japan were at the upper end. For formative assessment, Korea and Japan were at the bottom, while Malaysia, Hong Kong SAR, and Singapore more toward the higher end. Finally, for cognitive activation, Singapore and Hong Kong SAR were in the top half, while Malaysia and Korea toward the lower end. Echazarra et al. (2016) also looked at memorization practices reporting that their findings “challenges some deep-rooted beliefs about the over-reliance on rote learning and memorisation in East Asian countries, particularly those sharing a Confucian heritage” (p. 75).

### 2.4 Views on instructional practices in policy and media

Although we have already seen a tension between on the one hand a view that there are “cultural clusters” and on the other hand, instructional practices are multifaceted and firmly set in a local

context, another factor in this debate is the fact that policies and media are still very much immersed in caricatures of instructional practices in East Asia. A recent policy example is the mention of Asia in a research review by the inspectorate in England (Ofsted, 2021). In discussing Asian practices, a mix of practices emerges. For example, Lim (2007) is cited several times in the review, who investigated characteristics of effective mathematics teaching in five Shanghai schools. She concluded that these classrooms showed teaching with variation, an emphasis on precise and elegant mathematical language, an emphasis on logical reasoning, mathematical thinking and proofing during teaching, order and serious classroom discipline, strong coherent teacher–student rapport, and a strong collaborative culture amongst mathematics teachers. The review also highlights Asian examples such as the Mathematics Teacher Exchange with Shanghai (Boylan et al., 2018) and homework practices in Singapore. In newspapers, results from ILSA often are set against those from the top-performing countries, for example, after the PISA 2018 results: “The global trend shows Asian school systems such as China and Singapore getting the best results” (Coughlan, 2019) with associated speculation about the many factors that purportedly cause this effectiveness. The overarching view seems to be that we still need further insight into the instructional practices in East and West.

## 2.5 The current study

The current study, then, combines several instructional elements regarding possible differences between East Asian countries and other countries. Firstly, it uses recent data from TIMSS 2019 to explore the issue. TIMSS is a dataset that is more tightly related to countries’ curriculum choices than for example PISA (Rindermann & Baumeister, 2015) and has data from secondary schools’ Grade 8. Secondly, TIMSS’ sampling of classrooms means that questionnaire results of mathematics teachers can be used to combine different instructional practices in coherent scales. As we use TIMSS 2019 data, instructional practices are those practices that either teachers or students do or are asked to do during a mathematics lesson in the classroom. Although the International Association for the Evaluation of Educational Achievement (IEA) has created a dedicated scale for “Teachers’ Emphasis on Science Investigation” in the Grade 8 TIMSS 2019 dataset (Martin et al., 2020, p. 16.320), there are no 2019 scales for instructional practices in mathematics classrooms, despite questionnaire items being available on this. Rather than base the composition of these scales on prior literature—we have already seen that there are many differing views on the essential instructional components of such scales—scales will be created by performing an EFA (RQ1). We then use the scales as predictors of mathematics achievement in multilevel models we create for each country (RQ2). Finally, we explore if there is a pattern in the predictive power of instructional practices with regard to “East” and “West” (RQ3). The questions that will be addressed are:

RQ1. What coherent scales for instructional practices can be created based on TIMSS 2019 Grade 8 data?

RQ2. To what extent do these scales for instructional practices predict mathematics achievement in the 46 jurisdictions of the TIMSS 2019 Grade 8 dataset?

RQ3. Are there consistent patterns in instructional practices in different areas of the world, with an emphasis on East Asia?

## 3. Methodology

The Trends in Mathematics and Science Study (TIMSS) in 2019 is the seventh assessment cycle of TIMSS, conducted by the IEA.<sup>1</sup> TIMSS 2019 was conducted in the fourth and eighth grades in 64 countries and 8 benchmarking systems. TIMSS has been conducted since 1995, with assessments conducted every four years. As part of their study, TIMSS collects mathematics achievement data,

and also contextual questionnaires from students, principals of schools, and mathematics teachers. In this analysis, data from the grade eight international database of the TIMSS 2019 cycle for 46 jurisdictions are used by combining data from students and mathematics teachers (Fishbein et al., 2021). TIMSS 2019 uses a two-stage sampling approach, first by sampling schools and then within the school by selecting one (or more) intact class from the target grade of each participating school.<sup>2</sup>

### 3.1 Analytical approach

To answer the first research question, an EFA of the teacher-reported data is conducted, which is reported in the section on instructional scales. To answer the second research question, multilevel models per country are constructed, because data are clustered as students nested in teachers/classrooms (Stapleton et al., 2016). The third research question is answered by observing the patterns in the descriptive statistics and results of our multilevel models. Through multilevel modeling, an adaptation of the general linear model for hierarchical datasets, we can explore the impact of student and teacher characteristics on students' achievement (Snijders & Bosker, 2012). We use the package BIFIEsurvey (Robitzsch & Oberwimmer, 2022), within the statistical packages R and Rstudio (R Core Team, 2021; RStudio Team, 2021), which enables us to compute hierarchical two-level models with random intercepts and random slopes. Full maximum likelihood estimation is conducted by means of an EM algorithm (Raudenbush & Bryk, 2002). In our approach, the complex sampling design of TIMSS 2019 data is taken into account (Rutkowski et al., 2010). This means that three aspects are included in the analyses. To cater for different probabilities of units being selected (classrooms/teachers and students) sampling weights are used. Sampling weights ensure that the choice of sampling design does not have an undesired effect on the analyses of data. In this case, weights for mathematics teachers are used at the classroom level (level 2, MATWGT), and weights at the student level (level 1) are calculated from the total weight and level 2 weights by the BIFIEsurvey package.<sup>3</sup> As TIMSS 2019 uses an incomplete and rotated-booklet design for testing children on the major outcome variables, five plausible values (PVs) for achievement are used. To correctly analyze, the BIFIEsurvey package combines the five plausible values into a single set of point estimates and standard errors using Rubin's rules (Rubin, 1987). A final aspect concerns the variance estimation. As standard variance estimation formulas are not appropriate for data obtained with a complex sample design (Rutkowski et al., 2010), a Jack-knife procedure is used. In this study, it is relevant to note that mathematics teachers were not randomly sampled; they are the mathematics teachers of randomly sampled classrooms. An R script was created that generates descriptive statistics and the model results.

### 3.2 Variables

The following variables are included in the analyses.

As dependent variable(s), the five plausible values are used for mathematics achievement, BSMMAT01 to BSMMAT05. These are imputed scores averaging 500 and a standard deviation of 100 (Martin et al., 2020).

