



# 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

Yan Zhu<sup>1</sup> · Gabriele Kaiser<sup>2,3</sup>

Accepted: 21 February 2022 / Published online: 24 March 2022  
© The Author(s) 2022

## Abstract

Teaching effectiveness is a core issue in educational research; however, there is little consensus about the most important results of classroom teaching from an international perspective. The effectiveness of teaching has remained a ‘black box’ for a long time. In the secondary study described in this paper we used empirical data for Shanghai taken from the international Organization for Economic Co-operation and Development (OECD) study *Global Teaching InSights (GTI)*—initially the Teaching and Learning International Survey (TALIS) Video Study—which was based on videotaped direct observations of classroom teaching. Eighty-five junior high school mathematics teachers and their students in Shanghai were observed to explore the impact of specific teaching practices on students' interest, self-efficacy, and mathematics achievement scores. The results revealed that social-emotional support and instruction quality were the key dimensions relating to the characteristics and differences of mathematics lessons in Shanghai. While the former had a significantly positive impact on students' general mathematics self-efficacy, the latter had a significantly positive impact on students' mathematics interest. Although specific teaching practices had no significant direct impact on students' mathematics achievement scores, social-emotional support and instruction quality considerably influenced students' academic performance in an indirect way via general self-efficacy.

**Keywords** Video-studies · Classroom observations · Instructional quality · Students' learning gains · Students' mathematics interest · Students' mathematics efficacy

## 1 Introduction and literature survey

Teaching effectiveness has been a core issue in educational research for decades, since it enhances our understanding of how classroom teaching influences students' learning and how classroom teaching can be designed to have a positive impact on learning. Comprehensive research in general education has suggested that teachers are the most important school-related factor affecting students' learning, including

learning of the subject of mathematics, but only a few studies have so far examined how mathematics teachers act in classrooms, how their teaching influences instructional quality, and what constitutes ‘good’ mathematics teaching (e.g., Baumert et al., 2010; Hill & Chin, 2018; König et al., 2021). Building on the pioneering work of the Third International Mathematics and Science Study (TIMSS; Stigler & Hiebert, 1999), carried out more than 20 years ago, the Organization for Economic Co-operation and Development (OECD) conducted the Global Teaching InSights (GTI) study. This ambitious and comprehensive project aimed to improve our understanding of what teaching practices teachers use, how they are interrelated, and which ones most influence students' cognitive and non-cognitive outcomes.

International comparative studies on students' achievements and possible influential factors have a long tradition, especially in mathematics, starting in the 1960s with the First International Mathematics Study (FIMS), which

✉ Gabriele Kaiser  
gabriele.kaiser@uni-hamburg.de

Yan Zhu  
yzhu@kcx.ecnu.edu.cn

<sup>1</sup> East China Normal University, Shanghai, China

<sup>2</sup> University of Hamburg, Hamburg, Germany

<sup>3</sup> Nord University, Bodø, Norway

compared mathematics achievement across 12 countries, revealing strong gender differences and a strong influence of family and affective factors (Husén, 1967). Later, in the 1970s, the comparison went further to collect data on teachers and teaching, primarily based on teachers' self-reports via questionnaires or observational coding schemes (McKnight et al., 1987). However, the instruments used often lacked validity, while purely qualitative approaches based on video analysis were unsuitable as a foundation for identifying general models of classroom teaching and learning. Under such circumstances, the first video survey, integrating videotaping and national sampling, was developed to complement the Third International Mathematics and Science Study (TIMSS), the TIMSS video study. In this study, a representative sample of 231 eighth-grade mathematics lessons from Germany, Japan, and the United States (US) were videotaped. Since the goal of the study was to review the instructional quality of the observed lessons and identify teaching patterns, that means recurring interactional patterns, in the three cultures, three dimensions of the lessons received particular attention in the analysis, as follows: the working environment (number of students in the class, group work or individual learning, access to and use of books and materials, interruptions, etc.); the involvement of students in class (skills, problem solving, level of mathematics, inner coherence, etc.); and the methods the teacher used (structuring the lessons, classwork vs. seatwork, teachers' roles in classes on various occasions, discourse in the class, performance expectations, etc.). As most of the general findings of the study showed, there were significant differences in teaching across the three countries, but differences within countries were smaller than those between countries. Stigler and Hiebert (1999) therefore concluded that teaching is a "cultural activity", which "is learned through informal participation over long periods of time" (p. 86). Stigler and Hiebert (1999) reconstructed culturally shaped teaching "scripts," which rested on beliefs about the nature of the taught subject and how teaching and learning should take place; for example, the German script was characterized by long teacher-guided whole-class activities, and the solution procedure of tasks contrasted strongly with the Japanese script, which focused on individual students' work on problems. The US script was characterized by teachers demonstrating the solution and the students practicing thereafter.

