AIMS Mathematics

AIMS Mathematics, 9(10): 28671–28697.
DOI: 10.3934/math.20241391
Received: 29 June 2024
Revised: 14 August 2024
Accepted: 06 September 2024
Published: 09 October 2024

http://www.aimspress.com/journal/Math

Research article

Dialogic teaching practices and student performance in mathematics in Asian countries

Volker Schöer[1,2] and Jose G. Clavel[3,*]

[1] School of Economics and Finance, University of the Witwatersrand, Johannesburg, South Africa
[2] School of Business and Economics, University of Tübingen, Germany
[3] Department of Quantitative Methods, Universidad de Murcia, Spain

* **Correspondence:** Email: jjgarvel@um.es; Tel: +34868883757; Fax: +34868883789.

Abstract: Improving the mathematics performance of school children is an objective for many policy-makers around the world. Student-centered interactive pedagogies like classroom discussions and other dialogic interaction practices have been considered the best practice to engage learners effectively in the learning process. However, dialogic teaching practices are least used in most Asian countries that on average achieve the highest mathematics scores in international assessments. Based on this conundrum, this paper utilized a large-scale education dataset for five Asian countries, namely the Trends in International Mathematics and Science Study (TIMSS) dataset for learner data from Chinese Taipei, Hong Kong, Korea, Malaysia, and Singapore to examine the relationship between dialogic classroom interaction teaching practices and the mathematics performance of 8th-grade learners. Using a within-learner-between-subject estimation strategy to account for endogeneity, we established that learners who are taught more frequently through dialogic interactive teaching practices in mathematics classes achieve higher mathematics scores. Our results confirm that interactive pedagogies do provide learning benefits, even in countries that use them sparingly. Thus, our findings challenge the held assumption that the efficacy of dialogic classroom interaction practices is context and learning culture specific. Nevertheless, our study also shows that the highly endogenous nature of the teaching and learning environment and learner performance limits the ability of any study that uses observational data to establish the true impact of dialogic practices on learner mathematics performance.

Keywords: TIMSS; plausible values; teaching mathematics strategies; student achievement; dialogic classroom interaction pedagogies, Asian teaching and learning culture, Chinese learner paradox

Mathematics Subject Classification: 97A40, 97C50, 97D10, 97D40

1. Introduction

Dialogic classroom interaction practices are considered a crucial component of quality teaching in standard teaching and learning models [18,29,34] including the teaching of mathematics [2,35]. These dialogic interaction practices include asking individual learners to report back to the class, as well as classroom discussions. However, only sparce, rigorous, empirical evidence exists to support the effectiveness of these practices in improving learning outcomes in mathematics. More importantly, the lack of frequent use of dialogic teaching practices in countries that perform well in international benchmark tests suggests that this relationship is context and learning culture specific. Specifically, various studies indicate that dialogic classroom practices do not work in Asian countries, confirming the so-called “Chinese learner paradox” [20] in other countries of that part of the world.

Our study aims to contribute to the debate on the effectiveness of dialogic classroom interaction practices by specifically investigating the impact of various dialogic classroom interaction practices and mathematics performances of eight -grade learners from five Asian countries.

1.1. Teaching and learning models and the importance of dialogic interactions

Various scholars have developed comprehensive models of teaching and learning. These models typically identify three key dimensions that facilitate an effective teaching and learning environment and therefore determine subsequent learning outcomes: 1) classroom management that establishes the learning environment in the classroom; 2) teacher expertise; and 3) cognitive activation of learners through teaching practices [4,5,10,25]. The classroom provides the space of these interactions. As Praetorius et al. [25] summarized:

“The classroom offers a socio cognitive as well as a material space where teachers and students engage in various forms of interactions and activities. These interactions and activities are deeply rooted in social practice that is co-constructed by students and teachers, using traditional patterns (such as social-spatial arrangements, teaching methods, ways to involve media, rules of collaboration, and rules of turn-taking) that have a strong cultural background. The teacher, as an expert in (general and subject-matter) didactics, chooses content elements, shapes learning tasks, and orchestrates the teaching–learning process. He or she has to find a balance between various educational goals and safeguard this balance on a moment-to-moment basis within the ongoing, contingent flow of classroom interactions” [25, pages 408–9].

According to these models, quality teaching of any subject implies that the teacher creates a classroom culture which fosters learning using teaching practices that encourage learners to engage with the material critically [5].

Part of the learning process includes that learners master the language of the subject through dialogic practices. Following this line of argumentation, [4] suggested that teaching practices should “build on students’ thinking, facilitating mathematical communication, encouraging the development of mathematical language, employing worthwhile mathematical tasks, making appropriate mathematical connections, assessing formatively and selecting appropriate tools and representations” [4]. Facilitating mathematical communication and the development of mathematical language allow learners to construct meaning for themselves [17,30]. Thus, implementing dialogic classroom practices provides learning spaces that "promote dialogic, engaged and 'safe' classroom environments where students are actively involved and feel free to contribute and take risks" [19].

Classroom management combined with cognitive activation have been shown to improve mathematics achievements [13]. Especially, frequent classroom discussions improve the learners' ability to engage with the mathematics material by allowing them to formulate their ideas, problems, and solutions as well as understanding the approaches of their peers.

However, several studies have provided evidence counter to the expected positive correlation between classroom dialogic interaction practices and improved mathematics achievement such as the Spanish case in [10]. More importantly, most of these studies are based on data from Asian countries and the findings are that specific teaching practices do not have any significant direct impact on student's mathematics achievement scores [28,38,39]. These studies suggest that interactive teaching methods are generally ineffective in Asian countries due to cultural differences (see, for instance, [37]).

This could be another form of the so-called "Chinese learner paradox", coined by Watkins and Biggs in 1996 [36]. Perhaps, one explanation for this observation could be that "putatively discredited practices" [20] actually work, i.e., that more authoritarian teaching and passive learning environments can produce superior learning outcomes. However, the fact that teachers in Asian countries do report the use of student-centered teaching strategies like dialogic practices, even if only sparingly, implies that some teachers do see value in the implementation of such pedagogies (see [6,11]).

This begs the question if the low uptake of interactive teaching practices across schools implies that using dialogic practices by Asian teachers do not have positive effects on learner outcomes. What is the actual impact of dialogic pedagogies on Asian students?

Our paper would like to answer these questions. Additionally, we contribute to the clarity of this debate by using the within-learner-between-subjects estimator to investigate the effect of dialogic practices in Asian countries. This estimation strategy avoids the limitations of most studies that do not sufficiently control for endogeneity of the teacher's choice of teaching practice in particular teaching and learning environments. It is plausible that the chosen classroom teaching practice is a function of teacher, student, classroom, school, and country characteristics that either facilitate or hinder the use of a particular teaching practice. The within-learner-between-subject estimator controls for most of the confounding variables and therefore allows us to provide a causal interpretation of the findings.

Our study contributes to the literature in three ways. First, using a quantitative cross-country analysis by comparing five different Asian countries, we test if dialogic classroom interaction practices affect mathematics achievement irrespective of cultural background. Second, we investigate separate individual dialogic practices—encouraging individual learners to express their ideas, asking learners to explain their answers to the rest of the class, as well as classroom discussions—and their impact on mathematics achievement. Finally, we control for potential endogeneity of the dialogic practice in the classroom environment. As such, we contribute to the debate around the Asian learner paradox literature through a rigorous analysis of a large and rich, internationally comparable dataset.

2. Materials and methods

2.1. Data and variables

We use the learner and teacher data from the "Trends in International Mathematics and Science

Study" (TIMSS) 2019, which represents the seventh assessment cycle conducted by the International Association for the Evaluation of Educational Achievement (IEA). This cycle includes assessments of learners in grade 4 and grade 8 classes across 64 countries and 8 benchmarking systems. For the purposes of our study, we focus on data from eighth-grade students in five Asian countries.

TIMSS 2019 for the 8th-grade classes provides us with an opportunity to investigate the classroom teaching and learning environment and the teaching activities of the mathematics and science teachers in more detail. These questionnaires were designed and validated by the International Association for the Evaluation of Educational Achievement (IEA)[1]. In grade 8, each mathematics and science teacher were asked individually to complete at least one questionnaire with questions regarding their teaching strategies and teaching context.

The teacher data contains one record for each of the classes taught either by a mathematics or a science teacher. In some schools, a student may have several teachers for the same subjects. Due to our identification strategy, we reduce our data to schools where students have only one mathematics teacher and one science teacher. This is the case for 88% of the sample. In Tables A1–A4 in the appendix, we present some descriptive statistics for the two groups of schools: schools with one subject teacher (included in our sample) versus schools with more than one subject teacher (excluded from our sample). Overall, the class environment of the one teacher group is more conducive to the implementation of dialogic practices. They have more enthusiastic teachers who spend more minutes per week on mathematics, while the students respect their teachers more, are more interested in the subject, and are less disruptive in class. It is therefore surprising that students with more than one teacher per subject perform on average better.

2.1.1. Measuring learner mathematics performance: The plausible values

TIMSS data has a complex structure that requires specific calculations to obtain reliable and representative results. Similar to other large-scale student assessments like PISA and PIRLS, the objective of measuring multiple subject domains results in hundreds of test items. Each student is administered only a fraction of these items during testing, following a balanced incomplete block design (BIBD). Item response theory (IRT), a method particularly well-suited to handle such data collection designs whereby not all students are tested on all items, provides plausible values that measure students' performance (see [33]).