### 3.3 Instructional scales (teacher level)

As independent variables, scales are created from variables included in the teacher questionnaire. In operationalizing the pertinent activities in the classroom, contrary to, for example, similar types of scales in PISA 2015, self-reported behaviors from mathematics teachers are deemed more trustworthy than that of students. A few questions in the mathematics teacher questionnaire are particularly relevant for instructional approaches.<sup>4</sup>

Firstly, Question 12 in the teacher questionnaire asked “how often do you do the following in teaching this class” with teachers asked to answer on a four-point scale (1 = every or almost every lesson, 2 = about half the lessons, 3 = some lessons, 4 = none) for seven activities, as mentioned in Table 1 (BTBG12A-G). A second question (Question 15, BTBM15A-F) asked teachers “In teaching mathematics to this class, how often do you ask students to do the following” with the same answer options. Two items in Question 15, pertaining to mixed or same-ability groups, were not used in this analysis, as they pertained to the way students were organized within the classroom, not specific instructional approaches.

All items were first reverse coded (4 became 1, 3 became 2, 2 became 3, and 1 became 4) so that “higher” denoted “more.” Then an EFA with varimax rotation was conducted with the 13 questionnaire items, using the *psych* package in R (Revelle, 2022). A scree plot indicated that a three- or four-factor solution would fit best, but upon trying a four-factor solution, one questionnaire item (BTNM15F) was not loaded on a factor. Furthermore, three questionnaire items loaded on two factors. Regarding model fit for the three-factor solution, the RMSEA index was 0.051; according to Hu and Bentler (1999), this indicates a close fit. The Tucker–Lewis Index was 0.936—an acceptable value considering it is over 0.9. Finally, the RMSR was 0.05, with values up to 0.08 being acceptable (Hu & Bentler, 1999). Using a threshold for loadings over 0.3, Table 1 indicates the three factors we translated in scales. BTBG12A-G averaged together to make a scale we called “Generative” (GEN) as we felt the items asked students actively generate knowledge ( $\alpha = 0.779$ ). BTBM15A-C averaged together to make a scale that refers to more teacher activity, so we called it “Teacher-Active” (TA,  $\alpha = 0.725$ ). Finally, we averaged BTBM15D-F to make a scale that referred to as “Practice-on-Own” (PO,  $\alpha = 0.609$ ).<sup>5</sup> Note that although the EFA resulted in disjunct scales, this study does not see the scales as mutually exclusive, as it is perfectly possible to employ generative instructional practices, teacher direction, and space for independent practice in the same lesson, something the TIMSS 1999 video study has also shown (Hiebert et al., 2005). Another question asked teachers “In a typical week, how much time do you spend teaching mathematics to the students in this class? (minutes)” (BTBM14), which is included as instructional time. As the study focuses on instruction, homework was not included. In line with the literature on “Opportunity to

**Table 1.** Operationalization of instructional scales GEN, TA, and PO.

In TIMSS 2019		Scale		
Name variable	Item	GEN	TA	PO
BTBG12A	Relate the lesson to students' daily lives	X		
BTBG12B	Ask students to explain their answers	X		
BTBG12C	Ask students to complete challenging exercises that require them to go beyond the instruction	X		
BTBG12D	Encourage classroom discussions among students	X		
BTBG12E	Link new content to students' prior knowledge	X		
BTBG12F	Ask students to decide their own problem-solving procedures	X		
BTBG12G	Encourage students to express their ideas in class	X		
BTBM15A	Listen to me explain new mathematics content		X	
BTBM15B	Listen to me explain how to solve problems		X	
BTBM15C	Memorize rules, procedures, and facts		X	
BTBM15D	Practice procedures on their own			X
BTBM15E	Apply what they have learned to new problem situations on their own			X
BTBM15F	Work problems together in the whole class with direct guidance from me			X

Note: GEN = generative scale; TA = teacher-active scale; PO = practice-on-own scale.



Learn,” one expectation was that instructional time might mediate both achievement and instructional practices (e.g., Bokhove et al., 2019; Carroll, 1963). The IEA created the “Classroom Teaching Limited by Students Not Ready for Instruction” (BTBGLSN) scale based on teachers’ responses to questions about barriers to teaching their class, including those on lacking prerequisite knowledge or skills, a lack of basic nutrition, a lack of sleep, absence from class, disruption by students, uninterested students, mental, emotional or psychological impairment; and students having difficulties understanding the language of instruction. The scale was standardized to a mean of 10 and a standard deviation of 2. With this scale, higher scores mean “better,” so, in this case, fewer barriers and limitations to teaching.<sup>6</sup> This variable is included in the analyses because it is expected that all instructional practices can potentially be influenced negatively because of such barriers.

### 3.4 Control variables

The Home Educational Resources (BSBGHER) scale was created based on students’ reports regarding the availability of three home resources which are included as a proxy for socioeconomic status (SES). SES is a longstanding predictor of academic achievement (e.g., for a review, see Sirin, 2005). The scale contains items on the number of books at home, the highest level of education of either parent, and the number of home study supports. The scale was standardized to a mean of 10 and a standard deviation of 2. With this scale, higher scores mean “better”: more resources and thus higher SES. Finally, gender is included as a control variable (ITSEX, 1 = female, 2 = male).

### 3.5 Model building

Several models per country are created with the five PVs for mathematics achievement as dependent variables: a null model, Model 1 which adds two student control level variables BSBGHER and ITSEX, Model 2 which adds the three newly created instructional scales GEN, TA, and PO, and a Model 3 which further adds the limitations to teaching scale and instructional time. Although our analysis script contains code for all interim models, this paper only reports on the final model. We do not treat missing data, but use listwise deletion. No centering was used. As two countries’ models did not converge, the final sample included 44 countries with 260,111 students for the descriptive statistics.

## 4. Results

Table 2 presents descriptive statistics for all included variables. Firstly, we can confirm that mathematics achievement in five of the six East Asian jurisdictions, namely Chinese Taipei, Hong Kong SAR, Japan, Korea, and Singapore are among the highest. Only Malaysia scores much lower, but this is in line with Malaysia not typically being included (yet) as an example of East Asian higher mathematics achievement. We can see substantial differences in instructional time between countries as well with Japan and Korea at the low end of total instructional time per week. This might give the impression that East Asian countries have less instructional time across the board, but that is not the case with the other jurisdictions more in the middle of the pack. Nevertheless, it is notable that more instructional time is often associated with higher achievement, but just like Japan in Bokhove et al. (2019), this is not necessarily the case. All values for the three instructional practices scales GEN, TA, and PO, are over the midpoint of 2.5 for all jurisdictions, suggesting that these practices occur frequently and are not mutually exclusive (i.e., when one score is high, the other can also be high). However, it also is clear that practices vary considerably across countries. For example, Dubai and Bahrain score highest for generative practices (GEN) with little spread, with Chinese Taipei and Hong Kong SAR scoring lowest. In fact, it is notable that East Asian countries all

