7 Relationships between teaching practices and student outcomes

Sy Doan, Kata Mihaly and Daniel F. McCaffrey

This chapter explores how the three domains of mathematics teaching described in prior chapters (classroom management, social-emotional support and instruction) are related to three different student outcome measures (academic achievement, personal interest in mathematics and self-efficacy with mathematics).

A central question in education research is understanding how and to what extent teaching influences student outcomes. Prior research indicates that teachers can have a great impact on a wide range of student outcomes, including academic achievement, social-emotional skills, and post-secondary and labour market attainment (Chetty, Friedman and Rockoff, 2014[1]; Jackson, 2018[2]; Kraft, 2019[3]).

Less understood is which dimensions of teaching are most closely associated with various student outcomes and whether the same teaching practices are effective in all educational settings. A deeper knowledge of the relationship between teaching and learning may help inform several policy decisions to better support student learning and well-being as well as spur reflection among the teaching communities on their own practices.

The present Study aims to shed light on these questions through a pre-post design that measures students' academic performance and dispositions both before and after the observed focal unit of the Study – quadratic equations – is taught. This chapter explores how the three domains of mathematics teaching explored in prior chapters (classroom management, social-emotional support and instruction) are related to three different student outcome measures (mathematic achievement, personal interest in mathematics and self-efficacy with mathematics).

To examine the relationship between teaching practices and student outcomes, regression models of both post-unit scores (levels) and post-unit scores after accounting for pre-unit scores and other student and classroom characteristics (growth) of each student outcome are considered (see the *Global Teaching InSights Technical Report* [hereafter the "Technical Report"]). Specifically, these models estimate the average difference in student outcomes associated with a one-unit increase in a classroom's level of management, social-emotional support and instruction.

Key findings

- Higher performing students were more likely to be in classrooms with higher observed levels of teaching practice. In particular, the differences in instructional quality between high and low performing classrooms were significant in Biobío, Metropolitana and Valparaíso (Chile) (hereafter "B-M-V [Chile]"), England (UK) and Mexico. However, after accounting for students' background and their prior mathematical knowledge, teaching practices were significantly associated with test scores only in Colombia.
- In half of the countries/economies, students were significantly more likely to have a higher interest in mathematics when their teachers provided social-emotionally supportive environments. A unit increase in social-emotional support was associated with a 0.1 to 0.3 unit increase on a 1-4 scale.
- Classroom management, social-emotional support and instruction were all significant predictors of student self-efficacy in Mexico and Shanghai (China), with a unit increase in these domains relating to a 0.2 to 0.6 unit increase on a 1-4 scale of self-efficacy.
- Student-reported opportunities to learn were statistically significant predictors of student outcomes in several countries/economies, in particular those related to "reasoning".

Teaching practices are related to levels but not growth in student test scores

Students' mastery of academic content can be a valuable building block for their later success in the classroom and throughout their careers, with teaching being among the most important within-school levers for improving these academic skills (Chetty, Friedman and Rockoff, 2014[1]; McCaffrey et al., 2003[4]). Across eight countries/economies, classroom-level observation scores of three domains of teaching practice (classroom management, social-emotional support and instruction) were examined alongside student performance on an assessment of quadratic equation to assess the extent to which, and which types of, teaching practices are most closely related to students' mathematics knowledge.

Measuring student mathematics knowledge

Students were given assessments to measure their mathematics knowledge prior to and after a focal unit on quadratic equations. The pre-unit measure of mathematics knowledge was constructed using a 30-item mathematics assessment of students' general knowledge on the subject. Students' post-unit knowledge was measured using a 24-item assessment that focused specifically on knowledge of topics related to quadratic equations. Specific subtopics measured on the post-test included the ability to solve quadratic equations through factoring, completing the square, and recognising and using the general quadratic equations through factoring. Students' pre- and post-unit test scores are scaled to range from 1-300. As with other outcome measures, students' post-unit test scores serve as the primary outcomes in each regression, with pre-unit scores used as a control variable to model students' growth .

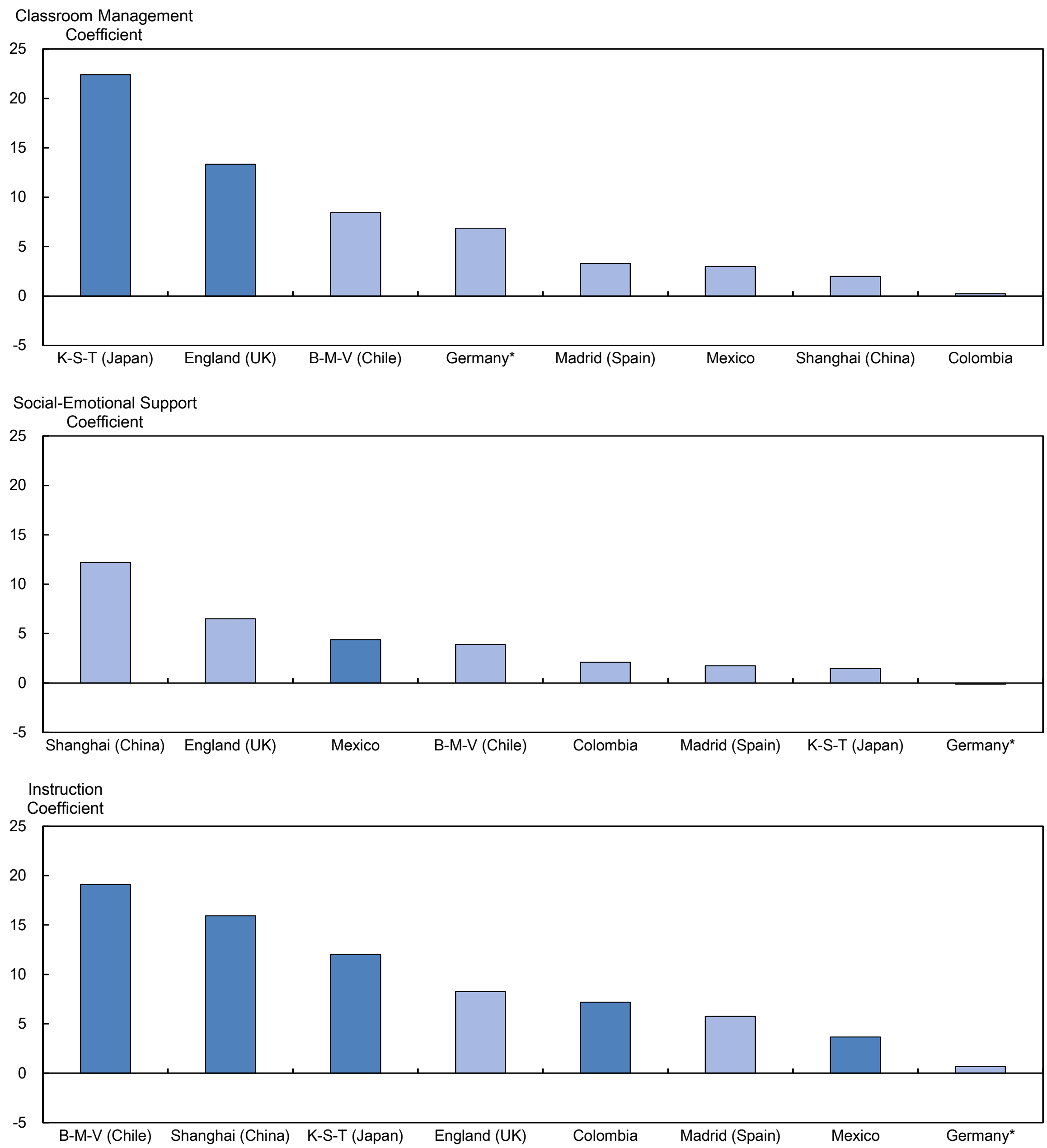
High performing classrooms had better instruction

Examining the relationship between classroom-level teaching and students' performance can help identify which practices are most present in classrooms with high mathematics performance. To do so, regression models are fit separately for each combination of country/economy, domain and outcome in order to examine the extent to which differences in observed domain scores are predictive of student outcomes. Two types of regression models are fit: a "levels" model which estimates the association between post-unit outcome measures and domain scores without accounting for additional characteristics, and a "growth" model which takes into account student's pre-unit scores, background characteristics and classroom averages of these student factors.[1] Figure 7.1 and Figure 7.3 show respectively the results of the "levels" and "growth" models (regression coefficients are available in Annex 7.A).[2]

Without accounting for students' pre-unit test scores and other characteristics, instructional practices were positively associated with post-test scores in many countries/economies. Specifically, in B-M-V (Chile), Colombia, Kumagaya, Shizuoka and Toda (Japan) (hereafter "K-S-T [Japan]"), Mexico and Shanghai (China), students in classrooms with higher levels of instruction, on average, performed significantly higher than their peers on the assessment of quadratic equation knowledge. The largest association was observed in B-M-V (Chile), where a one-unit increase in instruction scores was associated with a 19 point, or three-quarters of a standard deviation, increase in post-test scores.