In addition to the measurement errors arising from the BIBD and the IRT, there are potential sampling errors related to the sampling process. TIMSS uses the probability proportional to size sampling mechanism, which can impact the standard error of the estimations of the achievement variables. For example, the average performance for the studied countries is presented in Table 1. As shown, with the exception of Malaysia, the Asian countries in our sample outperform Anglophone countries in mathematics and science scores.

---

1 More information can be found in Chapter 3 of [21].The IEA Code of Conduct including their ethical principles is available at https://www.iea.nl.

Table 1. Average learner scores in mathematics and science.

Country	Mathematics score	Science score	Full sample
Singapore	615.77	607.55	4853
Chinese Taipei	612.24	574.26	4707
Korea	603.54	559.21	2265
Hong Kong	575.49	500.79	3153
Cyprus	558.05	533.34	458
Quebec, CA	543.61	536.97	3129
Ontario, CA	529.75	521.59	3776
Ireland	526.50	527.47	3853
Australia	517.70	528.69	8587
England	515.14	516.78	3325
United States	515.09	522.01	8655
New Zealand	481.20	499.47	5722
Malaysia	460.57	460.24	7065
South Africa	388.27	368.38	20,306
Avg/Total	531.64	518.34	79,854

2.1.2. Measuring frequent use of dialogical classroom interaction practices: Teachers' questionnaires

For grade 8, there are two teachers' questionnaires: one for the mathematics teacher and one for the science teacher. Besides collecting information on the individual characteristics of the teachers, it also asks about the instructional activities and strategies that teachers implement in the classroom. Specifically, it asks the teacher to report on the frequency of using the following practices in the classroom to encourage learners to engage with mathematics verbally:

Teacher questionnaire question 12: "How often do you do the following in teaching this class?": (1) Asking student to explain their answers. (2) Encourage students to express their ideas in class. (3) Encourage classroom discussions among students. The answer options are: Never (coded "1"), Some lessons (coded "2"), About half the lessons (coded "3"), and Every or almost every lesson (coded "4").

2.1.3. Measuring additional control variables to account for the three-dimensional teaching model: Students, schools, and teachers' questionnaires

The TIMSS data is a particularly rich dataset and can be summarized in three clusters: The student's context questionnaire includes questions on the learners' demographics (including age and gender), the learners' perceptions of the learning environment in the school and classroom, their educational resources at home, and their preferences for mathematics and science. In the school questionnaire, principals answered questions about the overall learning environment including demographic characteristics of the school's student body, the availability of instructional resources, and the general conduciveness of the environment for teaching and learning in their schools. Finally, the teachers' questionnaires collect information on the teacher demographic characteristics including age and gender as well as their level of education, professional development, and experience in teaching. Teachers were also asked about their perceptions of the classroom environment including the level of disruptive behavior by students, and the students' interest in the subject.

All in all, the information provided by the various TIMSS sections can be easily assigned to each of the three dimensions we are proposing in our framework. Classroom management establishes the learning environment in the classroom and the school. We capture if the classroom and the school are characterized by the level of learners behaving orderly or if classes are frequently disrupted. Additionally, we include the size of the class and the number of learners that struggle with the language to account for mediating effects that could explain the use of dialogical classroom practices. The expertise of the teachers is measured by their age, degree, years of teaching; and finally, the cognitive activation of learners through the teaching practices, our treatment variable in this paper as described above.

We restrict the sample to learner observations for which we have a full set of covariates (complete cases). The full sample of learners in the combined dataset contains 19,845 observations comprised of 4195 learners from Chinese Taipei, 2361 learners from Hong Kong, 2173 learners from Korea, 6340 learners from Malaysia, and 4806 learners from Singapore. We report in Table A5 (Appendix) the differences in characteristics between the included complete learner cases and the excluded incomplete learner cases. In terms of teacher and school characteristics, complete and incomplete learner cases differ at a statistically significant level. Specifically, complete cases seem to show a more conducive teaching and learning environment for better learning outcomes. Interestingly, classroom characteristics do not differ as much. Additionally, the two samples also differ in the frequency of dialogic practices with complete cases reporting higher frequencies of dialogic practice use. The Little's test confirms that missing values are not missing completely at random at a 1% level of significance. Thus, continuing with the delimited sample is likely to introduce sample selection bias. However, as incorrect model specification during imputation could similarly introduce bias given the highly endogenous and complex nature of the data structure, we refrain from imputing missing data and continue with the delimited dataset.

2.2. Descriptive statistics

For comparison purposes, we include in the first graph various Anglophone countries for which we have relevant dialogical teaching practice information. Confirming previous studies, Figure 1 shows that teachers in Asian countries report significantly lower frequencies of using classroom discussions in mathematics lessons. Interestingly, the ranking of countries by the frequency of using classroom discussion almost mirrors the ranking of the same countries' average mathematics performances as shown in Table 1. This pattern seems to be aligned with the claims that dialogic pedagogies do not necessarily produce a better understanding of mathematics. However, observing this pattern also does not disprove the possibility that dialogic pedagogies could still be useful for improving learner performances, even in countries that are already performing well.

The patterns of using dialogical teaching practices in mathematics classes and science classes show some variation across the Asian countries. Teachers in Malaysia report a significantly higher frequency of using classroom discussions in mathematics and science classes, while teachers in Korea, Hong Kong, and Singapore are less likely to use classroom discussions in mathematics classes. Science teachers in Singapore, on the other hand, report a higher frequency of using classroom discussions compared to their mathematics colleagues.

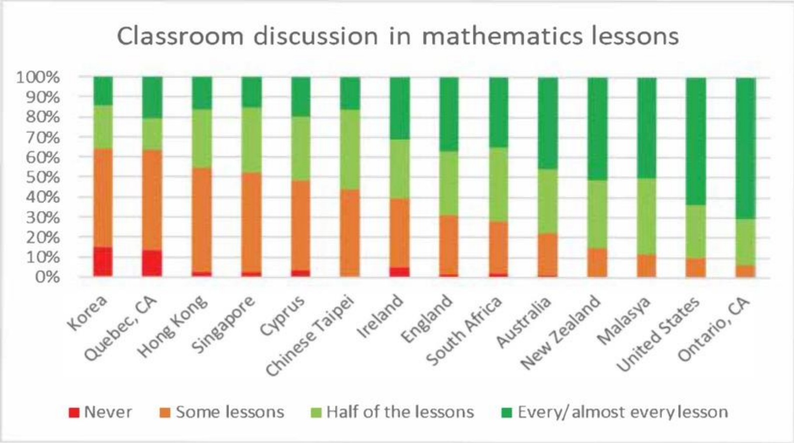

Figure 1. Frequency of using classroom discussions in mathematics lessons (multiple countries).

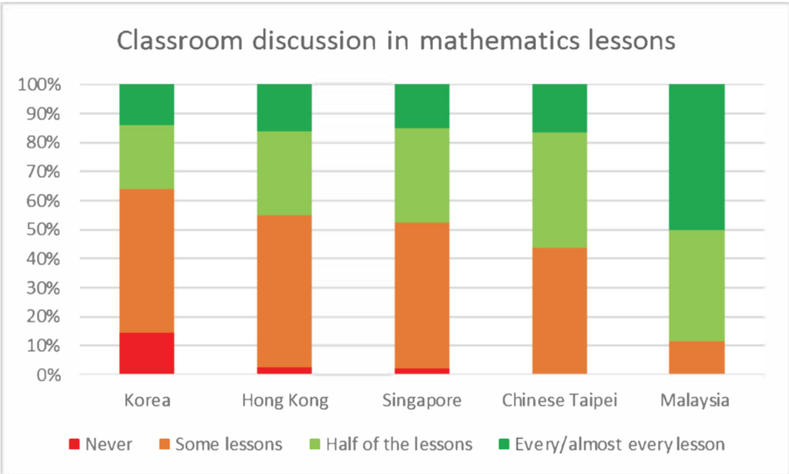

Figure 2. Frequency of classroom discussion in mathematics classes (II)—Asian countries.

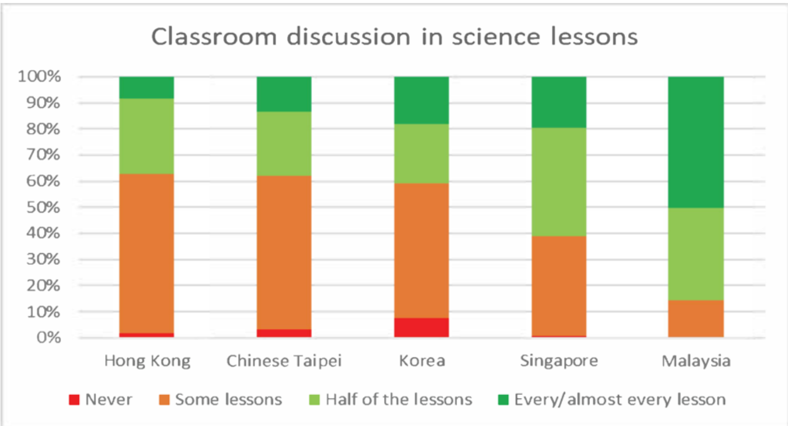

Figure 3. Frequency of classroom discussion in science classes.