These innovative results prompted a follow-up and expansion of the 1995 video study—the TIMSS 1999 Video Study—which investigated eighth-grade science and mathematics lessons in seven countries (including Japan and the US) that had high-performing students in the TIMSS 1995 assessment. Building on the methods and results of the original study, the mathematics component of the 1999 TIMSS Video Study included the analysis of 638 randomly chosen lessons. The study found similar general features of

mathematics teaching across the seven countries, such as the strong usage of textbooks or a much greater amount of talking by teachers than students. Furthermore, many culturally shaped differences were identified, such as a stronger emphasis on practicing new content in Hong Kong, a higher level of procedural complexity of the problems tackled in Japan, and different types of mathematics problems treated in Japan and Hong Kong compared to the other countries. Overall, the study clearly concluded that no single method of teaching was common to all the participating countries (Hiebert et al., 2003).

The Learner's Perspective Study, established in 1999 (Clarke, 2006; Clarke et al., 2006) as a bottom-up international comparative study, emphasized more strongly the differences between classrooms around the world and students' perspectives, which according to the results of this study was underrepresented in the above-mentioned studies. However, it also emphasized the following:

[How] culturally situated are the practices of classrooms around the world and the extent to which students are collaborators with the teacher, complicit in the development and enactment of patterns of participation that reflect individual, societal and cultural priorities and associated value systems. (Clarke et al., 2006, p. 1)

In this study, in contrast to the TIMSS Video Studies, several East Asian countries or regions participated—Japan, Shanghai (China), Singapore, Hong Kong SAR, and Korea—but the study used convenience sampling and a small dataset, which limited the generalizability of the results (Clarke et al., 2006).

Building upon the two influential TIMSS Video Studies and other existing research, the OECD initiated a new international effort aiming to disclose the characteristics of teaching practices in different cultural systems since 2016, taking into account aspects of teaching that relate to students' learning and development. The city of Shanghai participated in the study, which had high-achieving students according to the 2009 and 2012 cycles of the Program for International Student Assessment (PISA). This new study was initially referred to as the Teaching and Learning International Survey (TALIS) Video Study, but was renamed the Global Teaching InSights (GTI) study. The GTI study aimed to examine the effectiveness of classroom teaching via direct observation, building on TALIS, which asked teachers to report their own teaching in "target classes" via a self-report questionnaire. After a review of existing conceptualizations of teaching quality, including national standards and international research on teaching, three broad dimensions of instructional quality were identified, namely, classroom management, social-emotional support, and instructional practices (Opfer et al., 2020; for the theoretical foundation

of these three dimensions, see Charalambous & Praetorius, 2018; Praetorius et al., 2018). These dimensions then further guided the GTI in formulating a standardized classroom observation protocol consisting of six teaching practices (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e; see details on p. 8).

Research on teachers' effectiveness (Chetty et al., 2014; Hanushek & Rivkin, 2010) has indicated that the impact of teaching quality on students' learning outcomes is multidimensional (Hill & Chin, 2018; König et al., 2021). It is generally believed that highly qualified teachers not only support students in improving their achievement, but also provide instructional contexts that help students to develop their social skills, and help teachers to manage classroom behaviors, deliver accurate subject knowledge, and cultivate critical thinking (Cohen, 2011; Lampert, 2001; Pianta & Hamre, 2009). As for academic achievement, research has also revealed considerable differences between teachers regarding their abilities to affect students' social-emotional development and a variety of in-school behaviors (Backes & Hansen, 2015; Gershenson, 2016; Jackson, 2012; Jennings & DiPrete, 2010; Koedel, 2008; Kraft & Grace, 2016; Ladd & Sorensen, 2015; Ruzek et al., 2015). For instance, Jennings and DiPrete's (2010) investigation of teachers' roles in cultivating kindergarten and grade one students' social and behavioral achievement reported that the effect the teachers had on social and behavioral achievement (0.21 Standard Deviation [SD]) was greater than the effect on students' academic performance (0.12 SD–0.15 SD, depending on grade level and subject content). In a survey of 35 mathematics teachers, Ruzek et al. (2015) found that teachers' influences on students' motivation were small but meaningful (between 0.03 SD and 0.08 SD). Kraft and Grace's (2016) study also revealed that teachers' influence on students' self-reported persistence and students' efforts in the classroom was between 0.14 and 0.17 SD. Other studies have investigated teachers' influences on students' in-school behaviors, including absenteeism, suspensions, grades, promotions, and graduation (e.g., Backes & Hansen, 2015; Gershenson, 2016; Jackson, 2012; Koedel, 2008; Ladd & Sorensen, 2015). Overall, research has indicated that the influence of teachers' instructional skills goes beyond the impact on their students' test scores for academic achievement alone.