**Table 2.** Descriptive statistics for key variables.

Country	BSMMAT		BSBGHER		ITSEX		GEN		TA		PO		BTBGLSN		BTBM14		
	Student level								Instructional variables		teacher level						
	N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Australia	9,281	517.56	88.99	11.12	1.44	1.50	0.50	3.18	0.47	3.28	0.59	3.35	0.53	9.86	2.04	213.90	46.36
Bahrain	5,725	481.09	95.26	10.20	1.42	1.51	0.50	3.59	0.38	3.63	0.54	3.60	0.47	9.91	1.46	265.67	52.34
Chile	4,319	441.56	76.61	9.88	1.39	1.51	0.50	3.47	0.46	3.33	0.62	3.46	0.56	9.20	1.87	306.38	106.93
Chinese Taipei	4,929	612.12	98.20	10.37	1.59	1.50	0.50	2.78	0.50	3.27	0.62	2.84	0.63	9.58	1.95	233.71	43.19
Cyprus	3,521	501.08	81.53	10.93	1.40	1.51	0.50	3.34	0.42	3.49	0.48	3.31	0.56	9.54	1.97	185.20	57.23
Finland	5,045	508.10	73.40	11.14	1.35	1.52	0.50	3.07	0.47	3.20	0.52	3.14	0.52	10.30	1.68	175.39	49.34
France	4,094	481.99	68.53	10.56	1.47	1.51	0.50	3.00	0.34	3.27	0.59	3.11	0.50	9.78	1.59	191.35	49.36
Georgia	3,315	461.30	88.05	10.83	1.48	1.52	0.50	3.31	0.36	3.49	0.70	3.26	0.57	9.77	1.44	193.99	46.25
Hong Kong SAR	3,282	579.70	90.99	10.31	1.56	1.54	0.50	2.74	0.50	2.98	0.60	2.86	0.59	10.47	1.56	241.88	79.88
Hungary	4,569	516.54	89.41	10.91	1.51	1.50	0.50	3.23	0.44	3.01	0.65	3.28	0.57	10.61	1.94	192.45	52.16
Islamic Republic of Iran	5,980	446.17	92.80	9.50	1.79	1.53	0.50	3.24	0.45	3.53	0.47	2.99	0.57	8.92	1.77	185.53	63.27
Israel	3,778	520.95	100.37	11.08	1.45	1.48	0.50	3.30	0.50	3.26	0.63	3.42	0.55	9.01	2.47	250.77	58.78
Italy	3,619	497.48	71.28	10.33	1.55	1.50	0.50	3.38	0.49	3.37	0.67	3.27	0.56	10.40	2.01	233.56	60.33
Japan	6,113	595.25	84.16	10.89	1.32	1.48	0.50	2.75	0.52	3.33	0.70	3.26	0.60	12.30	2.07	154.25	28.75
Kazakhstan	4,532	488.68	82.37	10.20	1.27	1.51	0.50	3.55	0.37	3.63	0.61	3.51	0.46	10.34	2.17	207.15	57.39
Jordan	7,176	420.27	86.66	9.63	1.41	1.52	0.50	3.39	0.41	3.61	0.58	3.39	0.54	9.33	1.49	216.32	58.33
Republic of Korea	4,125	607.08	97.07	11.63	1.45	1.52	0.50	2.92	0.54	3.34	0.53	3.49	0.51	9.08	1.80	163.56	54.04
Kuwait	4,574	402.75	87.72	9.92	1.40	1.47	0.50	3.23	0.46	3.48	0.63	3.42	0.58	9.68	1.44	232.78	66.31
Lebanon	4,730	429.31	72.49	9.74	1.56	1.51	0.50	3.31	0.50	3.57	0.56	3.20	0.63	9.51	1.53	305.68	74.58
Lithuania	3,826	520.43	81.64	10.82	1.27	1.50	0.50	3.18	0.41	3.56	0.61	3.13	0.56	9.33	2.09	203.06	44.43
Malaysia	7,065	460.57	90.67	9.60	1.51	1.49	0.50	3.10	0.54	3.08	0.66	3.04	0.61	9.34	1.95	194.41	48.54
Morocco	8,458	388.19	71.64	8.28	1.96	1.50	0.50	3.14	0.47	3.56	0.57	3.36	0.61	9.20	1.52	241.94	52.03
Oman	6,751	410.66	98.96	9.73	1.63	1.52	0.50	3.43	0.38	3.70	0.52	3.48	0.48	9.14	1.68	279.99	57.16
Norway	5,142	501.35	78.10	11.36	1.37	1.51	0.50	3.00	0.43	2.69	0.58	2.99	0.56	9.87	1.66	168.24	49.89
Portugal	3,377	500.32	74.02	10.24	1.61	1.50	0.50	3.16	0.47	3.39	0.56	3.41	0.54	9.49	1.53	218.21	52.43
Qatar	4,078	447.63	95.68	10.37	1.45	1.52	0.50	3.49	0.46	3.41	0.71	3.52	0.58	9.80	1.70	239.22	85.08
Romania	4,494	478.98	97.59	10.28	1.52	1.49	0.50	3.34	0.45	3.41	0.59	3.38	0.63	10.29	1.83	222.71	72.56
Russian Federation	3,946	543.42	81.03	10.56	1.31	1.51	0.50	3.24	0.42	3.28	0.60	3.05	0.53	9.86	2.24	241.68	73.94
Saudi Arabia	5,680	393.77	77.97	9.47	1.55	1.51	0.50	3.44	0.45	3.52	0.63	3.45	0.53	9.52	1.72	222.02	99.62

(continued)

**Table 2.** (continued)

Country	N	BSMMAT		BSBGHER		ITSEX		GEN		TA		PO		BTBGLSN		BTBM14	
		Student level								Instructional variables		teacher level					
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Singapore	4,853	615.77	88.69	10.56	1.46	1.51	0.50	2.82	0.52	3.18	0.58	3.20	0.58	10.60	1.82	217.96	72.01
South Africa	20,936	389.55	76.90	9.06	1.54	1.48	0.50	3.19	0.51	3.25	0.66	3.37	0.59	8.93	1.53	277.19	64.26
Sweden	3,996	502.52	76.53	10.98	1.51	1.51	0.50	3.13	0.47	3.10	0.62	3.11	0.51	10.27	1.67	174.50	25.20
United Arab Emirates	23,950	478.72	103.32	10.42	1.50	1.52	0.50	3.59	0.43	3.16	0.75	3.47	0.57	10.43	2.09	280.14	93.92
Turkey	4,077	495.63	109.36	9.47	1.83	1.50	0.50	3.38	0.46	3.35	0.48	3.51	0.55	9.32	2.07	238.09	89.40
Egypt	7,210	412.88	95.45	9.45	1.62	1.45	0.50	3.27	0.43	3.56	0.49	3.42	0.55	9.35	1.49	214.79	40.91
United States	8,698	515.44	98.93	10.70	1.56	1.51	0.50	3.38	0.45	3.10	0.65	3.50	0.50	9.72	1.82	255.65	91.02
England	3,445	514.52	89.29	10.73	1.49	1.47	0.50	3.14	0.44	3.21	0.51	3.31	0.53	10.39	1.90	200.81	53.47
United Arab Emirates (Dubai)	6,725	540.22	89.87	10.80	1.42	1.50	0.50	3.68	0.34	3.02	0.78	3.51	0.55	11.05	2.09	254.17	74.28
United Arab Emirates (Abu Dhabi)	8,656	441.00	105.13	10.28	1.52	1.53	0.50	3.45	0.49	3.13	0.73	3.41	0.57	9.82	1.91	293.83	103.98
Canada (Ontario)	3,776	529.75	78.80	11.22	1.31	1.50	0.50	3.51	0.43	2.84	0.65	3.48	0.49	9.70	1.73	298.20	55.85
Canada (Quebec)	3,201	542.72	66.74	10.91	1.26	1.49	0.50	2.93	0.53	3.21	0.61	3.04	0.68	10.54	1.91	246.75	60.75
Russian Federation, Moscow	3,972	578.85	85.83	11.36	1.26	1.52	0.50	3.31	0.43	3.24	0.62	3.21	0.55	9.94	2.40	267.64	80.52
South Africa (Gauteng)	5,633	420.70	76.67	9.49	1.49	1.45	0.50	3.24	0.52	3.26	0.65	3.33	0.59	8.93	1.61	279.64	81.86
South Africa (Western Cape Province)	5,458	440.74	88.72	9.30	1.72	1.45	0.50	2.97	0.50	3.32	0.65	3.28	0.64	8.91	1.88	264.74	60.15