Figure 7.1. Relationships between observed teaching practices and student post-test scores

Regression estimates of the expected change in a scale of students' post-unit mathematics achievement given a one-point increase in classroom management, social-emotional support and instruction domain scores



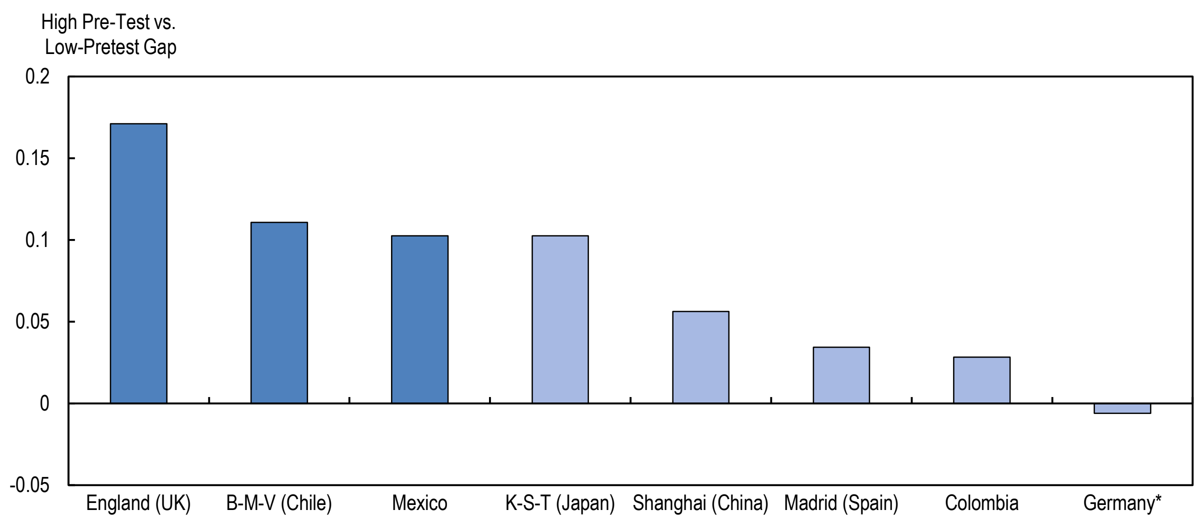
Notes: Bars represent the estimated increase in student post-test scores associated with a one-unit increase in classroom management (top panel), social-emotional support (middle panel) and instruction (bottom panel) in each country/economy. Darker shading indicates that a given association is statistically significant at the $p < .05$ level. Bars are sorted according to the magnitude of the coefficient, with the largest, most positive coefficients on the left of the figure.
Students' raw responses on both the post-unit assessments were scaled using IRT methods to range from 100 to 300, with a score of 200 representing the average test performance across students and a standard deviation equal to 25 points.
*Germany refers to a convenience sample of volunteer schools.
Countries and economies are ranked in a descending order by coefficient size.
Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188120

Classroom management and social-emotional practices were also positively associated with student post-test scores, albeit in fewer countries/economies at a statistically significant level. In both England (UK) and K-S-T (Japan), classroom management scores were significant positive predictors of post-test scores, with the size of these associations being 22 and 13 test score points, respectively. Social-emotional support was a significant predictor of post-test scores only in Mexico though was positively associated with post-unit scores in all countries/economies except Germany*[3]. This is not surprising given that the instruction domain, as defined in Chapter 5, is focused on teaching practices most directly linked to academic content knowledge (Pianta and Hamre, 2009[5]).

Figure 7.2. Comparing observed instruction between low and high pre-test classrooms

Difference in classroom-level instruction scores between classrooms with above-median ("high") average pre-test test scores and below-median ("low") average pre-test scores



Notes: Figure shows, for each country/economy, the difference in instruction domain scores between classrooms with an above-median average pre-test score for that country-economy ("High Pre-Test") and classrooms with a below-median average pre-test score for that country-economy ("Low Pre-Test"). Darker shading indicates that the difference between low and high pre-test classrooms is statistically significant at the $p < .05$ level. Countries/economies are sorted according to the size of the difference between high pre-test and low pre-test classrooms. Instruction domain scores range from 1-4 and pre-test and post-test scores range from 100-300.

*Germany refers to a convenience sample of volunteer schools.

Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188139

Students enter the classroom with different amounts of knowledge, disposition and home resources. Numerous studies indicate that these "background" characteristics are both highly predictive of not only students' outcomes (Hanushek and Woessmann, 2017[6]), but also, of the types of schools and classrooms students are likely to attend (Kalogrides, Loeb and Beteille, 2013[7]; Givord, 2019[8]).

Indeed, in many participating countries/economies, students' pre-unit measures of knowledge, dispositions and demographic characteristics were often associated with the levels of teaching practice they encountered during the quadratic equations unit. Figure 7.2 shows the difference in the average instruction domain scores among classrooms with above-median ("high") average pre-test scores and classrooms with below-median ("low") average pre-test scores. More positive values indicate that high pre-test classrooms, on average, received higher instruction scores than low pre-test classrooms. This gap is largest among classrooms in England (UK), where high pre-test classrooms had instruction scores that were, on average, .17 units higher than low pre-test classrooms. Overall, the gap in instruction between high and low pre-test classrooms is statistically significant in three countries/economies: B-M-V (Chile), England (UK) and Mexico.

Certain types of students being more likely to attend classrooms with higher (or lower) levels of teaching quality may reinforce and exacerbate existing gaps and inequities in education. Additionally, the link between students' backgrounds and their exposure to different teaching practices complicates the task of identifying how these practices relate to student outcomes.

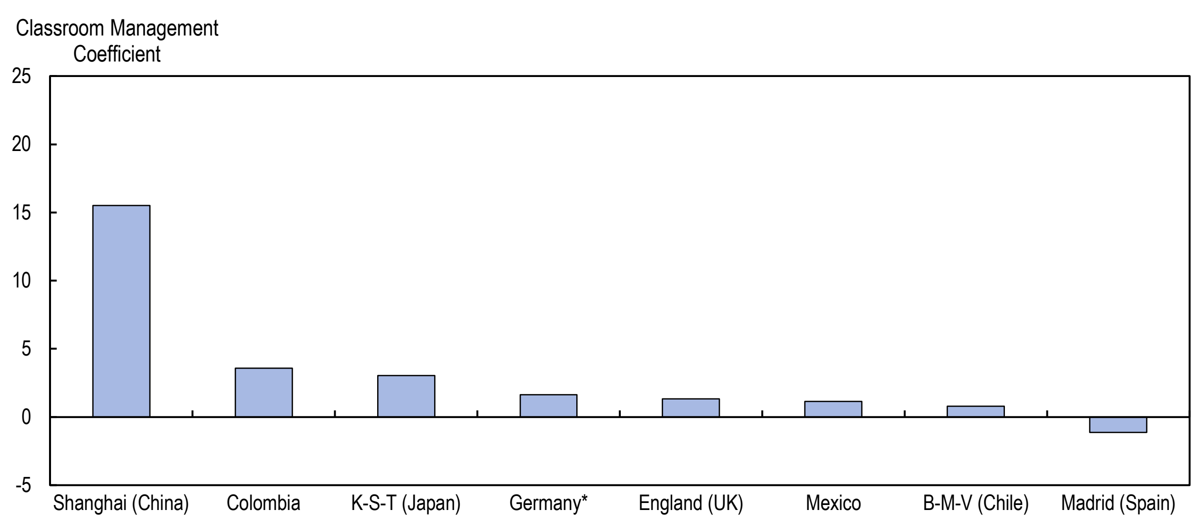
Therefore, to examine the impact of teaching practices on student learning, it is important to take their background into consideration. While domain scores, particularly in instruction, are significantly related to student post-test scores in many countries/economies, these associations may reflect students' background characteristics rather than the quality of teaching if the former are not accounted for. To better examine the effect of teaching on student performance, regression models are fit that account for student's pre-test score in addition to the following: students' home possessions, immigration background, one or both of their parent's highest level of education, and classroom-level averages of pre-unit and background measures. This expanded model is referred to as the "growth" model, as it accounts for students' pre-unit scores on the outcome being analysed.[4]