To summarize, while the promotion of students' affective and cognitive development is the ultimate goal of education worldwide, the relationship between teachers' competencies and the instructional quality of their teaching activities in the classroom is still under-researched, especially in East Asian countries or regions. To the best of our knowledge, only a few studies worldwide have evaluated instructional quality and related it to teachers' competencies and students' learning outcomes. However, these few known studies—COACTIV (Baumert et al., 2010), TEDS-Instruct/

Validate (König et al., 2021), and Mathematical Knowledge for Teaching (Hill & Chin, 2018; Hill et al., 2005)—have not reviewed teaching and learning patterns, but have evaluated instructional quality at an aggregated level (i.e., focusing on student–teacher interaction at the classroom level). The already described TIMSS Video Study and its 1999 extension provided more detailed analysis and found interesting teaching patterns and cultural differences; however, they did not relate teaching patterns to students' learning outcomes.

## 2 Research questions

Based on the above-described international development studies, the overall question evaluated in this secondary study was how far these originally identified culturally dependent teaching patterns would hold for China Mainland, since China Mainland did not participate in the previous studies. Furthermore, the data from the TIMSS video study are more than two decades old, and the data from the Learners' Perspective Study are more than 15 years old, so they may no longer reflect the current situation worldwide (especially not the situation in China Mainland). To analyze the influence of teachers' activities and their teaching strategies on the affective and cognitive development of their students in China Mainland, the present study is based on a secondary analysis of the just-released GTI data (Retrieved from <https://www.oecd.org/education/school/global-teaching-insights-technical-documents.htm>) to explore in-depth which aspects of teaching related to students' learning and non-cognitive outcomes in mathematics classrooms, focusing on Shanghai as part of China Mainland.

In particular, three research questions were addressed:

1. How can teaching in mathematics classrooms in Shanghai be characterized compared to those in other participating educational systems?
2. To what extent does teachers' classroom teaching relate to students' learning interest, self-efficacy, and mathematical achievements?
3. To what extent does teachers' classroom teaching relate to students' overall learning outcomes?

To answer these three questions, this study used the GTI dataset, which provided video-based classroom observation data on teachers' teaching of quadratic equations as a focal curricular mathematics unit, including the following: students' pre-instruction test scores for general mathematics knowledge and post-instruction test scores for quadratic equation knowledge; various aspects of students' learning before and after the focal curricular unit; and teachers' information about their background and education, beliefs, motivations, and perceptions of the school environment.

Learning achievement in this study included the following three indicators: students' self-reported mathematics interest, their general mathematics self-efficacy, and their scores for mathematics tests. The reason for using the three indicators as student outcome measures is not only that all of them were investigated before and after the videotaped lessons via questionnaires, but also that they are important learning outcomes to which researchers, policy makers, and parents pay considerable attention (Borghans et al., 2008; Chetty et al., 2011; Farrington et al., 2012). While the focus of the present analysis was on Shanghai, the data from the other seven participating countries/economies were used as a reference, particularly for answering the first research question.

### 3 Framework and design of the GTI study and the study relating to Shanghai

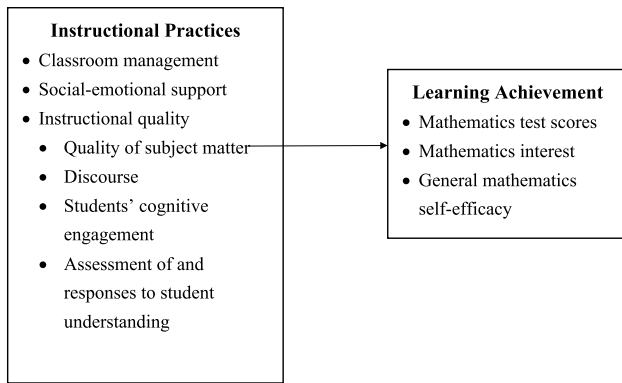
In this section we describe the framework of the GTI study, since this determined the data available and used in our own secondary study; then we present our own secondary analysis.

#### 3.1 Conceptual and analysis framework of the GTI study and its design

Based on national standards and international research on teaching, the GTI study was intended to identify common aspects and patterns of teaching across different countries of reference (Klieme, 2020). The referenced sources mainly included participating countries' or regions' local conceptualizations of teaching, research literature on teaching quality, and the relevant OECD conceptual frameworks from the most recent TALIS 2018 and PISA 2018 studies. Referring to existing research on the effectiveness of teaching and teaching quality, the GTI study enabled the development of a framework consisting of six "domains of teaching practices," which guided the development of the questionnaires, classroom observations, and artifacts. The six domains of teaching practices include classroom management, social-emotional support, quality of subject matter, discourse, students' cognitive engagement, and assessment of and responses to students' understanding (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e; see the Technical Report for additional details about the creation of the analytic domains; for an overview of existing frameworks to which the study refers, see Praetorius & Charalambous, 2018). In particular, *classroom management* refers to the processes that ensure lessons run smoothly and effectively and maximize teachers' and students' in-class time, focusing on academic and social emotional learning (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e). This domain consists of classroom routines, teacher monitoring, and classroom disruptions as the three