Figures 4–7 illustrate that mathematics and science teachers are more likely to use individual dialogical teaching practices like asking learners to share their ideas or to explain their answers to the rest of the class. With the exception of science teachers in Chinese Taipei, the majority (60%+) of mathematics and science teachers across the five Asian countries report using these teaching strategies during half or almost every lesson.

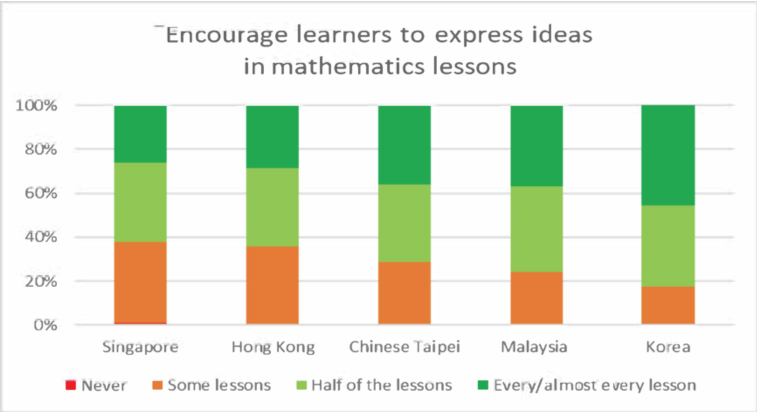

Figure 4. Frequency of asking learners to share ideas in mathematics classes.

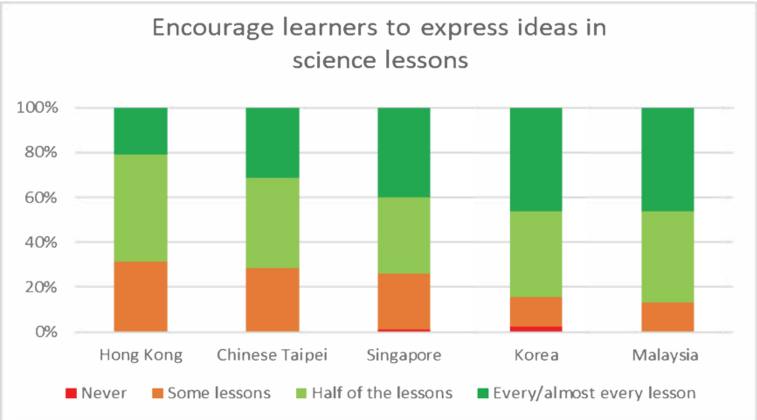

Figure 5. Frequency of asking learners to share ideas in science classes.

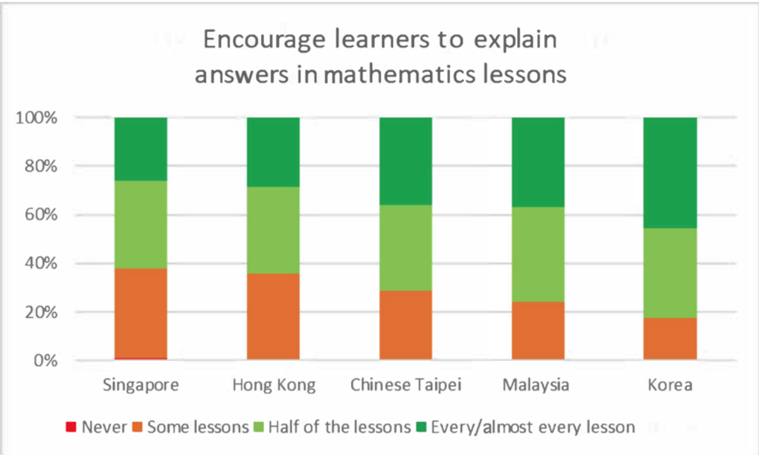

Figure 6. Frequency of asking learners to explain their answers in mathematics classes.

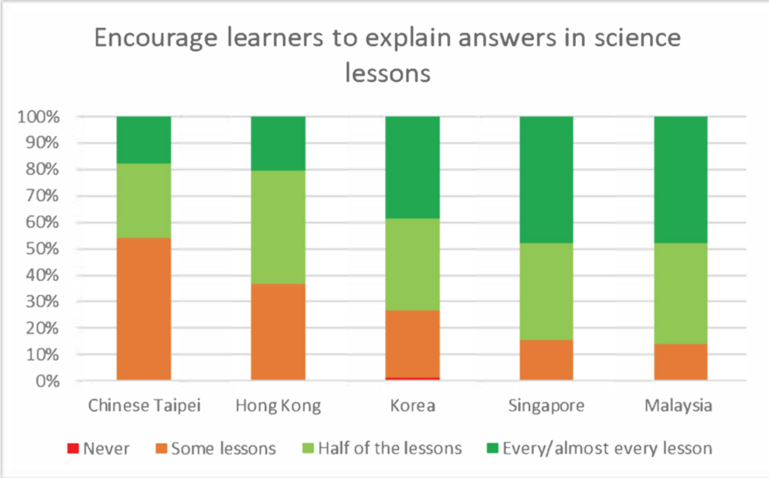

Figure 7. Frequency of asking learners to explain their answers in science classes.

Given that the different teaching practices were part of the same answer options for question 12 in the respective subject teacher questionnaires, it is possible that we simply capture teachers that choose to use dialogical practices in various forms. As is shown in Table 2, the frequencies of using the three dialogical practices are only weakly to moderately correlated (0.39–0.52.). We therefore continue to investigate the impact of each dialogical teaching practice separately instead of their combined usage. This also allows us to exploit more variation across the use of the three teaching practices between the mathematics teachers and the science teachers.

Table 2. Correlation matrix: Use of different dialogical practices in mathematics classes.

	Classroom discussion	Encourage students to express ideas	Ask students to explain answers
Classroom discussion	1.00		
Encourage students to express ideas	0.42	1.00	
Ask students to explain answers	0.39	0.52	1.00

2.3. The econometric approach

The aim of this paper is to investigate the impact of different dialogical teaching practices in mathematics classes on learner performance in the mathematics test section of TIMSS. To do so, we have two different goals to cover: 1) What are the educational characteristics (teachers, classrooms, and schools) that facilitate the use of dialogic practices?; and 2) Do dialogic practices have any impact on student performance in mathematics? To answer the first question, given the ordinal nature of the dialogic teaching practice variable, we use an ordered probit model. For the second question, to avoid the expected endogeneity problem and to establish evidence of casualty, we implement a within-learner-between-subject estimation approach.

Beginning with the first objective, as the three-dimensional teaching model by Praetorius et al. [25] outlines, teachers are likely to choose a pedagogical strategy that reflects the teacher's pedagogic abilities and is feasible in the context of the overall school and classroom dynamics. As such, each teaching strategy is likely to be a function of the teacher's characteristics (gender, teaching experience, enthusiasm, additional pedagogical training, and subject qualification), the teacher's perception of classroom dynamics (perceived level of classroom discipline and level of learner engagement with the teacher), classroom and lesson characteristics (size of the class and total number of minutes that the subject is taught per week), as well as school characteristics that determine the overall learning and teaching environment of the institution (perceived level of school discipline measured in terms of overall safety, obedience, and respect).

First, using an ordered probit model, we test if the frequency of using the three dialogical teaching practices is correlated with the above teacher, classroom, and school characteristics which are likely to either facilitate or hinder the teacher's choice of implementing them in the classroom.

Subsequently, in order to estimate the causal effect of the dialogical teaching practices on learner mathematics test scores, the second objective mentioned at the beginning of this section, we need to account for the potential endogeneity problem highlighted by Praetorius et al.'s [25] model and our findings of the ordered probit estimation. While the rich TIMSS dataset would allow us to control for observed learner, teacher, classroom, and school characteristics, it is possible that the TIMSS questionnaire did not capture all relevant factors that influence the teacher's decision to pursue dialogical teaching practices and simultaneously determine the learners' academic performance in the subject.

To address the problem of estimation bias due to unobservable learner, teacher, classroom, school, and country characteristics, we use the within-learner-between-subject estimation approach (see [27]) and compare the learners' performances in the mathematics test section to their performances in the science test section. Thus, we take the difference between the learners' mathematics and science test scores and regress them against the difference in the frequency of each dialogic teaching practice used by the mathematics teachers and the science teachers. This approach allows us to control for any unobserved factors at the country, school, classroom, and learner level that are subject invariant but simultaneously affect the teachers' choices of using dialogical practices as well as learner performance.

Specifically, we estimate the following first differences:

$$Math_i - Scie_i = diff_i^{M-S} = \beta\left(FreqDiscPrac_{MT,j} - FreqDiscPrac_{ST,j}\right) + \epsilon_i^{M-S}, \quad (1)$$

where the difference between the learner's mathematics and science test scores ($diff_i^{M\text{-}S}$) is determined by the difference in the mathematics ($FreqDiscPrac_{\mathbf{MT},j}$) and science ($FreqDiscPrac_{\mathbf{ST},j}$) teachers' reported frequency of using the dialogical practice j (classroom discussion, encouraging learners to express ideas, and asking learners to explain their answers). In this case, we compare the variation of the learner being exposed to different frequencies of dialogical teaching practices on the difference in the learner's performance in the two subjects (see Figures A1–A4 in the Appendix for variation in the use of dialogical practices and test scores across subjects).