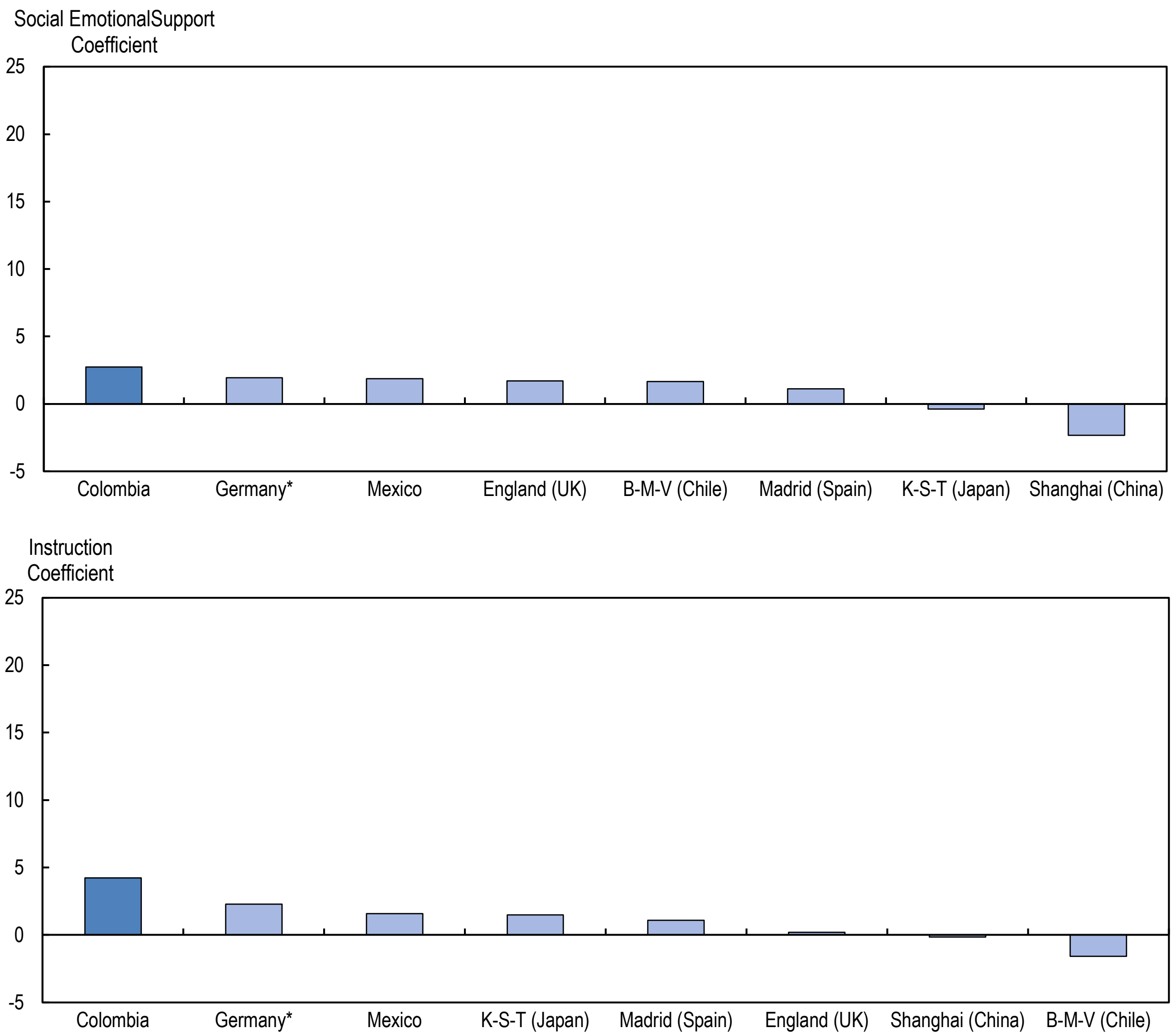
Teaching practices did not explain to students' performance when accounting for their background

Figure 7.3 shows the relationships between teaching practices and student growth in test scores. Specifically, each bar represents the estimated association a unit increase in observed teaching practices in each domain and students' post-test scores, adjusted for students' pre-test scores and background characteristics.

Figure 7.3. Relationships between observed teaching practices and growth in student test scores

Regression estimates of the expected change in a scale of students' post-test scores given a one-point increase in classroom management, social-emotional support and instruction domain scores, adjusting for student and classroom baseline characteristics





Notes: Bars represent the estimated difference in student post-test scores associated with a one-unit increase in classroom management (top panel), social-emotional support (middle panel) and instruction (bottom panel) in each country/economy, adjusted for students' pre-test scores, demographic characteristics and classroom averages of these characteristics. Darker shading indicates that a given association is statistically significant at the $p < .05$ level. Bars are sorted according to the magnitude of the coefficient, with the largest, most positive coefficients on the left of the figure. The coefficient represented by each bar was estimated in a separate country/economy-domain-outcome regression model.

*Germany refers to a convenience sample of volunteer schools.

Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188158

The relationships with student growth in test scores in Figure 7.3 follow the same direction (i.e. increases in teaching practices are generally positively associated with student test score levels and growth) across countries/economies and domains, but the magnitude and significance of these relationships differ considerably. Whereas instruction was significantly associated with student test score levels in five of eight participating countries/economies, this domain was a significant predictor of test score growth only in Colombia, where a one-unit increase in instruction was associated with a modest 4.2 unit increase in student test score growth. The only other significant test score growth relationship for the other two domains of teaching practice was also found in Colombia, where social-emotional support had a significant, but similarly modest, positive association (2.7 unit increase).

Students predisposed to perform well on the post-test assessment, by virtue of higher pre-test scores, were more likely to be in classrooms that enjoy high levels of each teaching practice (Figure 7.2). After controlling for students' pre-test scores and other background characteristics, the relationship between teaching practices and student test scores diminishes greatly, a result that echoes prior work conducted by Carnoy, et al., (2016[9]).

Several potential reasons might explain the lack of significant relationships between teaching practices and student performance, once students' previous achievement and other background characteristics are accounted for. First, while observed teaching practices may be telling of the levels of achievement of the students in that classroom, these differences in achievement may be more reflective of students' background characteristics rather than the direct contribution of these practices to their achievement. Additionally, the potential student growth over the brief period of time in which the focal unit is taught may be limited, and therefore, more difficult to detect.