Note. East Asian jurisdictions in grey. N Refers to the Maximum Sample Size. M = mean; SD = standard deviation; BSMMAT = mathematics achievement. BSBGHER = home educational resources scale; ITSEX = gender (female = 1, male = 2); GEN = generative scale; TA = teacher-active scale; PO = practice-on-own scale; BTBGLSN = "Classroom Teaching Limited by Students Not Ready for Instruction" scale; BTBM14 = instructional time in minutes per week.

score well below average, with five out of six jurisdictions, under 3. In contrast, Middle Eastern countries, except Georgia, have higher than average GEN scores. For the TA scale, Bahrain, Kazakhstan, and Oman score highest, while Norway and Hong Kong SAR score lowest for the measure. Finally, the PO scale shows Chinese Taipei and Hong Kong SAR as the lowest; and Bahrain and Qatar as the highest. These findings show that relatively high or low scores for the scales can coexist; in other words, the teaching practices are not mutually exclusive. They further show that although there are substantial differences between countries, clusters of countries seem to be apparent. However, a lower or higher level of practice does not say much about the predictive value of these practices with regard to mathematics achievement. We therefore now turn to the multilevel models. Table 3 shows the results for the final model for all countries.<sup>7</sup> Although it is not the main focus of this paper, it firstly can be noted that in many jurisdictions there are significant differences between males and females, within most countries males outperforming females, but many Middle-Eastern countries showing the opposite result. SES is a positive predictor of Grade 8 mathematics achievement, in line with prior literature. GEN practices are only a significant predictor in 12 countries, with four negative and eight positive. In East Asian countries, only in Singapore generative practices are a significant negative predictor of mathematics achievement. Even fewer of the betas for TA and PO are significant, with most being negative. For our East Asian focus, only in Hong Kong SAR, the “Teacher-Active” scale is a significant negative predictor of mathematics achievement, while in Singapore the “Practice-on-Own” scale is a significant positive predictor of mathematics achievement. In almost all countries, the absence of limitations for teaching (BTBGLSN) is a positive predictor of mathematics achievement. Notably, though, this is not the case in Korea. Finally, instructional time (BTBM14) is a significant predictor of mathematics achievement in 19 out of 43 countries, six negatively and 13 positively.<sup>8</sup> Looking at East Asian jurisdictions, we see instructional time being a positive predictor of mathematics achievement in Chinese Taipei and Japan, a negative predictor of mathematics achievement in Singapore, and no significant predictor in Hong Kong SAR and Korea. These results do not show a consistent pattern of country clusters.

## 5. Discussion

Across all countries, the results show that the extent to which instructional variables are predictive of Grade 8 mathematics achievement, differs considerably, even within clusters of countries that typically are seen as “related.” Between East-Asian countries, there is quite some variance, with for example Singapore showing generative practices (GEN) as a significant negative predictor, and PO as a significant positive predictor. However, other jurisdictions such as Chinese Taipei, Japan, and Korea show no significant predictors (Table 3). East Asian jurisdictions *do* report relatively fewer GEN practices, than most other countries (Table 2). Two other instructional predictors, limitations for teaching and instructional time, also show no consistent pattern, both across all countries as East Asian countries in particular. For example, although in Chinese Taipei, Japan, and Singapore limitations for teaching and instructional time both are significant predictors, the latter is negative in Singapore and positive in the other two. In Hong Kong SAR, only the limitations for teaching are a significant positive predictor. In Korea, neither variables are predictive of Grade 8 mathematics achievement. So, although we posited two current views on “clusters of countries,” one view that says that there is some coherence and another view that within East Asia there are many differences, our results are more in line with the latter. To echo previous results by Cheng (2014), one approach cannot be said to be uniformly more predictive of positive outcomes than the other. Our findings contradict suggestions by Leung (2017) that East Asian countries form a cultural cluster. If the question is, as posited in Leung’s (2017) title “Does culture really matter?”, then we agree that the answer still is “probably,” with this study providing some further evidence toward this. However, although there might be common elements between certain cultures, they also are inherently unique. In addition,