main components. *Social-emotional support* refers to a positive learning climate that encourages students to take risks and challenges them at an intellectual and sometimes emotional level, based on respect between teachers and students, encouragement and warmth, and students' willingness to take risks, as its main components (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e). *Quality of subject matter* refers to content and tasks being clear and accurate, and students and teachers being able to make explicit connections between subject content, procedures, viewpoints, and representations or equations that are clear and appropriate, with components including explicit connections as well as explicit patterns and generalizations (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e). *Discourse* refers to extended conversations between and among teachers and students, with students having an adequate amount of talking time, including students engaging in cognitive reasoning on a range of levels, with components concerning the nature of language, questioning, and explanations (Praetorius et al., 2020a, 2020b). *Students' cognitive engagement* refers to students engaging in cognitively rich analysis, creation, or evaluation work that requires thoughtfulness, and it consists of cognitive engagement with demanding subject matter, multiple approaches to/perspectives on reasoning, and understanding of subject matter procedures and processes as the main components (.). *Assessment of and responses to student understanding* refers to teachers aligning instruction with students' thinking to elicit students' understanding, assess it, and provide feedback, with components including eliciting students' thinking, teacher feedback, and instructional alignment with students' thinking (Praetorius et al., 2020a, 2020b). The latter four domains were further grouped in the GTI study into the broad analytic domain of *instruction*. The development and descriptions of the six domains of teaching practice examined in the GTI study can be found in detail in the technical report of the project (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e).

Regarding students' learning outcomes, the GTI study included both cognitive student test results and a set of non-cognitive dispositions. While the latter consists of *interest in mathematics*, *self-efficacy in mathematics*, *self-concept in mathematics*, *instrumental motivation*, *learning goal orientation*, *effort and perseverance*, only the first three were surveyed twice in a pre-post-design (i.e., before and after the teaching of quadratic equations as a focal unit) (Praetorius et al., 2020a, 2020b). Moreover, as self-efficacy is more specific and circumscribed than self-concept, the former is selected as one outcome indicator in this analysis, which is used to reveal the immediate influence of classroom teaching on students' learning. As a result, three learning outcome indicators (i.e., achievement test scores, students' self-reported mathematics interest, and their general mathematics self-efficacy) were included in the analysis within



**Fig. 1** Analytical framework of the relationships between specific classroom instructional practices and learning achievement

our study. Details of the three indicators are described in the explanation of our own study.

The components of the framework of the GTI study connecting specific instructional practices in the classroom and students' learning outcomes are displayed in Fig. 1. Using the Shanghai data from the GTI study, in the present study we explored the direct and indirect relationships between specific instructional practices and students' learning achievement, as well as the corresponding strengths of these relationships.

### 3.2 Our secondary study: data sources and analysis methods

About 700 mathematics teachers and their 17,500 students from the following countries or from regions out of the country, participated in the GTI study: Chile (three cities), Columbia, United Kingdom (England), Germany (seven Federal States), Japan (three regions), Spain (Madrid), Mexico, and China (Shanghai). The Shanghai sample consisted of 85 mathematics teachers and their 2613 students. Of these teachers from Shanghai, about 76.5% were female, with an average length of teaching experience of around 16.2 years. Moreover, about 14.1% of these Shanghai mathematics teachers possessed a master's degree or higher. Quadratic equations was chosen as the focal mathematics topic within the GTI study, since this topic was taught in all participating countries. In Shanghai, quadratic equations are usually taught in the eighth grade; consequently, the average age of the participating students from Shanghai was around 13.6 years, and about 46.6% of them were girls. Using the GTI study's Shanghai data, which are publicly available on the OECD website, this secondary analysis aimed to explore the influence of specific classroom instructional practices on students' learning achievement (i.e. their test scores, mathematics interest, and general mathematics self-efficacy).

In the following, we describe the process of the selection of the GTI variables for the present analysis, which was highly important, especially for the secondary analysis of a large-scale study. Teachers' classroom instructional behaviors were the key research theme for the GTI study. Based on the videotaped teachers' teaching of quadratic equations, the raters evaluated the six dimensions of classroom instructional practices (i.e., classroom management, social-emotional support, quality of subject matter, discourse, students' cognitive engagement, and assessment of and responses to students' understanding), assigning scores for 16 related elements, and these scores were then condensed into three dimensional scores for classroom management (CLASSMAN), social-emotional support (SOCILEMO), and instructional quality (INSTRUCT), respectively. These dimensional scores ranged between 1 and 4, with higher scores indicating teachers' higher achievement levels for the corresponding dimension of instructional practice (Bell et al., 2020a, 2020b, 2020c, 2020d, 2020e).