Because we compare performances within the same learner, the learner fixed effect model accounts for all observed and unobserved subject-invariant learner, classroom, school, and country factors that could affect the teacher's choice to use dialogical teaching practices and the learner's performance in the two subjects.

Nevertheless, it is still feasible that subject variant teacher and learner characteristics simultaneously determine the subject teacher's use of the dialogical teaching practice and the learner's performance in the subject. For instance, the learner's preference for a particular subject could make the learner more willing to actively participate in classroom interactions. Thus, the learner exerts more effort in doing well in the subject and the teacher can implement a dialogical teaching practice because learners are more engaged. Similarly, as we compare across different teachers, subject specific teacher characteristics like enthusiasm could still bias our findings. In these two examples, not controlling for confounding variables that simultaneously affect not only the frequency of dialogical practices but also student performance positively would lead to an overestimate of the true impact of dialogical practices on student performance and produce an upward-biased estimate. We therefore control for some of the potential subject-variant teacher and learner characteristics in order to reduce the risk of omitted variable bias.

$$diff_i^{M-S} = \beta\left(FreqDiscPrac_{MT,j} - FreqDiscPrac_{ST,j}\right) + \sigma diffX_{tsi} + \sigma diffZ_{si} + \epsilon_{ist}^{M-S} \quad (2)$$

Specifically, we include the learner's reported preference for each subject as well as subject teacher characteristics including the teacher's gender, years of teaching experience, minutes spent teaching the subject per week, their level of enthusiasm, as well as the teacher's perception of the classroom dynamics.

Because we are limited by the information captured in the TIMSS learner and teacher questionnaires, it is still possible that we do not capture all subject variant factors that could determine the learner's test scores as well as the teacher's choice to use dialogical teaching practices. Nevertheless, we are confident that our identification and estimation strategies do control for most of the bias that could have been introduced through omitted variables. We test the sensitivity of our findings for omitted variable bias following [23] and confirm that our estimates remain moderately sensitive to unobserved confounding variables. We acknowledge this limitation but continue to report our findings in terms of causal interpretations.

3. Results

3.1. The frequency of dialogic teaching practice as a function of the teaching and learning environment: Ordered probit results

In line with the three-dimensional model of the Praetorious et al. [36] model, the learning and teaching environment in the classroom is determined by factors that either facilitate or hinder the use of pedagogic practices (see Table 3). In terms of teacher characteristics and experience, we can see that female mathematics teachers, teachers who have majored in the mathematics, who have in the last 2 years attended professional development courses that focus on teaching critical thinking skills, and who report to be highly motivated are more likely to use dialogical classroom teaching practices. On the other hand, teachers with more years of experience are less likely to use dialogical teaching practices. In terms of classroom characteristics and dynamics, teachers are less likely to use dialogical interactions in larger classes, where they perceive students to be uninterested in the subject, and when students are seen as disruptive. However, the more time the teacher has for teaching the subject each week and the more they feel respected by the learners, the higher the frequency of using dialogical practices during mathematics lessons.

Similarly, school characteristics, as described by the principal of the school and country effects, are also correlated with the teacher's choice of using dialogical teaching methods.

The findings of our ordered probit regression confirm that the teacher's choice to implement dialogical classroom interaction practices is a function of various dimensions that either facilitate or hinder the perceived usefulness of such teaching methods. To the best of our knowledge, none of the existing studies have convincingly controlled for endogeneity bias. Thus, the reported empirical findings of positive and negative associations of dialogic teaching practices and learner performances might simply reflect confounding variables similar to the set of characteristics shown in our ordered probit.

We therefore proceed with the within-learner-between-subject estimator to reduce the potential impact of omitted variable bias in our analysis.

Table 3. Frequency of using dialogical classroom practices in mathematics classes.

	Frequency of using: *Classroom discussions*	Frequency of using: *Learners explain their answers*	Frequency of using: *Learners express their ideas*
Teacher characteristics			
Mathematics teacher female	0.176***	0.215***	0.176***
	(0.0193)	(0.0193)	(0.0195)
Years teaching mathematics	-0.0133***	-0.0207***	-0.0234***
	(0.00112)	(0.00117)	(0.00113)
Mathematics major	0.139***	0.0347	0.117***
	(0.0246)	(0.0234)	(0.0224)
Teacher enthusiastic	0.258***	0.230***	0.369***
	(0.0152)	(0.0157)	(0.0151)
Professional development training: critical thinking pedagogy	0.124***	0.250***	0.380***
	(0.0186)	(0.0187)	(0.0187)
Classroom characteristics and dynamics			
Class size	-0.0127***	-0.00634***	-0.00853**
	(0.00106)	(0.00111)	(0.00103)
Mathematics classes: minutes per week	0.00151***	0.000938***	0.000499***
	(0.000150)	(0.000159)	(0.000139)
Students disruptive	-0.0808***	0.0408**	-0.0154
	(0.0164)	(0.0171)	(0.0171)
Students uninterested	-0.225***	-0.203***	-0.181***
	(0.0189)	(0.0185)	(0.0185)
Students respect teacher	0.227***	0.0592***	0.0894***
	(0.0191)	(0.0196)	(0.0194)
School characteristics			
Safe school index	-0.0868***	0.0127	-0.149***
	(0.0217)	(0.0222)	(0.0225)
Share of learners at school with LoTL at home	0.0133*	-0.0467***	-0.0508***
	(0.00756)	(0.00793)	(0.00786)
School: promotes mathematics	-0.0306*	0.0325*	0.0267
	(0.0184)	(0.0176)	(0.0182)

Continued on next page

	Frequency of using: *Classroom discussions*	Frequency of using: *Learners explain their answers*	Frequency of using: *Learners express their ideas*
School: classroom disturbance	0.0495***	-0.0906***	-0.0520***
	(0.0178)	(0.0170)	(0.0165)
School: students are verbally abused	-0.124***	0.0229	-0.0760***
	(0.0216)	(0.0204)	(0.0198)
School: teachers are verbally abused	0.107***	0.0795***	0.236***
	(0.0234)	(0.0243)	(0.0248)
Country FE			
Chinese Taipei	0.380***	-0.126***	0.438***
	(0.0359)	(0.0390)	(0.0381)
Hong Kong	0.191***	-0.180***	0.278***
	(0.0361)	(0.0391)	(0.0373)
Korea	-0.0434	0.225***	0.680***
	(0.0514)	(0.0509)	(0.0501)
Malaysia	1.189***	0.279***	0.203***
	(0.0341)	(0.0338)	(0.0361)
/cut1	-1.208***	-2.395***	-2.022***
	(0.122)	(0.140)	(0.127)
/cut2	0.770***	0.0179	0.313***
	(0.121)	(0.127)	(0.121)
/cut3	1.828***	1.026***	1.313***
	(0.121)	(0.127)	(0.121)
Observations	18,870	18,870	18,861

Notes: Cluster adjusted, robust standard errors in parentheses (clustered at the class level), *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; sample is weighted using Senate Student weights.

3.2. Dialogic practices and student mathematics performance: First difference model results

Table 4 reports the findings of the first difference estimator. Model 1 refers to Eq (1) where we simply regress the difference in the learner's mathematics test score minus the learner's science test score against the difference of the subject teacher's reported frequency of using one of the three dialogical teaching methods. Model 2 includes additional controls for subject variant teacher and learner characteristics as outlined in Eq (2) in the Methodology section.

Across the three pedagogical methods, learners that are exposed to a higher frequency of dialogical classroom interaction practices experience on average between 11%–14% higher test scores compared to learners that are less exposed. A higher frequency of using classroom discussions as well as encouraging learners to express their ideas have, on average, similar impact on the performance of learners, while asking learners to explain their answers to the rest of the class has a larger positive impact on learner performance. Across the three pedagogical methods, a one-unit increase in the frequency of dialogical practices is on average correlated with learners achieving between 3.1–4.2 units more in mathematics scores compared to learners who do not experience any change in the frequency of dialogical practices between their mathematics lessons and science lessons (as reflected by the constant). Thus, learners that are exposed to a positive change in the frequency of dialogical classroom interaction practices experience on average between 11%–14% higher change in test scores compared to learners who do not experience a change in the frequency of dialogical practices.

Our findings contrast with the study by Zhu and Kaiser [39] who did not find any significant influence of teaching practices on learner performance. However, we need to keep in mind that their approach was quite different using the interviews from the *Global Teaching InSights (GTI)* and data for Shanghai only.

Table 4. First difference of mathematics and science performance by dialogical practice.