**Table 3.** Model results.

	Intercept	ITSEX	BSBGHER	GEN	TA	PO	BTBGLSN	BTBML4
Australia	174.55 ***	10.28 ***	10.94 ***	8.62 ***	5.92	-6.16	18.77 ***	-0.06
Bahrain	366.77 ***	-12.54 ***	8.44 ***	11.82 ***	-6.54	-4.36	3.40 *	0.04
Chile	221.88 ***	10.41 ***	9.38 ***	-16.58 **	-9.68	8.60	13.50 ***	0.15 ***
Chinese Taipei	240.39 ***	-0.99 ***	19.18 ***	5.62 ***	3.74	-4.45	9.46 ***	0.25 ***
Cyprus	185.92 ***	-1.57 ***	21.51 ***	3.58 ***	-6.61	10.20	3.88 ***	0.14 ***
Finland	180.09 ***	1.92 ***	15.68 ***	9.48 ***	-7.71 *	6.16	10.58 ***	0.06 ***
France	211.12 ***	10.42 ***	17.95 ***	-5.03 ***	3.42	0.08	6.02 ***	0.05 ***
Georgia	308.05 ***	9.39 ***	13.74 ***	21.32 ***	-0.14	-25.31 ***	-3.73	0.20 *
Hong Kong SAR	227.33 ***	10.36 ***	1.96 ***	16.62 *	-12.80 *	7.33	27.89 ***	-0.03 ***
Hungary	107.15 ***	12.34 ***	20.25 ***	12.16 *	-5.16	-0.89	11.21 ***	0.13 ***
Islamic Republic of Iran	270.24 ***	-9.00 ***	10.78 ***	19.89 ***	-15.57 ***	1.67	5.49 ***	0.12 **
Israel	267.63 ***	11.75 ***	10.38 ***	16.30 *	-24.67 ***	-0.91	15.64 ***	0.04 ***
Italy	349.93 ***	14.27 ***	13.80 ***	-7.15 ***	-2.56	-4.98	2.19 *	0.04 ***
Japan	296.37 ***	2.28 ***	19.32 ***	4.38 ***	0.65	-2.07	4.18 ***	0.17 ***
Kazakhstan	360.77 ***	0.75 ***	7.73	5.68 ***	5.21	-12.10	2.06	0.19 ***
Jordan	256.09 ***	-13.24 *	7.83	0.51 ***	-3.16	9.40	10.83 ***	-0.05 ***
Republic of Korea	359.40 ***	3.17 ***	21.25 ***	4.44 ***	-4.41	0.64	-0.28	-0.02 ***
Kuwait	185.55 ***	6.07 ***	7.05	5.31 ***	-7.83	-2.13	13.66 ***	0.09 ***
Lebanon	308.88 ***	2.55 ***	5.37 ***	16.31 ***	-14.57 *	2.13	4.93 *	0.02
Lithuania	211.87 ***	6.38 ***	22.12 ***	6.79 ***	-6.07	6.55	3.93 ***	0.00 ***
Morocco	189.72 ***	5.67 ***	3.21 ***	5.29 ***	1.98	0.56	6.70 ***	0.31 ***
Oman	268.52 ***	-15.64 ***	14.12 ***	-1.92 ***	-8.45	3.43	7.88 ***	-0.06 ***
Norway	226.72 ***	4.70 ***	18.91 ***	0.90 ***	1.13	-1.07	5.41 ***	-0.01 ***
Portugal	183.88 ***	14.59 ***	14.00 ***	8.30 ***	7.95	-6.83	9.21 ***	0.15 ***
Qatar	354.13 ***	8.87 ***	13.11 ***	-11.51 ***	-27.16 ***	0.60	11.45 ***	-0.08 ***
Romania	189.03 ***	-8.79 ***	20.58 ***	8.34 ***	-11.78	1.06	9.78 ***	-0.02 ***
Russian Federation	360.75 ***	7.97 ***	9.73 ***	14.76 ***	-11.23	5.18	3.70 **	0.01
Saudi Arabia	265.79 ***	-15.54 ***	8.13 ***	10.82 ***	-2.33	2.69	3.50	0.02
Singapore	462.46 ***	6.80 ***	1.76 ***	-27.10 ***	-10.41 ***	19.11	20.76 ***	-0.23 ***
South Africa	273.34 ***	-0.84 ***	1.52 ***	-12.41 ***	-2.77 *	12.69 *	14.68 ***	-0.05 ***
Sweden	156.61 ***	1.37 ***	19.71 ***	3.01 ***	1.25	3.14	7.27 ***	0.17 *
United Arab Emirates	277.83 ***	5.25 ***	6.31 ***	28.08 ***	-3.99	-0.66	12.62 ***	-0.22 ***

(continued)

Table 3. (continued)

	Intercept	ITSEX	BSBGHER	GEN	TA	PO	BTBGLSN	BTBM14
Turkey	223.35 ***	-5.84 ***	18.10 ***	-8.12 ***	-5.27	9.85	11.03 ***	0.08
Egypt	326.34 ***	-14.99 ***	6.54 ***	10.44 ***	-15.30	-11.75	4.97	0.30 ***
United States	222.22 ***	7.28 ***	11.54 ***	1.30 ***	-11.33 *	15.01 *	9.13 ***	0.10 ***
England	122.27	12.42 ***	12.11 ***	41.70 ***	-3.86 **	1.69	14.44 ***	-0.17
United Arab Emirates (Dubai)	436.49 ***	5.25 ***	5.86 **	11.38 ***	-3.10	-12.30 *	8.01 ***	-0.16 ***
United Arab Emirates (Abu Dhabi)	198.51 ***	0.87 ***	6.61 ***	36.27 ***	7.94 ***	4.96	11.88 ***	-0.30 ***
Canada (Ontario)	277.68 ***	7.17 ***	16.54 ***	-21.12 ***	-1.55 ***	4.28	11.42 ***	0.04
Canada (Quebec)	220.56 ***	13.06 ***	10.02 ***	27.13 ***	12.26 *	-17.27 ***	11.20 ***	0.02
Russian Federation, Moscow	446.32 ***	18.97 ***	10.98 ***	-3.67 ***	-35.37 ***	41.79 ***	1.28 ***	-0.11 ***
South Africa (Gauteng)	301.46 ***	0.58 ***	2.05 ***	-16.38 ***	-5.05	11.66	15.25 ***	0.00
South Africa (Western Cape Province)	255.33 ***	12.70 ***	4.17 ***	-32.79 ***	-12.94 ***	12.96	32.33 ***	-0.19 ***

Note: Values are reported betas for the final multilevel models. East Asian jurisdictions in grey. Malaysia's models did not converge and therefore have been omitted. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . Dependent variable BSMAT01-05 = mathematics achievement; ITSEX = gender (female = 1, male = 2); BSBGHER = home educational resources scale; GEN = generative scale; TA = teacher-active scale; PO = practice-on-own scale; BTBGLSN = "Classroom Teaching Limited by Students Not Ready for Instruction" scale; BTBM14 = instructional time in minutes per week.

cultural influences are hard to determine in even just one country, let alone numerous cultures across the world. We therefore need to continue our research in effective mathematics education practices, but realize that there is not one “silver bullet” that works across all available contexts. Having said that, this also does not mean we cannot *learn* from other countries’ prevalent instructional practices, with some practices supported by sociological, pedagogical, and psychological research.

Regarding our analytical approach, we acknowledge that our findings will be influenced by the choice of variables. In this case, we used instructional variables available in the Grade 8 TIMSS 2019 dataset, but the way we combine them in scales is subject to analytical choices, which can have consequences on our conclusions. For example, many analysts have looked at inquiry-based learning in science based on PISA 2015 data. Two separate analyses of the data, using different compositions of the inquiry scales, yielded different results (Jerrim et al., 2020; Oliver et al., 2021). This article presents data-driven choices, based on factor analysis, with regard to scales we dubbed “generative,” “teacher-active,” and “practice-on-own.” Although we describe our analytical choices as transparently as possible, they still are subject to some arbitrariness. In fact, analyses of secondary data like these, inherently include many analytical choices, which in turn lead to analytical variability (Bokhove, 2022). In other words, different variable choices might yield different results. To mitigate issues with analytical variability, in future research secondary data analyses could be preregistered, with analysis scripts deposited in, for example, the Open Science Framework,<sup>9</sup> or even a registered report.