Classroom management and social-emotional practices were positively associated with student interest

Developing students' interest in academic content is an important precursor for not only their acquisition of that content, but also, their sustained engagement with academic endeavours in the future (Harackiewicz and Hulleman, 2010[10]). The promotion of students' interest is linked to increases in attention, recall, task persistence, and effort toward both proximate tasks such as classwork and reading comprehension (Ainley, Hidi and Berndorff, 2002[11]; Hidi and Renninger, 2006[12]). Additionally, personal interest is associated with more distal student outcomes such as selection of hobbies and academic courses (Harackiewicz et al., 2002[13]). Beyond performance and attainment, developing student interest in their daily activities serves a critical role in promoting student well-being and satisfaction (Sheldon and Elliot, 1999[14]).

Measuring student personal interest in mathematics topics

To measure their personal interest in mathematics, students were asked to indicate their agreement with a series of statements both prior to and after the focal unit. The pre-unit measure of personal interest was formed by averaging student responses across four items that asked to indicate, on a 1-4 scale ranging from "strongly disagree" to "strongly agree", about their general interest in mathematics under their previous teacher (e.g. "I was interested in mathematics" and "I often thought that what we were talking about in my mathematics class was interesting"). The post-unit measure, on the other hand, used three items that asked about students' interest specifically in quadratic equations with their current teacher (e.g. "I was interested in the topic of quadratic equations" and "I often thought that what we were talking about in my mathematics class during the unit on quadratic equations was interesting") using the same "strongly disagree" to "strongly agree". Student responses to each item were averaged to form a single scale of "personal interest in quadratic equations", ranging from 1-4, with higher values in the scale corresponding to greater personal interest in mathematics. The standard deviation of the post-unit personal interest ranged from 0.6 to 0.8 across countries/economies.

Classroom management and social-emotional support are related to student interest in several countries/economies

Using analogous techniques for test score analyses, regression models of students' personal interest in mathematics are fit to examine whether the variation in personal interest is explained by differences in student exposure to diverse teaching practices.

Teaching practices, particularly those in the classroom management and social-emotional support domains, were related to students' reported interest in quadratic equations. Additionally, unlike for student test scores, the relationships between teaching practices and personal interest were observed for both levels or growth of personal interest. Because of this, estimates examining the relationships between teaching practices and growth in student personal interest are focused upon. Figure 7.4 shows the estimated relationship between a one-unit increase in each domain of teaching practice and student growth in personal interest in quadratic equations in a regression model that takes into consideration students' prior interest in mathematicchiles and their background characteristics. (Annex 7.A presents associations between teaching practices and both levels and growth of personal interest).

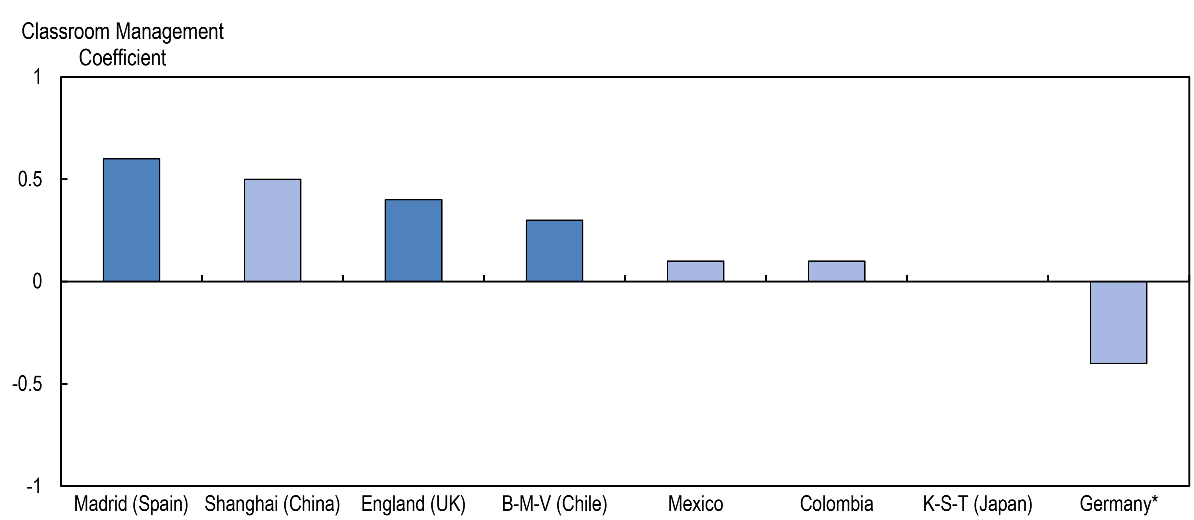
Students in several of the Study countries/economies are significantly more likely to have higher interest in mathematics when their teachers provide for social-emotionally supportive environments. Specifically, half of the countries/economies (B-M-V [Chile], Colombia, Madrid [Spain] and Shanghai [China]) show statistically significant relationships between the level of social-emotional support practices and personal interest in mathematics. In these countries/economies, a unit increase in social-emotional support (on a 1-4 scale) is associated with a 0.1 to 0.3 point increase (on a 1-4 scale) in students' personal interest.

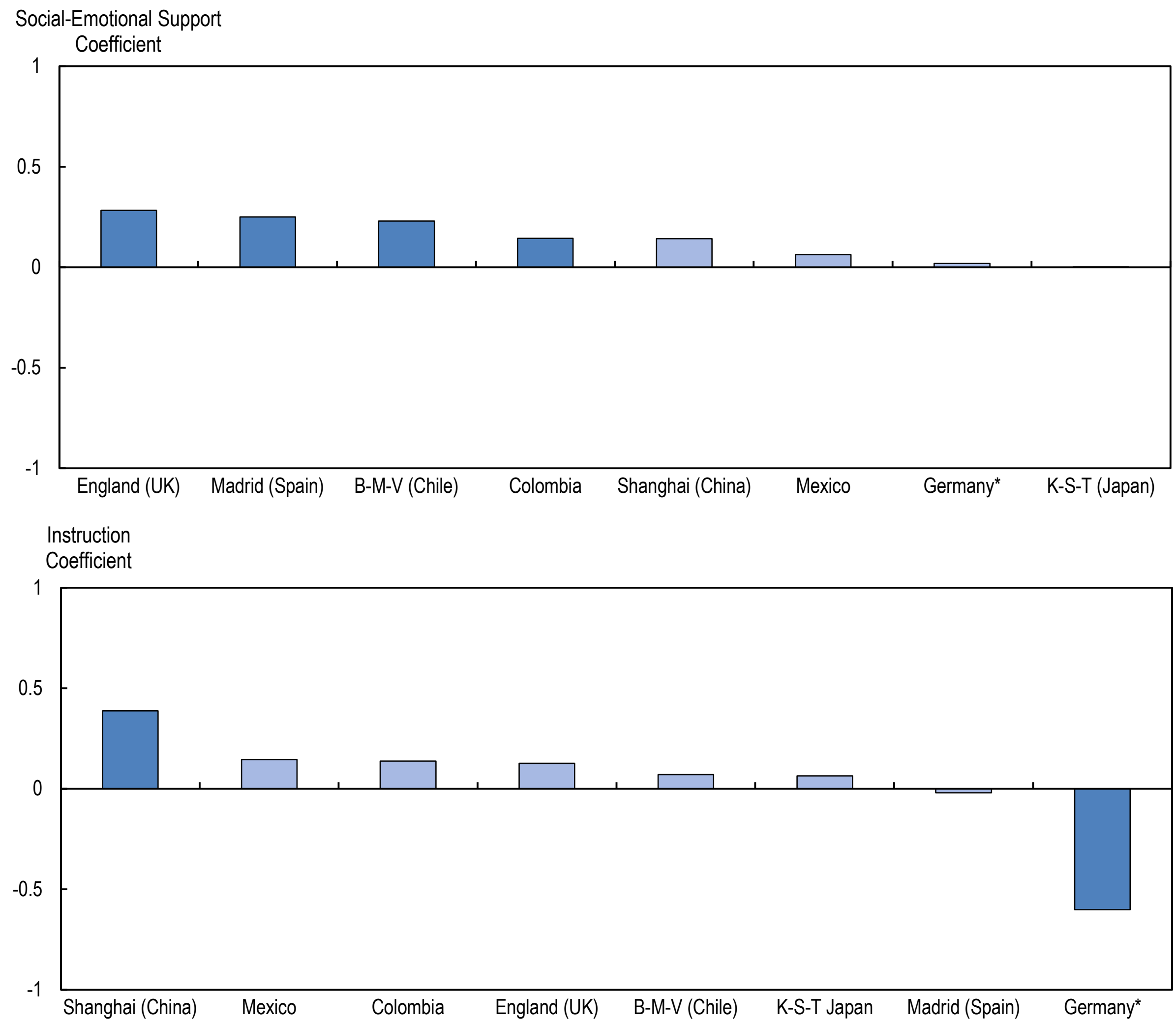
Additionally, students are also likely to significantly report greater interest in mathematics when their classrooms are well-managed. Classroom management scores are significant, positive predictors of student's growth in personal interest in mathematics in B-M-V (Chile), England (UK) and Madrid (Spain), with these associations ranging in size, from a 0.3 to 0.6 unit increase in personal interest for a one-unit increase in classroom management.