For students' learning outcomes, the GTI study administered achievement tests and questionnaire surveys to participating students before and after the focal unit instruction. In particular, two weeks before the instruction on quadratic equations, all the participating students sat a 30-item mathematics pretest, which mainly assessed students' general mathematics knowledge. Within two weeks of concluding the focal curricular unit, the students were again assessed using a 25-item mathematics test specifically on quadratic equations (excluding quadratic functions). In the same period, students' mathematics interest and general mathematics self-efficacy were surveyed via pre- and post-instruction questionnaires. Regarding the questionnaire design, the items were phrased in exactly the same way and consisted of three items on mathematics interest (e.g., "After a mathematics class, I was often curious about the next mathematics class") and five items on general mathematics self-efficacy (e.g., "I was confident I could master the mathematical skills being taught"). The only difference between the pre- and post-instruction questionnaires was that the former referred to mathematics teaching and learning in general, while the latter referred to the particular constructs as implemented during the focal unit on quadratic equation (Praetorius et al., 2020a, 2020b). In GTI questionnaires used the terms, "previous" and "current", to differentiate the two conditions. Moreover, the pre-instruction questionnaire asked students for their background information, such as their gender (FEMALE), parents' highest schooling level (PARED), and family possessions. These factors were used in this study as control variables.

Since the GTI study was an international research project, the observation scoring tools, questionnaires, and mathematics achievement tests needed to maintain a high level of consistency across countries and regions to ensure the data

were comparable. Correspondingly, the GTI examined the differential item functioning (DIF) of the achievement test items using a multi-group item response theory (IRT) model and standardized students' raw scores into IRT scores on a 100–300 scale (TEST), with 200 representing the average test score across all countries in the sample with a standard deviation of 25 points. Similarly, students' ratings of home possessions were standardized into IRT scores (HOMEPOS\_IRT), and their personal interest in mathematics (PINT) and general efficacy in mathematics (GENSELFEFF) were calculated as the means of ratings for the corresponding survey items (Doan & Mihaly, 2021).

Concerning the analysis methods, the present study of the Shanghai data from the GTI study first used descriptive analysis (e.g., means and standard deviations) of teachers' scores on various classroom instructional practices to show the overall classroom teaching level of the Shanghai mathematics teachers compared to those in other participating nations/regions. As teaching experience and gender are among the teacher characteristics whose relationship with teaching style has been explored in various studies (e.g., Baleghizadeh & Shakouri, 2014; Cho & Baek, 2019; Feldman, 2007; Shah & Udgaonkar, 2018), the present analysis was intended to investigate the predictive power of the two factors in specifying the quality of teaching in the case of Shanghai; in other words, the aim of this analysis is to reveal the variations of teaching practice among teachers related to their personal characteristics. In particular, *t*-tests were used to investigate whether male and female teachers had significantly different instructional teaching patterns, and correlation analysis was then conducted to identify the possible relationships between specific instructional practices and the length of teachers' teaching experience.

To reveal the relationships between specific classroom teaching practices and students' learning achievements, this secondary analysis first investigated the correlation between the two components. Since the specific classroom teaching practices is evident at the class level and students' learning achievement and personal characteristics are individual, the related analysis had to be hierarchical. Moreover, the influence of specific teaching practices on students' learning in this study could be analyzed only based on students' performance in the post-instruction test, which focused on students' knowledge and understanding of quadratic equations as a focal curricular unit. Consistently, students' mathematics interest and general mathematics self-efficacy in the post-instruction questionnaire related to their current mathematics teachers; therefore, before running the correlation analysis for specific teaching practices and learning outcomes, the three learning achievement indicators in the post-instruction test or post-questionnaire were aggregated to the class level. Thereafter, the relationships between the class averages and teachers' instructional dimensional scores

were examined via Pearson's correlation analysis. Next, with respect to the student-level learning achievement indicators, three two-level path models were constructed for the relationships between each of the indicators and classroom instructional practices. Due to the multidimensional nature of learning, different learning outcomes could also have close relationships; therefore, in this study we established an integrated two-level path model to include mathematics interest, general mathematics self-efficacy, and mathematics test scores, to explore the overall influence of specific classroom teaching practices on students' learning.