A further limitation of our analysis is that secondary data analyses with TIMSS data cannot determine causality. For this, we would need longitudinal data, and as TIMSS only provides cross-sectional data, this prevents strong causal assertions, unless we try to cater for this in our analyses. Using older TIMSS data, Gustafsson (2013) asserted that causal relations can be explored if threats to valid causal inference from observational data are addressed. In relation to the focus of teaching practices, it could be the case that teachers simply adopt different practices for students with different prior proficiency. For example, in PISA 2015 data, a negative association between inquiry-based instruction and science achievement was found (OECD, 2016) which led to a conclusion that some of the arguments against hands-on activities should not be disregarded. However, the findings also noted that teachers may be using hands-on activities to make science more attractive to disengaged students. This point regarding causality is part of wider criticisms some scholars have addressed at ILSA providers (e.g., Zhao, 2017, 2020) and the challenges to disentangle all the country, school, and other contextual variables that make up an education system.

Although this study’s use of teacher data was partly predicated on students not always knowing best with regard to their learning (e.g., see Kirschner & van Merriënboer, 2013), it also isn’t necessarily the case that teacher’s perceptions of their teaching approaches will match students’ experiences (Beusaert et al., 2013). Referring to other studies (e.g., Nijhuis et al., 2005), they conclude that it is important and relevant to measure students’ perceptions of the teachers’ approaches to teaching instead of using teachers’ self-reports and call for more research on students’ perceptions. This would present ILSA analysts of secondary education mathematics with a quandary as the two main studies TIMSS and PISA, each have their own emphases. That challenge extends to other contextual variables as well. For example, this analysis did not include all possible confounding variables. Although we, for example, included a proxy for SES and instructional time, there are numerous other relevant variables, including demographic factors, characteristics of school systems, time in school, time spent teaching particular topics, emphasis on testing, societal and family cultures, level of education for teachers, teacher professional development, inequities within a country, variation of funding/access to schooling within a country, affective measures such as confidence and many more. Furthermore, this study focused on mathematics achievement as an outcome, but there are many more noncognitive outcomes of mathematics education as well.

Notwithstanding this, using current data from TIMSS 2019 does provide some more insight into the association between instructional practices and mathematics achievement across multiple countries. Such large-scale international comparisons would be hard to conduct in any other way. Although we capture between-teacher variance in our models, we do not account for school-level variables, as the TIMSS dataset samples one or two classrooms per school, and hence might not be representative of instructional practices of the whole school. One challenge in comparing results from different countries lies in the extent different curricula more or less incentivize particular practices. Of course, individual teachers play an important role in this, but it could be argued that some curricula perhaps enable more opportunities for learning and teaching “the right content” for TIMSS outcomes. In that sense, TIMSS’ curriculum focus could obscure certain effects. To check such a suspicion, we turned to the TIMSS 2019 Test-Curriculum Matching Analysis (TCMA; Mullis et al., 2020). The TCMA analyses the extent to which the TIMSS 2019 mathematics assessment items are relevant for each country’s curriculum, calculating comparative outcome scores when only achievement items that are judged to be relevant to its own curriculum, are included. Although these analyses show *some* more differences between countries, none of them are notable with regard to this study’s findings, hinting like Jerrim (2015) that East Asia’s high levels of achievement also arise because of culture, not just curriculum. One such cultural aspect might be the use of out-of-school tutoring. This paper focuses on instructional practices in the classroom, but one aspect that Leung et al. (2015) and others have pointed out is the role of tutoring, for example, in Juku schools in Japan, and Hagwons in Korea.

Finally, although this study focused on the overall Grade 8 TIMSS 2019 mathematics achievement, it could be argued that traditional ILSA assessment items are less suitable for evaluating conceptual knowledge. It therefore was interesting that in the TIMSS 2019 cycle, a pilot study with problem-solving and inquiry tasks was conducted (Mullis et al., 2021). Although limited in scope, some countries’ achievement scores differed slightly upon including the items in the overall score (Mullis et al., 2021). The patterns of differences, however, again were not consistent, with for example Chinese Taipei and Korea scoring lower, but Singapore and Hong Kong SAR at a similar level as the general achievement score. The limited scope of the pilot study does not allow for firm conclusions on this, but could indicate that the way we assess mathematics in ILSA influences outcomes as well. Further research could look into the potential of further diversifying assessment items in large-scale studies, with an eye for implementation and feasibility. For example, recent results with randomized controlled trials have shown that “comparative judgement” could be a valid, reliable, and practical tool to measure conceptual understanding (Jones et al., 2019).

In conclusion, this article again confirms that diversity in instructional practices in international mathematics classrooms is not easy to capture with teacher-reported questionnaire data, but that if we would conclude one thing, it is that instructional practices differ considerably, across the world, and even within regions such as East Asia. In reporting on country differences, we need to be aware of this diversity.

### **Acknowledgments**

The author thanks the reviewers for their constructive comments.

### **Declaration of Conflicting Interests**


The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author received no financial support for the research, authorship, and/or publication of this article.



## ORCID iD

Christian Bokhove  <https://orcid.org/0000-0002-4860-8723>

## Notes

1. See <https://www.iea.nl/about>.
2. See [https://timssandpirls.bc.edu/timss2019/methods/pdf/T19\\_MP\\_Ch3-sample-design.pdf](https://timssandpirls.bc.edu/timss2019/methods/pdf/T19_MP_Ch3-sample-design.pdf) for more details on the sampling design.
3. The use of MATWGT at level 2 corresponds with the recommendations by Mang et al. (2021).
4. The context questionnaires can be found at <https://timssandpirls.bc.edu/timss2019/questionnaires/index.html>.
5. Note that from these questionnaire items, BTBM15F was not easy to fit in that description, as the text seemed quite ambiguous. Does “Work problems together in the whole class with direct guidance from me” refer to multiple students actively working on problems together, or are they working on their own, together in one classroom, under the guidance of the teacher? This might also explain the only just acceptable, but lower alpha value.
6. To not have to communicate the rather lengthy scale name, we refer to this scale with variations of limitations/barriers to teaching/instruction.
7. Note that the model results for Ireland, New Zealand, and Malaysia did not converge and therefore are not in the table. Only descriptive statistics for Malaysia were included in Table 2 because of the East Asian focus of the paper.
8. The betas are small here because the instructional time was measured in minutes, so the betas represent the change in mathematics achievement with an increase of one minute.
9. See <https://osf.io/>.