The quality of instruction, however, did not appear to have a widely prevalent relationship with students' personal interest in mathematics. Instructional practices are only significantly positively related to students' personal interest in Shanghai (China) and are negatively related in Germany*, where a unit increase in instruction is associated with a 0.6 point decrease in personal interest.

Figure 7.4. Relationships between observed teaching practices and growth in student personal interest

Regression estimates of the expected change in a scale of students' post-unit personal interest in the topic of quadratic equations given a one-point increase in classroom management, social-emotional support and instruction domain scores, adjusting for student and classroom baseline characteristics





Notes: Bars represent the estimated difference in student post-unit personal interest scores associated with a one-unit increase in classroom management (top panel), social-emotional support (middle panel) and instruction (bottom panel) in each country/economy, adjusted for students' pre-unit personal interest scores, demographic characteristics and classroom averages of these characteristics. Darker shading indicates that a given association is statistically significant at the $p < .05$ level. Bars are sorted according to the magnitude of the coefficient, with the largest, most positive coefficients on the left of the figure. The coefficient represented by each bar was estimated in a separate country/economy-domain-outcome regression model.
*Germany refers to a convenience sample of volunteer schools.
Countries and economies are ranked in a descending order by coefficient size.
Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188177

Teaching practices were positively related to students' self-efficacy

Lastly, research has found that students' self-efficacy, or the extent to which individuals feel confident in their ability to perform tasks that are required of them, is related to their academic and social-emotional outcomes, such as academic achievement and classroom task participation on the one hand, and depression and anxiety on the other (Bandura, 1977[15]; OECD, 2019[16]; Rhew et al., 2018[17]). Both as a scaffold for future learning and in its own right, student self-efficacy is another valuable outcome for stakeholders to consider how best to support through classroom teaching.

Measuring student self-efficacy toward mathematics

Similar to personal interest, two separate measures of self-efficacy are constructed: a “pre-score” measure of students’ perceived self-efficacy with regard to general mathematics topics under the previous teacher and a post-score measure of self-efficacy tailored specifically to students’ own feelings of self-efficacy toward quadratic equations under their current teacher. Students reported on a four-point scale, ranging from 1 (“Not at all true of me”) to 4 (“Extremely true about me”), their agreement with statements such as “I believe I would receive an excellent mark for mathematics/the topic of quadratic equations” and “I was confident I would understand the most difficult material covered in mathematics/during the unit on quadratic equations”. Students’ scores on each item are averaged together to form pre-unit and post-unit scales of self-efficacy, ranging from 1-4, with the standard deviation of post-unit self-efficacy ranging from 0.6 to 0.8 units across countries/economies.

Teaching practices in all three domains were linked to self-efficacy in half of countries/economies

Teaching practices across all three domains were significantly related to levels and growth of student self-efficacy, with all three domains being significant predictors in Shanghai (China) and Mexico. Figure 7.5 shows that observed classroom management, social-emotional support and instructional practices are positively related to student growth in self-efficacy, even after accounting for student and classroom-level pre-unit measures of self-efficacy and other demographic and socio-economic characteristics. Similar to the presentation of results pertaining to students’ personal interest in mathematics, regression results examining growth in student self-efficacy, which are modelled with controls for pre-unit self-efficacy and student background characteristics, are emphasised below. Full results for both the levels and growth models of self-efficacy are available in Annex 7.A.

Classroom management was significantly related to growth in self-efficacy in four out of eight countries/economies (Mexico, England [UK], Madrid [Spain] and Shanghai [China]), with the association of a one-unit increase in classroom management ranging from a 0.2 unit (Mexico) to a 0.6 unit (Shanghai [China]) increase on the 1-4 self-efficacy scale.

In half of the participating countries/economies, social-emotional support was significantly related to student growth in self-efficacy (Colombia, Madrid [Spain], Mexico, Shanghai [China]). Yet, across all countries/economies, the magnitude of these associations was quite modest, with significant relationships ranging from a 0.1 to 0.2 unit increase in self-efficacy for a one-unit increase in social-emotional support.

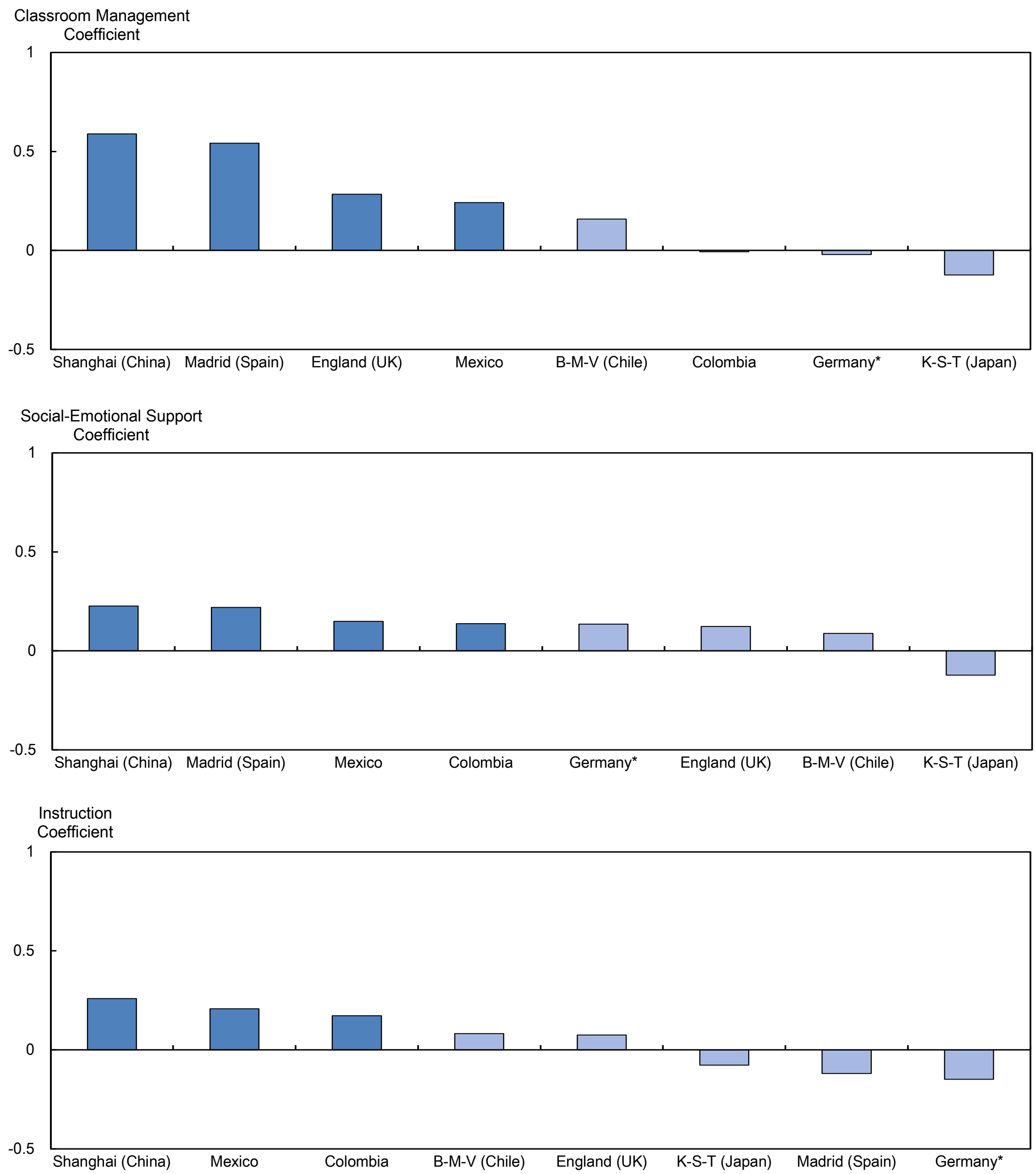
Lastly, students’ exposure to different levels of instruction was a significant predictor of self-efficacy growth in Shanghai (China), Mexico and Colombia. Specifically, a unit increase in classroom instruction scores is associated with between a 0.2-0.3 unit increase in student-reported self-efficacy.