## 4 Findings and discussion of the in-depth study in Shanghai

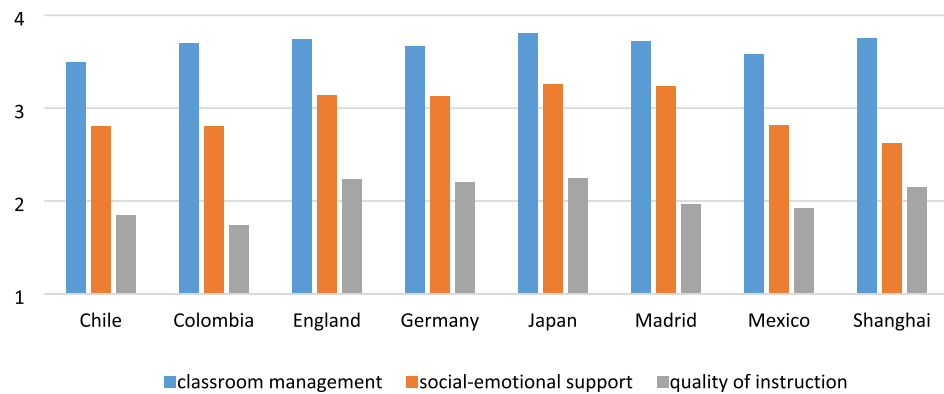
In the following, we present the results of our in-depth study, which was based on a secondary analysis of the Shanghai data from the GTI study.

### 4.1 International comparison of Shanghai classroom instructional practices

To answer the first research question, the three teaching domains from the GTI study were evaluated by comparing the Shanghai teachers' performance with that of the teachers from the other participating countries. Shanghai teachers scored significantly higher for classroom management ( $M=3.75$ ) than social-emotional support and instructional quality. Such a pattern was evident for all eight participating systems (see Fig. 2). The high scores for classroom management in all these systems indicated that the majority of the classroom routines were well organized (Bell, Qi, Witherspoon, Howell, & Barragan, 2020d). The variation in classroom management between Shanghai teachers was found to be at the lowest level ( $SD=0.06$ ), followed by Japan ( $SD=0.12$ ). In fact, nearly all the classrooms from the two East Asian educational systems had a high mean score for classroom management (i.e., above 3.5). A greater variation was observed in the two Central and South American educational systems: Mexico ( $SD=0.23$ ) and Chile ( $SD=0.23$ ). This suggested that Shanghai mathematics classes were more like each other than those in other educational systems in terms of the disciplinary climate.

Across the three main evaluation dimensions, Shanghai teachers scored lowest for instruction quality ( $M=2.15$ ). This low quality of instruction was observed in all eight participating educational systems, and four educational systems, including Shanghai, received a mean score above 2.0 on a four-point scale. Again, Shanghai classrooms had a higher level of similarity than those in other educational systems for the quality of instruction ( $SD=0.18$ ), although in Shanghai, compared to other systems, the related

**Fig. 2** Mathematics teachers' scores for various teaching domains



variations across classrooms (range of 0.18 to 0.27) were larger than those for classroom management (a range of 0.06 to 0.24).

Although Shanghai teachers' mean score for social-emotional support ( $M=2.62$ ) was higher than for their instruction quality, they received the lowest score among the participating educational systems. For this domain, four educational systems received a mean score above 3.0, with Japan having the highest score ( $M=3.26$ ). Compared to the other two teaching domains, the differences across classrooms in all the educational systems were much larger, indicating that in all of them, there were strong and weak classrooms in terms of social-emotional support.

Compared to the high level of classroom management with a small between-class variation, the results for the Shanghai teachers' social-emotional support and instruction quality were relatively low and offer room for improvement.

In order to know how much variation exists within classroom teaching behaviors among Shanghai teachers related to their characteristics, teachers' background characteristics on specific classroom teaching practices were examined. The results highlight that the length of teaching experience did not show a significant correlation with teaching practices ( $p > 0.05$ ), but Shanghai female teachers ( $M=2.18$ ,  $SD=0.18$ ) scored significantly higher than their male colleagues ( $M=2.05$ ,  $SD=0.15$ ),  $t(83)=3.07$ ,  $p=0.003$ ,  $d=0.78$ . Although similar differences were not observed for the other two kinds of teaching practice, a further comparison of social-emotional support between male and female teachers found that the top eight teachers receiving the highest scores and the bottom six teachers receiving the lowest scores were all female. Consistently, the standard deviations for the two gender groups showed a certain discrepancy ( $SD_F=0.24$ ,  $SD_M=0.15$ ). The corresponding differences in classroom management and instruction quality were not obvious, suggesting that there was a larger inter-individual difference in social-emotional support among female teachers. Also, social-emotional support had a negative correlation with the length of female teachers' teaching experience

( $r=-0.22$ ,  $p=0.08$ ), which was not observed for male teachers ( $p=0.94$ ).

An investigation of the correlations among the three classroom instruction practice indices showed that teachers' instruction quality was significantly positively correlated with their social-emotional support and its magnitude reached a moderate level ( $r=0.59$ ,  $p < 0.001$ ), while the correlations between classroom management and instruction quality ( $r=0.09$ ,  $p=0.39$ ), as well as social-emotional support ( $r=0.19$ ,  $p=0.08$ ), were much weaker. This indicated that, in Shanghai, teachers' classroom management had no significant impact on the improvement of instruction quality and social-emotional support, which may relate to the fact that Shanghai classrooms rarely have serious disciplinary problems. Overall, this result suggested that social-emotional support and instructional quality were the key dimensions relating to the characteristics and differences of mathematics classroom teaching in Shanghai.