## References

- Beattie, V., IV, Collins, B., & McInnes, B. (1997). Deep and surface learning: A simple or simplistic dichotomy? *Accounting Education*, 6(1), 1–12. <https://doi.org/10.1080/096392897331587>
- Beusaert, S. A. J., Segers, M. S. R., & Wiltink, D. P. (2013). The influence of teachers’ teaching approaches on students’ learning approaches: The student perspective. *Educational Research*, 55(1), 1–15. <https://doi.org/10.1080/00131881.2013.767022>
- Biggs, J. B. (1994). What are effective schools? Lessons from East and West. *The Australian Educational Researcher*, 21(1), 19–39. <https://doi.org/10.1007/BF03219558>
- Biggs, J. (1998). Learning from the Confucian heritage: So size doesn’t matter? *International Journal of Educational Research*, 29(8), 723–738. [https://doi.org/10.1016/S0883-0355\(98\)00060-3](https://doi.org/10.1016/S0883-0355(98)00060-3)
- Bokhove, C. (2022). The role of analytical variability in secondary data replications: A replication of Kim et al. (2014). *Educational Research and Evaluation*, 27(1–2), 141–163. <https://doi.org/10.1080/13803611.2021.2022319>
- Bokhove, C., Miyazaki, M., Komatsu, K., Chino, K., Leung, A., & Mok, I. A. C. (2019). The role of “opportunity to learn” in the geometry curriculum: A multilevel comparison of six countries. *Frontiers in Education*, 4, 63. <https://doi.org/10.3389/educ.2019.00063>
- Boylan, M., Maxwell, B., Wolstenholme, C., Jay, T., & Demack, S. (2018). The mathematics teacher exchange and ‘mastery’ in England: The evidence for the efficacy of component practices. *Education Sciences*, 8(4), 202. <https://doi.org/10.3390/educsci8040202>
- Brod, G. (2021). Generative learning: Which strategies for what age? *Educational Psychology Review*, 33, 1295–1318. <https://doi.org/10.1007/s10648-020-09571-9>
- Brownell, W. A. (1945). When is arithmetic meaningful? *Journal of Educational Research*, 38(7), 481–498. <https://doi.org/10.1080/00220671.1945.10881369>
- Bryan, C. A., Wang, T., Perry, B., Wong, N. Y., & Cai, J. (2007). Comparison and contrast: Similarities and differences of teachers’ views of effective mathematics teaching and learning from four regions. *ZDM—The International Journal on Mathematics Education*, 39(4), 329–340. <https://doi.org/10.1007/s11858-007-0035-2>

- Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation*, 49, 30–41. <https://doi.org/10.1016/j.stueduc.2016.03.005>
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8), 723–733. <https://doi.org/10.1177%2F016146816306400801>
- Chen, X., & Li, Y. (2010). Instructional coherence in Chinese mathematics classroom—A case study of lessons on fraction division. *International Journal of Science and Mathematics Education*, 8(4), 711–735. <https://doi.org/10.1007/s10763-009-9182-y>
- Cheng, Q. (2014). Quality mathematics instructional practices contributing to student achievements in five high-achieving Asian education systems: An analysis using TIMSS 2011 data. *Frontiers of Education in China*, 9(4), 493–518. <https://doi.org/10.1007/BF03397038>
- Coughlan, S. (2019, December 3). *Pisa tests: UK rises in international school rankings*. BBC News. <https://www.bbc.co.uk/news/education-50563833>
- Dahlin, B., & Watkins, D. (2000). The role of repetition in the processes of memorizing and understanding: A comparison of the views of German and Chinese secondary school students in Hong Kong. *British Journal of Educational Psychology*, 70(1), 65–84. <https://doi.org/10.1348/000709900157976>
- Echazarra, A., Salinas, D., Méndez, I., Denis, V., & Rech, G. (2016). How teachers teach and students learn: Successful strategies for school. *OECD Education Working Papers*, No. 130. <https://doi.org/10.1787/5jm29kpt0xxx-en>
- Fan, L., & Bokhove, C. (2014). Rethinking the role of algorithms in school mathematics: A conceptual model with focus on cognitive development. *ZDM – Mathematics Education*, 46(3), 481–492. <https://doi.org/10.1007/s11858-014-0590-2>
- Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 user guide for the international database*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/international-database>
- Guo, K. (2014, May 21–23). *Relation between time for studying mathematics and achievement* [Paper presentation]. The First Chinese Congress on Mathematics Education, Beijing, China.
- Gustafsson, J. E. (2013). Causal inference in educational effectiveness research: A comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement*, 24(3), 275–295. <https://doi.org/10.1080/09243453.2013.806334>
- Hess, R. D., & Azuma, M. (1991). Cultural support for schooling: Contrasts between Japan and the United States. *Educational Researcher*, 20(9), 2–8. <https://doi.org/10.3102%2F0013189X020009002>
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., Hollingsworth, H., Wearne, D., & Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, 27(2), 111–132. <https://doi.org/10.3102%2F01623737027002111>
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jerrim, J. (2015). Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent. *Oxford Review of Education*, 41(3), 310–333. <https://doi.org/10.1080/03054985.2015.1028525>
- Jerrim, J., Oliver, M., & Sims, S. (2020). The relationship between inquiry-based teaching and students' achievement. New evidence from a longitudinal PISA study in England. *Learning and Instruction*, 80, 101310. <https://doi.org/10.1016/j.learninstruc.2020.101310>
- Jones, I., Bisson, M., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal*, 45(3), 662–680. <https://doi.org/10.1002/berj.3519>