Interestingly, while, for any given country/economy, only select domains were significant predictors of students’ post-unit test scores and personal interest in mathematics, all three domains of teaching (classroom management, social-emotional support, instruction) were significant predictors of student self-efficacy in two countries/economies: Mexico and Shanghai (China). There were no significant relationships between observed teaching practice in any domain and self-efficacy in B-M-V (Chile), Germany* and K-S-T (Japan).

The relationships with student self-efficacy shown in Figure 7.5, along with those for personal interest growth (Figure 7.4) suggest that there are robust associations between teaching practices and student dispositions that appear regardless of whether levels or growth in these outcomes are examined. As increasing attention is paid to how teaching can influence students in ways beyond academic achievement, the results in this report suggest that teaching practices is potentially an important lever for affecting these key measures of student dispositions.

Figure 7.5. Relationships between observed teaching practices and growth in student self-efficacy

Regression estimates of the expected change in a scale of students' post-unit self-efficacy in the topic of quadratic equations given a one-point increase in classroom management, social-emotional support and instruction domain scores, adjusting for student and classroom baseline characteristics



Notes: Bars represent the estimated difference in student post-unit self-efficacy scores associated with a one-unit increase in classroom management (top panel), social-emotional support (middle panel) and instruction (bottom panel) in each country/economy, adjusted for students' pre-unit self-efficacy scores, demographic characteristics, and classroom averages of these characteristics. Darker shading indicates that a given association is statistically significant at the $p < .05$ level. Bars are sorted according to the magnitude of the coefficient, with the largest, most positive coefficients on the left of the figure. The coefficient represented by each bar was estimated in a separate country/economy-domain-outcome regression model.

*Germany refers to a convenience sample of volunteer schools.

Countries and economies are ranked in a descending order by coefficient size.

Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188196

Teaching materials and opportunities to learn were not strongly related to student outcomes

In addition to video observations, the Study also measured teaching practice through the use of teaching materials (as reported in Chapters 3, 4 and 5), and both teacher-reported and student-reported measures of Opportunities to Learn (OTL) in four subtopics: real-world contexts, reasoning, quadratic functions and algebraic operations (as reported in Chapter 6). These measures capture different aspects of how students experience teaching in each classroom that are distinct from video observations, and therefore, possibly capture distinct channels through which teaching affects student outcomes. To test whether this is the case, the relationships between teaching materials, teacher and student reported opportunities to learn and student outcomes were modelled using the same "levels" and "growth" regression models used for the video observation domains above.

Generally, teachers' teaching materials and their self-reported measures of opportunities to learn showed little, if any, consistent significant relationships with either levels or growth of the three student outcome measures (see results in the Technical Report). The lack of strong connections between teaching materials and student outcomes potentially signal differences between planned and enacted curriculum. Many teachers' teaching materials indicate how teachers intend to provide instruction but do not, unlike video observations, capture how teaching actually unfolds in the classroom. While teachers' teaching materials are intimately connected to the instruction they eventually provide in the classroom (Charalambous and Hill, 2012[18]), these results suggest that it is the instruction that students ultimately experience in the classroom that is more closely linked to their outcomes.

Similarly, while teacher-reported opportunities to learn was not strongly related to student outcomes, student reports were significant predictors of student outcomes in several countries/economies. This is consistent with earlier work on the connection between student-reported opportunities to learn, student achievement and mathematics literacy (Schmidt et al., 2015[19]). In particular, when looking at growth of student outcomes, the student-reported "reasoning" was a significant positive predictor of test scores, personal interest and self-efficacy in certain countries/economies (Table 7.1). Notably, student-reported "reasoning" was a significant predictor of all three student outcomes in England (UK).

Table 7.1. Relationships between student-reported "reasoning" OTL and growth in student outcomes

Regression estimates of the expected change student post-test scores, post-unit personal interest and post-unit self-efficacy for a one-unit icnrease in student-reported "reasoning" OTL subscale, adjusted for student and classroom baseline characteristics

Country/economy	Test scores		Personal interest		General self-efficacy	
B-M-V (Chile)	6.229	**	0.034		0.179	
	(2.037)		(0.120)		(0.100)	
Colombia	2.403		0.246	**	0.167	*
	(1.415)		(0.088)		(0.077)	
England (UK)	4.045	*	0.252	*	0.172	*
	(1.690)		(0.104)		(0.073)	
Germany*	-3.140		0.149		0.281	*
	(1.729)		(0.159)		(0.128)	
K-S-T (Japan)	-3.869		0.046		0.117	
	(2.834)		(0.105)		(0.083)	
Madrid (Spain)	3.521		0.051		-0.042	
	(2.088)		(0.110)		(0.112)	
Mexico	0.035		0.301	**	0.123	
	(1.291)		(0.104)		(0.081)	
Shanghai (China)	20.627		0.081		0.495	*
	(10.968)		(0.262)		(0.231)	

Notes: * p < . 05, ** p < .01, *** p < .001. Coefficients represent the estimated increase in a given post-score student outcome associated with a one-unit increase in the student-reported OTL "Reasoning" subscale, adjusted for students' pre-unit scores, demographic characteristics and classroom averages of these characteristics.
*Germany refers to a convenience sample of volunteer schools.
Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188215

OTL, and subtopics within OTL, may impact student learning in two ways. First, OTL can impact student learning by shaping and enabling quality teaching practices. Certain types of content matter may allow for either more efficient teaching practices to be implemented or ubiquitous teaching practices being implemented with higher quality. Framed differently, learning is supposed to depend on the quality of teaching, and the quality of teaching may in turn depend on the implemented curriculum.

Second, how students report their classroom experiences likely has a direct linkage to how well they acquire classroom content. This relationship rests on the assumption that content implemented in class and perceived by the students has a higher probability of being taken up and learnt by the students than content which is either not implemented or not perceived by students. Prior research from Carnoy et al. (2016[9]) which uses international data collected from TIMSS and PISA to connect the amount of time students spend on specific content and its prevalence in the curriculum suggests that this is a valid assumption.

Connections to prior literature on teaching practices and student outcomes

There is a rich literature on the linkage between observed teaching practices and student outcomes. However, few, if any, single studies match the breadth of settings, student outcomes and measurement modes contained in this analysis. Taken together, prior research finds that, similar to this study, observable measures of teaching are modestly related to student outcomes; but, ultimately, the size and strength of these relationships are highly dependent on the context and design of any particular study.

Lipowsky et al. (2009[20]) examine the relationship between teaching practices and student achievement over the duration of a unit on the Pythagorean theorem, finding that two of the three domains of teaching practice they examined (classroom management and cognitive activation, the latter being akin to the instruction domain used in the present study) were significantly, but modestly, associated with student post-test performance after accounting for pre-test measures and other student characteristics.