#### 4.2 Correlations between specific classroom practices and students' learning outcomes

To answer the second research question, the correlations between classroom teaching practices and various students' learning outcomes in Shanghai were analyzed and showed considerable variation. Regarding classroom management, there were no significant correlations with students' mathematics test scores, mathematics interest, and general mathematics self-efficacy in the post-instruction assessment ( $p > 0.05$ ), but the three learning outcomes had significant correlations with social-emotional support and instructional quality. In particular, the correlation between teachers' social-emotional support and students' general mathematics self-efficacy ( $r=0.44$ ,  $p < 0.05$ ), as well as that between teachers' teaching quality and students' mathematics interest ( $r=0.48$ ,  $p < 0.001$ ), were relatively high. Moreover, these two specific classroom practices also had a significantly positive correlation with students' mathematics post-instruction test scores, although the magnitude was weaker than

**Table 1** Correlation analysis of specific classroom teaching practices and students' learning outcomes

	Class average of mathematics post-instruction test scores	Class average of post-instruction mathematics interest	Class average of post-instruction general mathematics self-efficacy
Classroom management	- 0.01	0.19	0.15
Social-emotional support	0.28**	0.32**	0.44**
Instructional quality	0.23*	0.48***	0.41***

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

that with the other two learning outcomes. Comparatively, students' mathematics post-instruction test scores had the strongest correlation with teachers' social-emotional support ( $r = 0.28$ ).

### 4.3 Impact of specific classroom teaching practices on students' learning outcomes

To examine research question 3, we analyzed the correlations between teachers' classroom management and various students' learning and non-cognitive outcomes. Consistent with Table 1, which shows weak correlations between these dimensions, the results of the three-path analyses also showed that, after controlling for students' personal characteristics (including gender, parents' highest education level, and family possession level) and their pre-assessment performance, teachers' classroom management had no significant impact on students' mathematics test scores, mathematics interest, and general mathematics self-efficacy in the post-instruction assessment, and the relationships were positive. By comparison, the correlation between teachers' classroom management and students' mathematics test scores was obviously weaker than that with mathematics interest ( $\beta = 0.003$ ) and general mathematics self-efficacy ( $\beta = 0.22$ ).

The analysis revealed that the three types of classroom teaching practices had no significant impact on students' performance in the post-instruction mathematics test, although all the correlations were positive. The higher standardized coefficient of social-emotional support ( $\beta = 0.16$ ) compared with that of instructional quality ( $\beta = 0.11$ ) suggests a stronger impact of teachers' social-emotional support on students' cognitive test scores than teachers' instructional quality. These dimensions of classroom teaching practices could explain about 5.8% of between-class differences in students' mathematics post-instruction test scores, although the results were not significant.

For students' post-instruction mathematics interest, the standardized coefficient of teachers' scores for instructional quality reached a statistically significant level in a positive direction ( $\beta = 0.52$ ,  $p < 0.001$ ), indicating that the higher the instructional quality of mathematics lessons, the higher the level of students' personal interest in

mathematics. The three classroom teaching practices could explain 26.5% of between-class differences in students' post-instruction mathematics interest, which reached a significant level ( $p = 0.02$ ).

Of the three types of classroom teaching practices, the standardized coefficient of social-emotional support on students' general mathematics self-efficacy reached a statistically significant level ( $\beta = 0.34$ ,  $p < 0.05$ ), and together with the other two practices, they explained about 33.2% of between-class differences for this learning outcome index, which was statistically significant ( $p = 0.02$ ).

When examining the impact of specific classroom teaching practices on students' learning outcomes, in this study we considered the possible early impact of students' backgrounds on learning outcomes (i.e. their performance in the pre-instruction assessment, their gender, parents' highest education level, and family possession level) and performance for the corresponding indices in the pre-instruction assessment. As a result, students' family possession levels and pre-instruction assessment performance showed significant relations with all three learning outcomes, while neither held for parents' highest education level. Moreover, male and female students had highly consistent test scores and mathematics interest, but there was a significant difference in their general mathematics self-efficacy ( $\beta = -0.07$ ,  $p < 0.001$ ), with male students having an overall higher level of general mathematics self-efficacy. This implied that, when helping students improve their self-efficacy via social-emotional support, teachers should pay attention to between-student differences.