	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
VARIABLES	Test Score	Test Score	Test Score	Test Score	Test Score	Test Score
Encourage classroom discussion	3.192***	2.018**				
	(1.073)	(1.013)				
Encourage learners to express ideas			3.133***	2.166*		
			(1.216)	(1.221)		
Ask learners to explain answers					4.198***	3.396***
					(1.244)	(1.200)
Teacher female		3.913*		3.793*		3.752*
		(2.030)		(2.080)		(2.052)
Teaching experience (years)		-0.100		-0.0746		-0.0698
		(0.138)		(0.141)		(0.136)
Teacher enthusiasm		-0.553		-0.747		-0.808
		(1.566)		(1.550)		(1.557)
Minutes per week		0.0632***		0.0632***		0.0626***
		(0.0181)		(0.0179)		(0.0179)
Students uninterested in subject		-2.190		-2.271		-2.229
		(1.647)		(1.630)		(1.623)
Students respect teacher		-0.0839		-0.118		-0.304
		(1.757)		(1.828)		(1.772)
PD critical thinking		0.911		0.709		0.767
		(1.961)		(1.926)		(1.940)
Student’s favorite subject		16.17***		16.19***		16.17***
		(0.555)		(0.552)		(0.553)
Constant	29.15***	32.94***	29.58***	33.22***	29.34***	33.05***
	(1.431)	(1.463)	(1.437)	(1.454)	(1.436)	(1.470)
Observations	19,576	19,433	19,567	19,424	19,576	19,433
R-squared	0.003	0.121	0.003	0.122	0.006	0.124

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; sample is weighted using Senate Student weights.

As expected, the inclusion of subject-variant learner and teacher characteristics does account for some portion of the impact of the teaching strategy on learner performance as the size of our point estimates reduces. Specifically, the decrease of the point estimate and the significant improvement of the fit of the model (R-squared) with the inclusion of the additional controls suggests that our within-learner-between-subject estimator remains vulnerable to omitted variable bias and that the true impact is likely to be even smaller. Using the STATA module PSACALC, we report the deltas for each dialogical practice estimation in Table A6 (Appendix). Under the proportionality assumption, we see that the deltas for the individual dialogic practices vary between 0.0–0.60, suggesting high to moderate sensitivity of our estimate to omitted variable bias. While the findings for classroom discussions seem less vulnerable to omitted variable bias, encouraging students to express their ideas seems highly vulnerable to selection on unobservables. Thus, in future research, further analysis and the inclusion of additional controls may be needed to refine the estimates and address any remaining potential biases.

With our results, we provide empirical support to the findings by Shi [28]. Shi, who also used TIMSS data for five Asian countries/regions, discovered that the relationship between teacher instructional practices and learning outcomes vary within these countries. In her opinion, the teaching context in terms of preparation, academic standards, and teacher evaluation has a major influence of instructional practices on learning outcomes.

As expected, the learner's reported preference for a particular subject accounts for a significant portion of the difference in the learner's test performance between subjects. Similarly, the more time the learner is exposed to the subject, the better the learner's performance in the subject. In terms of teacher characteristics, only the teacher's gender is positively correlated with learner performance. Particularly, learners that are taught by female teachers perform better on average compared to learners that are taught by male teachers, holding other teacher and classroom characteristics constant.

As we saw in the descriptive statistics section, Malaysia is an outlier among the five Asian countries in our sample with regard to the learners' performances in mathematics and science as well as in terms of the use of dialogical teaching practices by mathematics and science teachers. It is possible that our findings are driven by the Malaysian experience. As a robustness check, we exclude Malaysia from the sample and run the regressions on the remaining four Asian countries only. As is shown in Table A7 in the Appendix, the positive impact of learners being exposed to a higher frequency of dialogical classroom interaction practices remains statistically significant for the reduced sample.

3.3. Gendered effects of dialogic teaching practices: Heterogenous effects model results

Female learners exhibit lower levels of confidence in their math abilities [7,12,16,22]. In this context, female learners might also respond differently to interactive classroom practices, especially when they are less confident about the correctness of their approaches and results. For instance, Aguillon et al. [1] reported that female learners are less likely to participate in biology class activities when teachers use active-learning methods, while male learners are overrepresented in dialogic activities. It is therefore possible that the benefits from interactive dialogic classroom participation accrue differently for female and male learners. Using a sub-group analysis, we investigate the differential impact of the three teaching practices for the sample of female and male learners separately.

As we report in Table 5, across the three classroom interaction methods, the point estimates suggest that Asian male learners experience higher gains from exposure to dialogical teaching practices compared to their female peers, particularly when teachers ask learners to explain their answers to the rest of the class. This would be in line with the findings by Aguillon et al. [1] implying that in the Asian context, male learners are not only more likely to participate in active learning practices but also seem to benefit more from exposure to such dialogic practices compared to their female peers. However, given the standard errors, the differences between the gender coefficients are not discernible at a statistically significant level.

Nevertheless, our results still show that both male and female learners achieve higher math scores when they are exposed to a higher frequency of dialogic teaching practices. Dialogical practices like classroom discussions and encouraging individual learners to explain their answers to the rest of the class might help learners in highly competitive learning environments to learn from their peers without exposing their own confusion and ignorance which might be seen as a weakness.

Table 5. Mathematics performance and dialogical practices by learner gender.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Female	Male	Female	Male	Female	Male
Classroom discussion	1.715	2.252*				
	(1.169)	(1.156)				
Explain answers			2.610**	4.148***		
			(1.222)	(1.441)		
Express ideas					2.169*	2.313
					(1.287)	(1.429)
Constant	33.93***	31.84***	34.06***	31.94***	34.19***	32.12***
	(1.530)	(1.724)	(1.542)	(1.729)	(1.530)	(1.719)
Subject variant controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9767	9666	9767	9666	9758	9666
R-squared	0.119	0.125	0.120	0.129	0.120	0.125

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; the specifications include additional controls for subject variant learner and teacher characteristics as outlined in model (2), and the sample is weighted using Senate Student weights.

4. Limitations and future research

Investigating the relationship between dialogic practices and learner performance is difficult. As we outlined above, the decision of the teacher to implement a particular teaching approach is determined by the overall teaching and learning environment of the school, classroom, learner, and teacher’s own background. Besides the complexity of this relationship and the data limitations, our study has faced a number of additional challenges.

For instance, our main variables of interest are self-reported by teachers. It is therefore possible that the ordinal nature of the frequency (“Never”, “Some lessons”, “About half the lessons”, and “Every or almost every lesson”) suffers from measurement error. This could be due to social desirability bias similarly to the bias in the FIT-choice scale [9] or variation in the interpretation of the answer options [24]. The accuracy of self-reported measures can be improved with repeated exposure of the teachers to the same questionnaire instrument [14] but, unfortunately, this is not the case in the TIMSS data collection process. Additionally, the dialogic classroom interactions can take on a variety of forms (see [8]), whereby the scope of the TIMSS questions and answer options of teaching practices might have been too vague to capture how teachers actually implemented these dialogic practices as the frequency itself does not capture the level of interaction between the learners themselves and the teacher. Future research should attempt to use questions that reduce interpretation bias and collect data on the actual implementation of the dialogic practice in more detail.

Finally, we have shown repeatedly the impact of endogeneity on establishing the true impact of dialogic practices on learner performance. While the TIMSS data already provides rich information on the teaching and learning environment, the pedagogical approaches of the teachers are not detailed enough to control for potential confounding variables. Thus, while the within-learner-between-subject estimator is a good step in the right direction, testing the impact of dialogical teaching practices on Asian learner performance in mathematics classes should be done using an experimental design, similar to Alexander [3] in the UK. Even with our attempts to reduce bias, our results remain vulnerable to omitted subject-variant factors at the learner, teacher, and classroom level.

5. Conclusions

The aim of our study was to test if a learner-centred pedagogy like dialogic teaching practices would yield positive learning gains in mathematics for students in countries that are generally described to exhibit a more collectivist and passive teaching and learning classroom environment. Specifically, we wanted to test the validity of the widely held perception that interactive dialogic classroom practices are ineffective in Asian countries. Additionally, by using an identification strategy that allows us to account for the inherent endogeneity between the teacher's choice of teaching practice, the classroom learning environment, and learner performance, we provide more rigorous empirical evidence of the relationship between interactive dialogic teaching practices and learner performance using a within-learner-between-subject estimation strategy.

Our findings challenge the notion that dialogic practices are ineffective in collective learning environments. While overall used sparingly by Asian teachers, we show that learners do benefit from more frequent use of dialogic teaching practices. Thus, the popular belief that the efficacy of certain teaching practices are culture and context specific [38,39] needs to be revisited. Following the survey findings of Loh and Teo [15], Asian learners see the value of such interactive teaching practices and therefore would be open to a more learner-centered teaching environment. As younger teachers seem to be more open to experimenting with learner-centered practices, it is likely that dialogic teaching methods are going to be used more frequently in Asian schools in the future [26].

With respect to heterogenous effects by gender, we find that professional development courses need to sensitize teachers to gender-responsive pedagogies. According to UNESCO [31,32], learner-centered teaching approaches need to be tailored to different demographic groups of learners. Specifically, gender-responsive pedagogy should ensure that "the learning materials, methodologies, content, learning activities, language use, classroom interaction, assessment and classroom set respond to specific needs of boys and girls in the teaching-learning process" ([32], page 6). Thus, the gender of the learners should be considered an important factor in the teacher's decision to adopt a particular teaching approach. While this is not a focus of our current study, our results show that male learners are more responsive to dialogic teaching practices and benefit more from higher exposure compared to their female peers.