With its use of a single focal unit, Lipowsky et al. (2009[20]) is structurally similar to the present Study, albeit only conducted in the context of one country (Germany*) compared to the eight countries/economies included in the present study in addition top other differences in study implementation. The ability to examine relationships, or lack thereof, across multiple educational settings and variation in study implementation across those settings, grants a deeper, more nuanced understanding of how teaching and learning are related that might be missed if studies are conducted only within a single context. Importantly, were this study to focus exclusively on Colombian classrooms, where two out of three domains of teaching practice were significantly associated with student test score growth, the results would be comparable to those in Lipowsky et al. (2009[20]); However, the ability to study classrooms across multiple countries/economies grants insight into the generalisability, or lack thereof, of the relationships between practice and outcomes, highlighting the importance of context when considering how best to support students through classroom instruction.

A number of studies in the United States. settings examine how observation-based measures of classroom instruction relate to student attainment, focusing on student growth over the course of an academic year. Bell et al. (2012[21]) use the Classroom Assessment Scoring System (CLASS) to study how teaching practices and student achievement are related in classrooms teaching algebra, finding that one (classroom organisation) of the three (emotional support, instructional support) domains of practice were significantly related to student test scores, modelled similarly to how the student growth analyses are conducted in this study. Similarly, the multi-state Measures of Effective Teaching (MET) Project (Bill & Melinda Gates Foundation, 2012[22]) calculates correlations between classroom observation scores and test score based "value-added" measures, finding that teachers' total observation scores were not significantly related to value-added measures in either mathematics or English language arts.

While much of the prior literature focused on the relationships between observed teaching practices and test scores, Blazar and Kraft (2017[23]) aggregated data over the course of three years to study how observed measures of mathematics teaching are related to both student test score and dispositional outcomes in nearly 300 elementary classrooms in the United States. The authors use a combination of the CLASS and MQI observation systems to define four domains of mathematics teaching and examine how these measures relate to student performance on both low- and high-stakes assessments of mathematics knowledge, in addition to student-reported dispositional measures of self-efficacy, happiness and behaviour. Similar to the present Study, the authors found that measures of instructional practice were most consistently related to student self-efficacy across the classrooms in their sample, with weaker relationships between practices, happiness and behaviour, and no significant relationships with student performance on a low stakes mathematics assessment.

Several studies support the collection of teachers' teaching materials as a valid process for examining classroom practice and one that can yield information that is distinct from observation-based measures (Borko et al., 2005[24]; Martínez, Borko and Stecher, 2012[25]), though there is a smaller body of literature that directly examines whether these materials successfully predict student achievement scores (Martínez, Borko and Stecher, 2012[25]). To this end, albeit with largely null findings, the present Study contributes substantially to the body of research on teaching materials and student outcomes. Future research on teaching materials should consider the ways teachers interact, modify and implement materials as possible mediating factors in how materials ultimately affect student outcomes.

One study that emphasises the importance of both international and national studies of the relationship between teaching and outcomes is the TALIS-PISA link study (Le Donné, Fraser and Bousquet, 2016[26]). Importantly, the TALIS-PISA link differs from this study in that the study is purely cross-sectional and does not include measures of students' baseline outcomes and it uses teacher reports, rather than video observation and teaching materials, to construct measures of teaching practice (active learning, cognitive activation and teacher-directed instruction). The authors identify significant positive relationships between teachers' reported use of cognitive activation and mathematics achievement in four out of the eight countries included in the Study, but inconsistent significance, magnitude and, for non-test score outcomes, even directionality of relationships between outcomes and teaching practices.

Ultimately, while prior literature in this area speaks to the potential for teaching practices to make a moderate impact on different types of student outcomes, how teaching and learning relate varies greatly across countries and modes of measurement. In finding evidence of prevalent, but not sample-wide, relationships between teaching practices and outcomes, the findings presented in this report emphasise the importance of understanding the unique educational context of different educational settings when developing methods for measuring and improving teaching.

References

Ainley, M., S. Hidi and D. Berndorff (2002), "Interest, learning, and the psychological processes that mediate their relationship", *Journal of Educational Psychology*, Vol. 94/3, pp. 545-561, https://doi.org/10.1037/0022-0663.94.3.545. [11]

Bandura, A. (1977), "Self-efficacy: Toward a unifying theory of behavioral change", *Psychological Review*, http://dx.doi.org/Psychological Review. https://doi.org/10.1037/0033-295X.84.2.191. [15]

Bell, C. et al. (2012), "An Argument Approach to Observation Protocol Validity", *Educational Assessment*, http://dx.doi.org/10.1080/10627197.2012.715014. [21]

Bill & Melinda Gates Foundation (2012), *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. [22]

Blazar, D. and M. Kraft (2017), "Teacher and Teaching Effects on Students' Attitudes and Behaviors", *Educational Evaluation and Policy Analysis*, http://dx.doi.org/10.3102/0162373716670260. [23]

Borko, H. et al. (2005), "Artifact packages for characterizing classroom practice: A pilot study", *Educational Assessment*, http://dx.doi.org/10.1207/s15326977ea1002_1. [24]

Carnoy, M. et al. (2016), "Revisiting the relationship between international assessment outcomes and educational production: Evidence from a longitudinal PISA-TIMSS sample", *American Educational Research Journal*, https://doi.org/10.3102/0002831216653180. [9]

Charalambous, C. and H. Hill (2012), "Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship", *Journal of Curriculum Studies*, http://dx.doi.org/10.1080/00220272.2011.650215. [18]

Chetty, R., J. Friedman and J. Rockoff (2014), "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood", *American Economic Review*, Vol. 104/9, pp. 2633-2679. [1]

Givord, P. (2019), “Does greater social diversity in schools have an impact on equity in learning outcomes?”*, PISA in Focus*, No. 97, OECD Publishing, Paris, https://dx.doi.org/10.1787/35d2ab24-en. [8]

Hanushek, E. and L. Woessmann (2017), “School resources and student achievement: A review of cross-country economic research”, in Rosen, M., K. Hansen and U. Wolff (eds.), *Cognitive abilities and educational outcomes*, Springer, Amsterdam. [6]

Harackiewicz, J. et al. (2002), “Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation”, *Journal of Educational Psychology*, Vol. 94/3, pp. 562–575, https://doi.org/10.1037/0022-0663.94.3.562. [13]

Harackiewicz, J. and C. Hulleman (2010), “The importance of interest: The role of achievement goals and task values in promoting the development of interest”, *Social and Personality Psychology Compass*. [10]

Hidi, S. and K. Renninger (2006), “The four-phase model of interest development”, *Educational Psychologist*, Vol. 41/2, pp. 111-127, http://dx.doi.org/doi:10.1207/s15326985ep4102_4. [12]

Jackson, C. (2018), “What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes”, *Journal of Political Economy*, Vol. 126/5, pp. 2072-2107. [2]

Kalogrides, D., S. Loeb and T. Beteille (2013), “Systematic sorting: Teacher characteristics and class assignments”, *Sociology of Education*, Vol. 86/2, pp. 103-123. [7]

Kraft, M. (2019), “Teacher effects on complex cognitive skills and social-emotional competencies”, *Journal of Human Resources*, Vol. 54/1, pp. 1-36. [3]

Le Donné, N., P. Fraser and G. Bousquet (2016), “Teaching Strategies for Instructional Quality: Insights from the TALIS-PISA Link Data”*, OECD Education Working Papers*, No. 148, OECD Publishing, Paris, https://dx.doi.org/10.1787/5jln1hlsr0lr-en. [26]

Lipowsky, F. et al. (2009), “Quality of geometry instruction and its short-term impact on students’ understanding of the Pythagorean Theorem”, *Learning and Instruction*, http://dx.doi.org/10.1016/j.learninstruc.2008.11.001. [20]

Martínez, J., H. Borko and B. Stecher (2012), “Measuring instructional practice in science using classroom artifacts: Lessons learned from two validation studies”, *Journal of Research in Science Teaching*, Vol. 49/1, http://dx.doi.org/10.1002/tea.20447. [25]