- Kaiser, G., & Blömeke, S. (2013). Learning from the Eastern and the Western debate: The case of mathematics teacher education. *ZDM—The International Journal on Mathematics Education*, 45(1), 7–19. [https://doi.org/10.1007/978-94-007-6437-8\\_24](https://doi.org/10.1007/978-94-007-6437-8_24)
- Kilpatrick, J. E., Swafford, J. E., & Findell, B. E. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press. <https://doi.org/10.17226/9822>
- Kirschner, P. A., & van Merriënboer, J. J. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, 48(3), 169–183. <https://doi.org/10.1080/00461520.2013.804395>
- Leung, F. K. S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics*, 47(1), 35–51. <https://doi.org/10.1023/A:1017936429620>
- Leung, F. K. S. (2006). Mathematics education in East Asia and the West: Does culture matter? In F. K. S. Leung, K.-D. Graf, & F. J. Lopez-Real (Eds.), *Mathematics education in different cultural traditions—A comparative study of East Asia and the West* (pp. 21–46). New York, NY: Springer. [https://doi.org/10.1007/0-387-29723-5\\_2](https://doi.org/10.1007/0-387-29723-5_2)
- Leung, F. K. S. (2017). Making sense of mathematics achievement in East Asia: Does culture really matter? In G. Kaiser (Ed.), *Proceedings of the 13th international congress on mathematical education* (pp. 201–218). Cham, Switzerland: Springer. [https://doi.org/10.1007/978-3-319-62597-3\\_13](https://doi.org/10.1007/978-3-319-62597-3_13)
- Leung, F. K. S., & Park, K. M. (2002). Competent students, competent teachers? *International Journal of Educational Research*, 37(2), 113–129. [https://doi.org/10.1016/S0883-0355\(02\)00055-1](https://doi.org/10.1016/S0883-0355(02)00055-1)
- Leung, F. K., Park, K., Shimizu, Y., & Xu, B. (2015). Mathematics education in East Asia. In S. J. Cho (Ed.), *The proceedings of the 12th international congress on mathematical education* (pp. 123–143). Cham, Switzerland: Springer. [https://doi.org/10.1007/978-3-319-12688-3\\_11](https://doi.org/10.1007/978-3-319-12688-3_11)
- Li, S. (1999). Does practice make perfect? *For the Learning of Mathematics*, 19(3), 33–35. <https://www.jstor.org/stable/40248308>
- Lie, S. (2014). *Similarities and differences between groups of countries concerning relative weaknesses and strengths*. IEA. [https://www.iea.nl/sites/default/files/2019-04/GA55\\_Weaknesses\\_strengths\\_groups\\_Lie.pdf](https://www.iea.nl/sites/default/files/2019-04/GA55_Weaknesses_strengths_groups_Lie.pdf)
- Lim, C. S. (2007). Characteristics of mathematics teaching in Shanghai, China: Through the lens of a Malaysian. *Mathematics Education Research Journal*, 19(1), 77–88. <https://doi.org/10.1007/BF03217450>
- Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mang, J., Küchenhoff, H., Meinck, S., & Prenzel, M. (2021). Sampling weights in multilevel modelling: An investigation using PISA sampling structures. *Large-Scale Assessments in Education*, 9(1), 1–39. <https://doi.org/10.1186/s40536-021-00099-0>
- Martin, M. O., Von Davier, M., & Mullis, I. V. (Eds.) (2020). *Methods and procedures: TIMSS 2019 technical report*. <https://timssandpirls.bc.edu/timss2019/methods/pdf/TIMSS-2019-MP-Technical-Report.pdf>
- Marton, F., Dall’Alba, G., & Tse, K.-T. (1996). Memorising and understanding: the keys to the paradox? In D. Watkins, & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological, and contextual influences* (pp. 69–83). Hong Kong, China: Comparative Education Research Centre; and Australian Council for Educational Research. <https://bibliography.lib.eduhk.hk/bibs/502412e1>
- Mullis, I. V., Martin, M. O., Fishbein, B., Foy, P., & Moncaleano, S. (2021). *Findings from the TIMSS 2019 problem solving and inquiry tasks*. TIMSS 2019. <https://timss2019.org/psi/>
- Mullis, I. V., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center website. <https://timssandpirls.bc.edu/timss2019/international-results>
- Nijhuis, J. F., Segers, M. S., & Gijsselaers, W. H. (2005). Influence of redesigning a learning environment on student perceptions and learning strategies. *Learning Environments Research*, 8(1), 67–93. <https://doi.org/10.1007/s10984-005-7950-3>
- OECD (2016). *PISA 2015 results: Policies and practices for successful schools* (Vol. II). Paris, France: OECD Publishing. <https://www.oecd.org/publications/pisa-2015-results-volume-ii-9789264267510-en.htm>
- OECD (2019). *PISA 2018 results: What students know and can do* (Vol. I). Paris, France: OECD Publishing. <https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>

- Ofsted (2021). *Research review series: mathematics*. Retrieved July 7, 2022, from <https://www.gov.uk/government/publications/research-review-series-mathematics/research-review-series-mathematics>
- Oliver, M., McConney, A., & Woods-McConney, A. (2021). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education, 51*(2), 595–616. <https://doi.org/10.1007/s11165-019-09901-0>
- Pampaka, M., & Williams, J. (2016). Mathematics teachers' and students' perceptions of transmissionist teaching and its association with students' dispositions. *Teaching Mathematics and its Applications: An International Journal of the IMA, 35*(3), 118–130. <https://doi.org/10.1093/teamat/hrw007>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- R Core Team (2021). *R: A language and environment for statistical computing* (Version 4.1.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2022). *psych: Procedures for Personality and Psychological Research* (Version 2.2.5) [Computer software]. Northwestern University. <https://CRAN.R-project.org/package=psychVersion=2.2.5>
- Rindermann, H., & Baumeister, A. E. (2015). Validating the interpretations of PISA and TIMSS tasks: A rating study. *International Journal of Testing, 15*(1), 1–22. <https://doi.org/10.1080/15305058.2014.966911>
- Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM – Mathematics Education, 49*(4), 599–611. <https://doi.org/10.1007/s11858-017-0834-z>
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587–597. <https://doi.org/10.1007/s10648-015-9302-x>
- Robitzsch, A., & Oberwimmer, K. (2022). *BIFIEsurvey: Tools for survey statistics in educational assessment* (R package version 3.4-15) [Computer software]. <https://CRAN.R-project.org/package=BIFIEsurvey>
- RStudio Team (2021). *RStudio: Integrated development for R* (Version 1.4.1717) [Computer software]. RStudio, PBC. <http://www.rstudio.com/>
- Rubin, D. (1987). *Multiple imputation for nonresponse in sample surveys*. New York, NY: John Wiley & Sons. <https://doi.org/10.1002/9780470316696>
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151. <https://doi.org/10.3102%2F0013189X10363170>
- Schoenfeld, A. H. (2004). The math wars. *Educational Policy, 18*(1), 253–286. <https://doi.org/10.1177%2F0895904803260042>
- Shimizu, Y., & Williams, G. (2012). Studying learners in intercultural contexts. In K. Clements, A. Bishop, J. Kilpatrick, F. Leung, & C. Keitel (Eds.), *Third international handbook in mathematics education* (pp. 145–167). New York, NY: Springer. [https://doi.org/10.1007/978-1-4614-4684-2\\_5](https://doi.org/10.1007/978-1-4614-4684-2_5)
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. <https://doi.org/10.3102/00346543075003417>
- Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20–26. [https://www.atm.org.uk/write/MediaUploads/Journals/MT077/Relational\\_Understanding\\_and\\_Instrumental\\_Understanding\\_%E2%80%93\\_Richard\\_R.\\_Skemp.pdf](https://www.atm.org.uk/write/MediaUploads/Journals/MT077/Relational_Understanding_and_Instrumental_Understanding_%E2%80%93_Richard_R._Skemp.pdf)
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.
- Stapleton, L. M., McNeish, D. M., & Yang, J. S. (2016). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist, 51*(3–4), 317–330. <https://doi.org/10.1080/00461520.2016.1207178>
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education, 36*(5), 404–411. <https://doi.org/10.2307/30034943>

- Wu, M. (2009). A comparison of PISA and TIMSS 2003 achievement results in mathematics. *Prospects*, 39(1), 33–46. <https://doi.org/10.1007/s11125-009-9109-y>
- Wu, Y. J., Carstensen, C. H., & Lee, J. (2020). A new perspective on memorization practices among East Asian students based on PISA 2012. *Educational Psychology*, 40(5), 643–662. <https://doi.org/10.1080/01443410.2019.1648766>
- Zhao, Y. (2017). What works may hurt: Side effects in education. *Journal of Educational Change*, 18(1), 1–19. <https://doi.org/10.1007/s10833-016-9294-4>
- Zhao, Y. (2020). Two decades of havoc: A synthesis of criticism against PISA. *Journal of Educational Change*, 21(2), 245–266. <https://doi.org/10.1007/s10833-019-09367-x>

### Author biography

**Christian Bokhove** is Associate Professor in Mathematics Education at the Southampton Education School, University of Southampton, UK. He specializes in secondary mathematics education, with an emphasis on international comparisons, as well as using innovative research methods to answer relevant educational research questions.