However, our analysis has also highlighted the difficulty of investigating the complex relationship between pedagogical practices and learner performance using observational data. Despite trying to reduce the impact of omitted variables on our estimates, various tests have shown that the interrelated relationship of the teaching and learning environment and learner performance is still vulnerable to omitted variable bias. If we want to understand the causal impact of dialogic teaching practices in different learning cultures like the Asian countries, the most promising strategy would be to implement a randomized controlled trial design.

Author contributions

Volker Schöer: Conceptualization (lead), Formal analysis (lead), Writing – original draft (equal), Writing – review and editing (equal); Jose G. Clavel: Conceptualization (supporting), Data curation (lead), Formal analysis (supporting), Writing – original draft (equal), Writing – review and editing (equal).

Conflict of interest

The authors have no conflicts of interest.

References

1. S. M. Aguillon, G. F. Siegmund, R. H. Petipas, A. G. Drake, S. Cotner, C. J. Ballen, Gender Differences in Student Participation in an Active-Learning Classroom, *CBE—Life Sci. Educ.*, **19** (2020), 2. https://doi.org/10.1187/cbe.19-03-0048
2. R. Alexander, *Towards dialogic teaching: Rethinking classroom talk*, 3rd Eds., Cambridge: Dialogos, 2008.
3. R. Alexander, Developing dialogic teaching: Genesis, process, trial, *Res. Pap. Educ.*, **33** (2018), 561–598. https://doi.org/10.1080/02671522.2018.1481140
4. P. Andrews, A. Ryve, K. Hemmi, J. Sayers, PISA, TIMSS and Finnish mathematics teaching: An enigma in search of an explanation, *Educ. Stud. Math.*, **87** (2014), 7–26. https://doi.org/10.1007/s10649-014-9545-3
5. E. Carter, E. Molina, A. Pushparatnam, S. Rimm-Kaufman, M. Tsapali, K. K.-Y. Wong, Evidence-based teaching: Effective teaching practices in primary school classrooms, *London Rev. Educ.*, **22** (2024), 8. https://doi.org/10.14324/LRE.22.1.08
6. D. Clarke, Contingent conceptions of accomplished practice: The cultural specificity of discourse in and about the mathematics classroom, ZDM *Math. Educ.*, **45** (2013), 21–33. https://doi.org/10.1007/s11858-012-0452-8
7. S. J. Correll, Gender and the career choice process: The role of biased self-assessments, *Am. J. Soc.*, **106** (2001), 1691–1730. https://doi.org/10.1086/321299
8. O. G. Drageset, Student and teacher interventions: a framework for analysing mathematical discourse in the classroom, *J. Math. Teacher Educ.*, **18** (2015), 253–272. https://doi.org/10.1007/s10857-014-9280-9

9. I. H. Erten, Social Desirability Bias in Altruistic Motivation for Choosing Teaching as a Career, *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, **30** (2015), 77–89.
10. M. Gil-Izquierdo, J. M. Cordero, V. Cristóbal, Teaching strategy specialization and student achievement, *Educ. Econ.*, **31** (2023), 755–773. https://doi.org/10.1080/09645292.2023.2169252
11. W.-M. Hsu, Examining the Types of Mathematical Tasks Used to Explore the Mathematics Instruction by Elementary School Teachers, *Creat. Education*, **4** (2013), 396–404. http://doi.org/10.4236/ce.2013.46056
12. E. Jouini, P. Karehnke, C. Napp, Stereotypes, under-confidence and decision-making with an application to gender and math, *J. Econ. Behav. O.*, **148** (2018), 34–45. https://doi.org/10.1016/j.jebo.2018.02.002
13. J. König, S. Blömeke, A. Jentsch, L. Schlesinger, C. Nehls, F. Musekamp, et al., The links between pedagogical competence, instructional quality, and mathematics achievement in the lower secondary classroom, *Educ. Stud. Math.*, **107** (2021), 189–212. https://doi.org/10.1007/s10649-020-10021-0
14. S. M. Koziol, P. Burns, Teachers' Accuracy in Self-Reporting about Instructional Practices Using a Focused Self Report Inventory, *J. Educ. Res.*, **79** (1986), 205–209. https://doi.org/10.1080/00220671.1986.10885678
15. C. Y. R. Loh, T. C. Teo, Understanding Asian Students Learning Styles, Cultural Influence and Learning Strategies, *J. Educ. Soc. Policy*, **7** (2017), 1.
16. M. McMurran, D. Weisbart, K. Atit, The relationship between students' gender and their confidence in the correctness of their solutions to complex and difficult mathematics problems, *Learn. Individ. Differ.*, **107** (2023), 102349. https://doi.org/10.1016/j.lindif.2023.102349
17. N. Mercer, L. Dawes, R. Wegerif, C. Sams, Reasoning as a scientist: Ways of helping children to use language to learn science, *Br. Educ. Res. J.*, **30** (2004), 3. https://doi.org/10.1080/01411920410001689689
18. E. Molina, A. Pushparatnam, S. Rimm-Kaufman, K. K. Wong, Effective Teaching Practices in Primary School Classrooms, Policy Research Working Paper 8656, 2018. Available from: https://documents1.worldbank.org/curated/pt/552391543437324357/pdf/WPS8656.pdf.
19. B. Moorhouse, Y. Li, S. Walsh, E-Classroom Interactional Competencies: Mediating and Assisting Language Learning During Synchronous Online Lessons, *RELC J.*, **54** (2021), 114–128. https://doi.org/10.1177/0033688220985274
20. K. Morrison, Paradox Lost: Toward a Robust Test of the Chinese Learner, *Educ. J.*, **34** (2006), 1.
21. I. V. S. Mullis, M. O. Martin, TIMSS 2019 Assessment Frameworks, 2017. Available from: http://timssandpirls.bc.edu/timss2019/frameworks/
22. M. Niederle, L. Vesterlund, Explaining the Gender Gap in Math Test Scores: The Role of Competition, *J. Econ. Perspect.*, **24** (2010), 129–144. https://doi.org/10.1257/jep.24.2.129
23. E. Oster, Unobservable Selection and Coefficient Stability: Theory and Evidence, *J. Bus. Econ. Stat.*, **37** (2017), 187–204. https://doi.org/10.1080/07350015.2016.1227711
24. E. Oz, Comparability of teachers' educational background items in TIMSS: A case from Turkey, *Large-scale Assess. Educ.*, **9** (2021), 4. https://doi.org/10.1186/s40536-021-00097-2
25. A. K. Praetorius, E. Klieme, B. Herbert, P. Pinger, Generic dimensions of teaching quality: The German framework of Three Basic Dimensions, *ZDM Math. Educ.*, **50** (2018), 407–426. https://doi.org/10.1007/s11858-018-0918-4

26. E. Saito, R. Takahashi, J. Wintachai, A. Anunthavorasakul, Issues in introducing collaborative learning in South East Asia: A critical discussion, *Manage. Educ.*, **35** (2021), 167–173. https://doi.org/10.1177/0892020620932367
27. P. Sancassani, The effect of teacher subject-specific qualifications on student science achievement, *Labour Econ.*, **80** (2023), 102309. https://doi.org/10.1016/j.labeco.2022.102309
28. Q. Shi, Relationship Between Teacher Efficacy and Self-Reported Instructional Practices: An Examination of Five Asian Countries/Regions Using TIMSS 2011 Data, *Front. Educ. China*, **9** (2014), 577–602. https://doi.org/10.3868/s110-003-014-0045-x
29. A. Tadesse, S. Lehesvuori, H. Posti-Ahokas, J. Moate, The learner-centred interactive pedagogy classroom: Its implications for dialogic interaction in Eritrean secondary schools, *Think. Skills Creat.*, **50** (2023), 101379. https://doi.org/10.1016/j.tsc.2023.101379
30. R. Tytler, G. Aranda, Expert Teachers' Discursive Moves in Science Classroom Interactive Talk, *Int. J. Sci. Math. Educ.*, **13** (2015), 425–446. https://doi.org/10.1007/s10763-015-9617-6
31. UNESCO, *Integration of Gender-Responsive Pedagogy in pre- and in-service teacher training courses in Ethiopia*, Bangkok: UNESCO Bangkok, 2017.
32. UNESCO Bangkok, Preparation of a comprehensive Gender-Responsive Pedagogy (GRP) Toolkit, 2017. Available from: https://bangkok.unesco.org/sites/default/files/assets/article/Teachers%20Education/GenderAssessment-May2017/Solomon-UNESCO_IICBA.pdf
33. M. Von Davier, E. Gonzalez, W. Schulz, Ensuring validity in international comparisons using state-of-the-art psychometric methodologies, In: *Reliability and Validity of International Large-Scale Assessment*, Cham: Springer, 2020, 187–219. https://doi.org/10.1007/978-3-030-53081-5_11
34. S. Walsh, *Classroom Discourse and Teacher Development*, Edinburgh: Edinburgh University Press, 2013.
35. M. Walshaw, G. Anthony, The teacher's role in classroom discourse: A review of recent research into mathematics classrooms, *Rev. Educ. Res.*, **78** (2008), 516–551. https://doi.org/10.3102/0034654308320292
36. D. A. Watkins, J. B. Biggs, *The Chinese learner: Cultural, psychological and contextual influences*, Hong Kong, Melbourne: Comparative Education Research Centre & Australian Council for Educational Research, 1996.
37. H. Wursten, C. Jacobs, The impact of culture on education. Can we introduce best practices in education across countries? *ITIM Int.*, **1** (2013), 1–28.
38. L. Xu, D. Clarke, Meta-rules of discursive practice in mathematics classrooms from Seoul, Shanghai, and Tokyo, *ZDM-Int. J. Math. Educ.*, **45** *(*2013), 61–72. https://doi.org/10.1007/s11858-012-0442-x
39. Y. Zhu, G. Kaiser, Impacts of classroom teaching practices on students' mathematics learning interest, mathematics self-efficacy and mathematics test achievements: A secondary analysis of Shanghai data from the international video study Global Teaching InSights, *ZDM Math. Educ.* **54** (2022), 581–593. https://doi.org/10.1007/s11858-022-01343-9

Appendix

Descriptive statistics of schools with one subject teacher versus schools with more than one subject teacher.