McCaffrey, D. et al. (2003), *Evaluating value-added models for teacher accountability*, RAND, Santa Monica, CA. [4]

OECD (2019), *PISA 2018 Results (Volume III) : What School Life Means for Students’ Lives*, OECD Publishing, Paris, https://doi.org/10.1787/acd78851-en. [16]

Pianta, R. and B. Hamre (2009), “Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity”, *Educational Researcher*, Vol. 38/2, pp. 109-119, http://dx.doi.org/10.3102/0013189X09332374. [5]

Rhew, E. et al. (2018), “The effects of a growth mindset on self-efficacy and motivation”, *Cogent Education*, http://dx.doi.org/10.1080/2331186X.2018.1492337. [17]

Schmidt, W. et al. (2015), “The role of schooling in perpetuating educational inequality”, *Educational Researcher*, Vol. 44/7, pp. 371-386, http://dx.doi.org/10.3102/0013189X15603982. [19]

Sheldon, K. and A. Elliot (1999), “Goal striving, need satisfaction, and longitudinal well-being: The self-concordance model”, *Journal of Personality and Social Psychology*, Vol. 76/3, pp. 482–497, https://doi.org/10.1037/0022-3514.76.3.482. [14]

Annex 7.A. Chapter tables

Annex Table 7.A.1. Relationships between observed teaching practices and post-unit student outcomes

	Test scores						Personal interest in mathematics						Self-efficacy					
	Classroom management		Social-emotional support		Instruction		Classroom management		Social-emotional support		Instruction		Classroom management		Social-emotional support		Instruction	
B-M-V (Chile)	8.43		3.90		19.10	**	0.31	**	0.24	**	0.12		0.24	*	0.11		0.32	*
	(5.52)		(4.37)		(6.44)		(0.11)		(0.08)		(0.11)		(0.11)		(0.08)		(0.13)	
Colombia	0.24		2.10		7.18	**	0.24		0.19	**	0.10		0.10		0.08		0.24	*
	(3.07)		(1.84)		(2.55)		(0.14)		(0.07)		(0.10)		(0.15)		(0.08)		(0.09)	
England (UK)	13.33	**	6.52		8.26		0.34		0.18		0.18		0.21		0.03		0.04	
	(4.39)		(3.55)		(5.08)		(0.18)		(0.10)		(0.10)		(0.18)		(0.09)		(0.12)	
Germany*	6.87		-0.12		0.66		-0.29		0.01		-0.29		-0.04		0.08		-0.06	
	(4.83)		(2.65)		(4.10)		(0.27)		(0.17)		(0.18)		(0.18)		(0.13)		(0.13)	
K-S-T (Japan)	22.41	**	1.48		12.00	***	0.04		0.02		0.14		-0.01		-0.10		0.04	
	(7.01)		(3.93)		(3.15)		(0.24)		(0.11)		(0.09)		(0.19)		(0.09)		(0.07)	
Madrid (Spain)	3.29		1.74		5.75		0.55	*	0.26	**	0.02		0.54	*	0.12		-0.03	
	(4.12)		(1.94)		(3.41)		(0.23)		(0.08)		(0.18)		(0.21)		(0.08)		(0.16)	
Mexico	3.00		4.38	*	3.67	*	0.09		0.06		0.15		0.21	*	0.16	*	0.22	**
	(1.99)		(1.68)		(1.81)		(0.10)		(0.09)		(0.09)		(0.09)		(0.06)		(0.07)	
Shanghai (China)	1.98		12.20		15.92	*	0.49		0.33	***	0.61	***	0.61		0.51	***	0.61	***
	(21.04)		(6.61)		(7.47)		(0.39)		(0.08)		(0.11)		(0.46)		(0.09)		(0.14)	

Notes: * p < . 05, ** p < .01, *** p < .001. Regression coefficients represent the estimated association between a given student outcome and a one-unit increase in the measure of teaching practice. Standard errors are in parentheses below regression coefficients.
*Germany refers to a convenience sample of volunteer schools.
Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188234

Annex Table 7.A.2. Relationships between observed teaching practices and growth in student outcomes, controlling by students' background and pre-unit scores

	Test scores						Personal interest in mathematics						Self-efficacy					
	Classroom management		Social-emotional support		Instruction		Classroom Management		Social-emotional support		Instruction		Classroom management		Social-emotional support		Instruction	
B-M-V (Chile)	0.78		1.65		-1.59		0.29	*	0.23	**	0.07		0.16		0.09		0.08	
	(2.02)		(1.51)		(2.75)		(0.12)		(0.08)		(0.12)		(0.10)		(0.06)		(0.12)	
Colombia	3.58		2.74	*	4.22	**	0.06		0.14	*	0.14		-0.01		0.14	*	0.17	*
	(2.04)		(1.17)		(1.46)		(0.15)		(0.07)		(0.08)		(0.12)		(0.06)		(0.07)	
England (UK)	1.32		1.70		0.18		0.42	*	0.28	**	0.13		0.28	*	0.12		0.07	
	(1.95)		(1.12)		(1.62)		(0.16)		(0.10)		(0.11)		(0.14)		(0.07)		(0.10)	
Germany*	1.63		1.94		2.29		-0.36		0.02		-0.60	***	-0.02		0.14		-0.15	
	(3.21)		(2.06)		(2.55)		(0.19)		(0.14)		(0.16)		(0.16)		(0.12)		(0.13)	
K-S-T (Japan)	3.04		-0.38		1.48		-0.04		0.00		0.06		-0.12		-0.12		-0.08	
	(4.72)		(2.18)		(2.33)		(0.18)		(0.09)		(0.09)		(0.16)		(0.07)		(0.07)	
Madrid (Spain)	-1.13		1.12		1.07		0.56	*	0.25	**	-0.02		0.54	**	0.22	*	-0.12	
	(3.22)		(1.60)		(3.09)		(0.23)		(0.09)		(0.15)		(0.19)		(0.09)		(0.16)	
Mexico	1.15		1.86		1.59		0.12		0.06		0.14		0.24	*	0.15	*	0.21	**
	(1.07)		(1.04)		(1.22)		(0.12)		(0.08)		(0.09)		(0.10)		(0.06)		(0.08)	
Shanghai (China)	15.51		-2.34		-0.16		0.50		0.14		0.39	***	0.59	*	0.23	**	0.26	*
	(12.24)		(3.76)		(4.78)		(0.27)		(0.07)		(0.11)		(0.24)		(0.07)		(0.11)	

Notes: * p < . 05, ** p < .01, *** p < .001. Regression coefficients represent the estimated association between a given student outcome and a one-unit increase in the measure of teaching practice, adjusted for students' pre-unit scores, demographic characteristics and classroom averages of these characteristics. Standard errors are in parentheses below regression coefficients.
*Germany refers to a convenience sample of volunteer schools.
Source: OECD, Global Teaching InSights Database.

StatLink https://doi.org/10.1787/888934188253

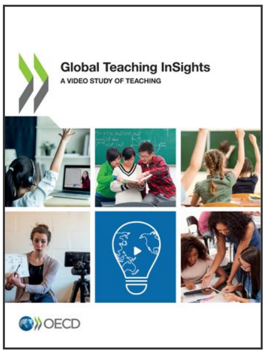
Notes

[1] All regression models use “zero imputation” methods to account for missing values of independent variables and use school clustered standard errors to account for the nesting of students within schools. Readers interested in full model specification should reference the Technical Report or Data User Guide for additional information.

[2] Data for Madrid (Spain) needs to be interpreted with caution. Fielding issues did not allow for a reliable comparison of pre- and post-measures at an individual or classroom level.

[3] Germany* refers to a convenience sample of volunteer schools.

[4] The term “growth model” is not to be confused with a longitudinal growth curve model, and rather, is used only to refer to the fact that students’ pre-unit scores are used as a control variable when estimating the relationship between teaching practices and post-unit scores. Readers interested in full model specification should reference the Technical Report or Data User Guide for additional information.


Global Teaching InSights