Table A0. Percentage of students with one or more than one teacher per subject.

	1 Scie Teacher	2 Scie Teachers	3 Scie Teachers	Total
1 Math teacher	88.09%	9.32%	0.53%	97.9%
2 Math teachers	0.89%	0%	1.17%	2.06%
Total	88.98%	10.49%	0.53%	100%

Table A00. Student performance on science and mathematics and the number of teachers per subject.

Mathematic	**Coef**	**Std.Err**	**t**	**t.param**	**p-value**
More Than One Teacher	612.4743	4.1839	146.3878	76.9901	0.0000
Only One Teacher	543.1924	2.3325	232.8801	67.2434	0.0000
Science	**Coef**	**Std.Err**	**t**	**t.param**	**p-value**
More Than One Teacher	563.4831	3.7650	149.6626	75.7935	0.0000
Only One Teacher	518.7908	1.9492	266.1596	75.1614	0.0000

Table A1. Dialogic practices of schools with one subject teacher versus schools with more than one subject teacher.

DIALOGIC PRACTICES										
	More than one teacher per subject					**One teacher for each subject**				
Variable	Obs	Mean	Std. err.	[95% conf. Interval]		Obs	Mean	Std. err.	[95% conf. Interval]	
Classroom discussion_Sci	3677	2.3695	0.0141	2.3418	2.3972	21,599	2.8722	0.0059	2.8606	2.8837
Classroom discussion_Mat	2165	2.6850	0.0188	2.6481	2.7218	21,823	2.8815	0.0060	2.8697	2.8933
Encourage students to express ideas_Sci	3711	3.1822	0.0141	3.1546	3.2097	21,599	3.2316	0.0052	3.2214	3.2419
Encourage students to express ideas_Mat	2165	3.3378	0.0144	3.3097	3.3660	21,855	3.1345	0.0054	3.1240	3.1451
Ask students to explain answers_Sci	3686	2.8608	0.0133	2.8347	2.8869	21,599	3.1224	0.0054	3.1118	3.1331
Ask students to explain answers_Mat	2165	3.1765	0.0154	3.1463	3.2068	21,864	3.0921	0.0053	3.0817	3.1025

Table A2. Teacher characteristics of schools with one subject teacher versus schools with more than one subject teacher.

TEACHER CHARACTERISTICS: MATHEMATICS										
	More than one teacher per subject					**One teacher for each subject**				
Variable	Obs	Mean	Std. err.	[95% conf. Interval]		Obs	Mean	Std. err.	[95% conf. Interval]	
Mathematics Teacher Female	2187	0.7105	0.0097	.6914915	.7295362	21,832	0.6397	0.0032	0.6333	0.6460
Years Teaching Mathematics	2162	14.1706	0.2095	13.75964	14.58146	21,775	15.2908	0.0597	15.1738	15.4078
Mathematics Mayor	2187	0.4334	0.0106	.4126396	.4542095	21,660	0.6695	0.0032	0.6632	0.6758
Teacher Enthusiastic	2187	3.3417	0.0131	3.316073	3.36732	21,834	3.3612	0.0044	3.3526	3.3697
Prof. Devel. Training: Critical Thinking	2187	0.4323	0.0106	.4114951	.453052	21,801	0.4383	0.0034	0.4317	0.4449

TEACHER CHARACTERISTICS: SCIENCE										
	More than one teacher per subject					**One teacher for each subject**				
Variable	Obs	Mean	Std. err.	[95% conf. Interval]		Obs	Mean	Std. err.	[95% conf. Interval]	
Science Teacher Female	3710	0.7339	0.0073	0.7197	0.7481	21,657	0.6716	0.0032	0.6653	0.6778
Years Teaching Science	3675	17.4128	0.1780	17.0639	17.7618	21,588	14.4901	0.0610	14.3705	14.6097
Science Mayor	3389	1.0000	0.0000	1.0000	1.0000	18,514	1.0000	0.0000	1.0000	1.0000
Teacher Enthusiastic	3691	3.3355	0.0106	3.3146	3.3563	21,650	3.3673	0.0043	3.3588	3.3758
Prof. Devel. Training: Critical Thinking	3688	0.4352	0.0082	0.4192	0.4512	21,637	0.4660	0.0034	0.4593	0.4726

Table A3. Classroom characteristics and dynamics of schools with one subject teacher versus schools with more than one subject teacher.

CLASSROOM CHARACTERISTICS AND DYNAMICS: MATHEMATICS										
	More than one teacher per subject					**One teacher for each subject**				
Variable	Obs	Mean	Std. err.	[95% conf. Interval]		Obs	Mean	Std. err.	[95% conf. Interval]	
Class size	2186	27.4717	0.1220	27.2325	27.7109	21,500	29.2154	0.0555	29.1066	29.3243
Mathematics classes: Minutes per week	2187	165.7342	1.1707	163.4384	168.0300	21,593	198.0848	0.4001	197.3005	198.8690
Student disruptive	2163	2.2983	0.0149	2.2692	2.3275	21,640	2.1407	0.0050	2.1309	2.1505
Students uninterested	2163	2.2948	0.0128	2.2696	2.3200	21,647	2.2686	0.0042	2.2604	2.2768
Students respect teacher	2165	2.9595	0.0148	2.9305	2.9886	21,753	3.0134	0.0042	3.0052	3.0216

CLASSROOM CHARACTERISTICS AND DYNAMICS: SCIENCE										
	More than one teacher per subject					**One teacher for each subject**				
Variable	Obs	Mean	Std. err.	[95% conf. Interval]		Obs	Mean	Std. err.	[95% conf. Interval]	
Class size	3683	29.4010	0.1335	29.1394	29.6627	21,464	28.9950	0.0437	28.9092	29.0807
Science classes: Minutes per week	3605	93.4759	0.7120	92.0799	94.8718	21,024	186.8775	0.4001	186.0932	187.6618
Student disruptive	3708	2.3971	0.0095	2.3784	2.4158	21,380	2.0510	0.0049	2.0413	2.0606
Students uninterested	3708	2.2740	0.0098	2.2548	2.2933	21,381	2.1683	0.0043	2.1600	2.1767
Students respect teacher	3711	2.9865	0.0103	2.9664	3.0067	21,569	3.0616	0.0044	3.0530	3.0702

Table A4. School characteristics of schools with one subject teacher versus schools with more than one subject teacher.

SCHOOL CHARACTERISTICS										
	More than one teacher per subject					**One teacher for each subject**				
Variable	Obs	Mean	Std. err.	[95% conf. Interval]		Obs	Mean	Std. err.	[95% conf. Interval]	
Safe School Index	2165	2.2558	0.0114	2.2334	2.2783	21,753	2.2835	0.0036	2.2764	2.2906
Share of learners with LoTL at home	5960	1.0835	0.0070	1.0698	1.0971	43,848	1.9934	0.0071	1.9796	2.0072
School: promotes mathematics	5960	3.5037	0.0072	3.4897	3.5177	43,494	3.5155	0.0027	3.5102	3.5208
School: classroom disturbances	5960	1.8668	0.0104	1.8465	1.8872	43,928	1.8405	0.0037	1.8334	1.8477
School: students are verbally abused	5960	1.8816	0.0105	1.8611	1.9022	43,802	1.6697	0.0035	1.6629	1.6765
School: teachers are verbally abused	5900	1.5089	0.0110	1.4874	1.5305	43,970	1.3344	0.0031	1.3283	1.3406

Table A5. Descriptive statistics of delimited sample: complete cases versus incomplete cases.

Variable	Complete Cases			Incomplete Cases				
	N	Mean	Std. dev.	N	Mean	Std. dev.	Difference	T stat
Frequency of dialogic practices in mathematics lessons								
Classroom discussion	19,845	2.9	0.85	1978	2.7	0.75	0.21	10.5
Encourage learners to express their ideas	19,836	3.1	0.8	2019	2.9	0.8	0.20	10.5
Ask learners to explain their answers	19,845	3.1	0.8	2019	2.9	0.8	0.20	10.6
Teacher characteristics								
Female teacher	19,845	0.6	0.5	1987	5	0.5	0.09	7.9
Years of experience	19,845	14.1	8.6	1930	16.4	9.7	−2.31	−11.1
Mathematics major	19.845	0.7	0.4	1815	0.6	0.5	0.11	10.6
Minutes per week	19,845	211.4	65.0	1748	224.0	60.6	−12.64	−7.8
Teacher enthusiasm	19,845	3.4	0.6	1989	3.2	0.7	0.14	9.2
Professional development: mathematics pedagogy	19,845	0.7	0.5	1924	0.6	0.5	0.10	9.0
Professional development: critical thinking	19,845	0.5	0.5	1956	0.4	0.5	0.04	3.7
Classroom characteristics								
Class size	19,845	30.2	8.1	1655	30.1	6.9	0.06	0.3
Students are difficult to understand	19,845	1.6	0.6	1823	1.7	0.6	−0.10	−6.2
Students are disruptive	19,845	1.9	0.7	1796	1.9	0.7	0.02	1.0
Students are uninterested	19,845	2.1	0.6	1802	2.2	0.6	−0.05	−3.4
Students behave orderly in class	19,845	3.1	0.7	1908	3.1	0.6	0.03	1.9
Students respect teacher	19,845	3.1	0.7	1908	3.1	0.6	−0.01	−0.6
School Characteristics								
Safe school index	19,845	2.4	0.6	1908	2.3	0.6	0.04	3.1
Share of learners in school: language of test as home language	19,845	2.8	1.8	2079	2.9	1.8	−0.07	−1.7
Promotes mathematics	19,845	3.6	0.5	1902	3.4	0.6	0.16	12.1
Classroom disturbance	19,845	1.7	0.7	2119	1.7	0.7	−0.03	−1.7
Verbal abuse of students	19,845	1.6	0.6	2056	1.5	0.6	0.09	5.9
Verbal abuse of teachers	19,845	1.2	0.5	2140	1.2	0.4	0.06	5.0
Teacher absenteeism	19,661	1.3	0.6	2132	1.3	0.5	−0.05	−3.7

Figures A1–A4 show the variation in the frequency of using discursive interaction practices between mathematics and science classes.

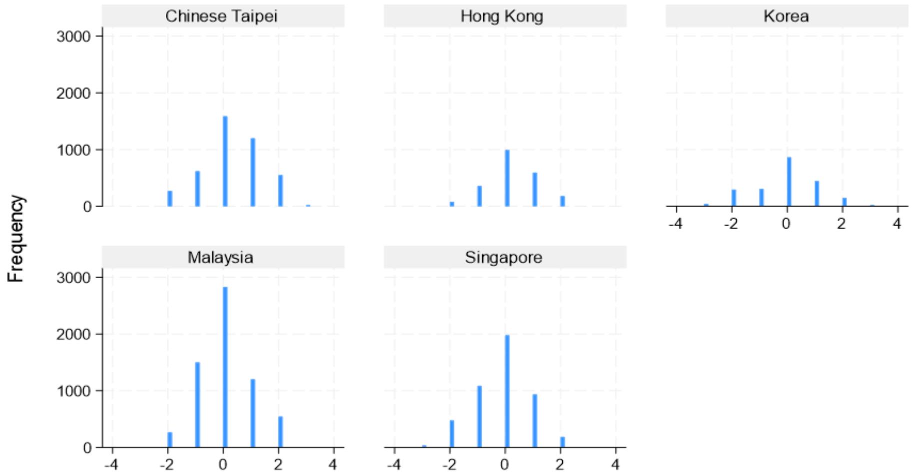

Figure A1. Differences in frequency of using classroom discussions.

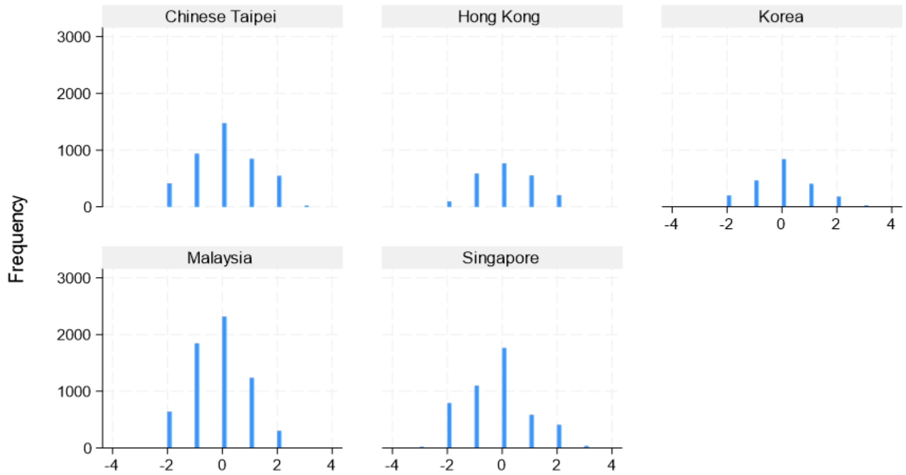

Figure A2. Differences in frequency of asking students to express ideas.

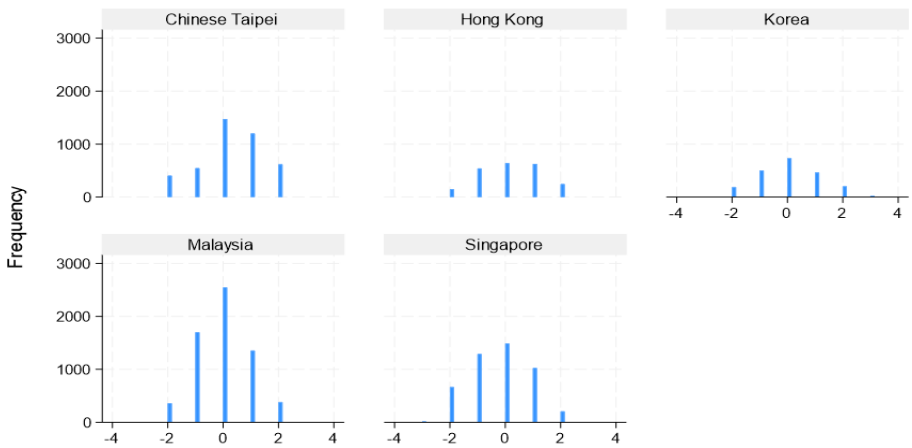

Figure A3. Differences in the frequency of asking learners to explain their answers.

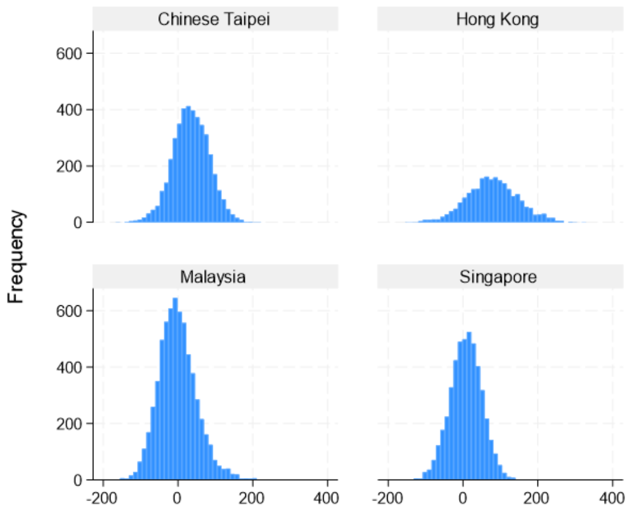

Figure A4. Difference in test scores: Mathematics-Science section.

Table A6. PSACALC results for dialogic practices.

Dialogic Practice	Estimate (β) (Uncontrolled)	R-Squared (Uncontrolled)	Estimate (β) (Controlled)	R-Squared (Controlled)	Delta (Δ)	Bias Direction
Classroom discussion	3.297	0.003	2.345	0.128	0.30	Yes
Explain answers	4.445	0.007	3.736	0.131	0.59	Yes
Express ideas	3.356	0.002	2.648	0.128	0.00	Yes

Notes: Other inputs: RMax = 1; Beta = 0.000.

Robustness check:

Table A7. First difference excluding Malaysia.

	(1)	(2)	(3)
VARIABLES	Test score	Test score	Test score
Classroom discussion	2.182**		
	(1.073)		
Explain answers		3.464***	
		(1.264)	
Express ideas			1.961
			(1.416)
Teacher female	3.195	3.028	3.080
	(2.087)	(2.096)	(2.143)
Teaching experience (years)	-0.101	-0.0616	-0.0716
	(0.143)	(0.145)	(0.150)
Teacher enthusiasm	0.530	0.216	0.402
	(1.787)	(1.796)	(1.767)
Minutes per week	0.0352*	0.0354*	0.0364*
	(0.0191)	(0.0188)	(0.0191)
Students uninterested in subject	-1.510	-1.593	-1.656
	(1.899)	(1.876)	(1.887)
Students respect teacher	-1.112	-1.300	-1.123
	(1.910)	(1.929)	(1.978)
PD critical thinking	0.438	0.229	0.356
	(2.172)	(2.140)	(2.157)
Student's favorite subject	15.33***	15.33***	15.37***
	(0.630)	(0.631)	(0.627)
Constant	40.77***	40.82***	40.97***
	(1.638)	(1.651)	(1.633)
Observations	13,131	13,131	13,122
R-squared	0.109	0.112	0.109

Notes: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.