### 4.4 An integrated model of the impact of specific teaching practices on learning outcomes

To further examine research question 3, we carried out additional analyses. The multidimensionality of learning was confirmed by the significant correlations between the three learning outcome indicators. In particular, the correlation between self-efficacy and personal interest was high ( $r = 0.65$ ), the correlation between students' post-instruction test scores and general mathematics self-efficacy was



**Table 2** The results of path analysis for students' post-instruction performance and specific classroom teaching practices

	Post-instruction mathematics test scores	Post-instruction mathematics interest	Post-instruction general mathematics self-efficacy
Student level			
Gender (Female) Post-instruction performance	0.02	-0.01	-0.07***
Parents' highest education level Post-instruction performance	0.01	0.01	0.04 <sup>†</sup>
Family possession level Post-unit performance	0.06**	0.14***	0.11***
Pre-instruction performance Post-instruction performance	0.56***	0.43***	0.64***
Teacher level			
Classroom management (Post) performance	0.00	0.18	0.22
Social-emotional support (Post) performance	0.16	-0.06	0.34*
Teaching quality (Post) performance	0.11	0.52***	0.25
Social-emotional support teaching quality	0.59***	0.59***	0.59***
Fit indices			
$\chi^2$ (df)	3.53 (2)	3.25 (2)	3.25 (2)
<i>P</i>	0.17	0.20	0.20
RMSEA	0.02	0.02	0.02
CFI	1.00	1.00	1.00
TLI	0.99	0.99	0.99

Asterisk (\*, \*\*, \*\*\*) indicates the significance level

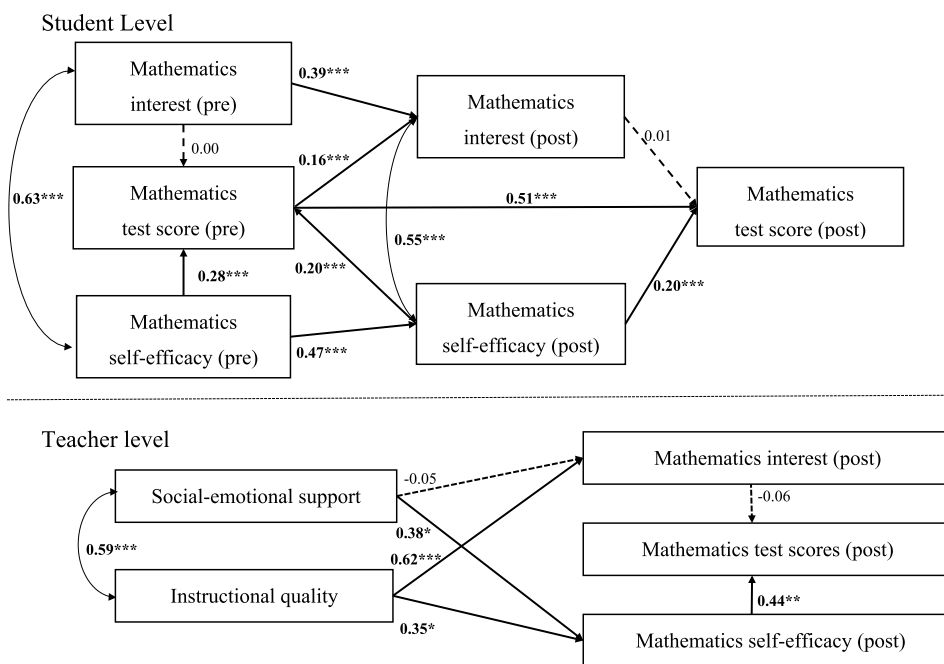
<sup>†</sup>  $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

close to 0.40, and its relationship with post-instruction personal interest was relatively weak<sup>1</sup> ( $r = 0.27$ ); therefore, it was necessary to integrate all these learning outcome indicators into one model to reveal the overall impact of specific classroom teaching practices on students' learning. Due to the weak correlations between teachers' classroom management and various learning outcomes, and the path analysis in Table 2 also showing the insignificant impact of this practice on the three types of learning achievement, in the construction of the integrated model we investigated only the impact of teachers' social-emotional support and instructional quality on students' post-instruction performance, using students' post-instruction mathematics test scores as the dependent variable in the model. Moreover, since classroom teaching practices had no significant impacts on students' post-instruction performance, as shown in Table 2, which concerned the direct effect of teaching practices on learning achievements, the construction of the integrated model also took these into account. Students' pre-instruction performance was also included in the integrated model as one of the controlling conditions (see Fig. 3), since the separate models indicated their important influences on students' post-instruction performance.

The fit indices for the integrated path analysis model,  $\chi^2$  ( $df = 7$ ) = 132.01,  $p < 0.001$ , RMSEA = 0.08, CFI = 0.98, TLI = 0.92, suggested that the model had a good overall fit. The variation in students' post-instruction mathematics test scores explained about 47.9% of the variation at the teacher level, and the corresponding explained proportions for mathematics interest and general mathematics self-efficacy were 44.3% and 36.8%, respectively. All these relations reached a significant level. Consistent with the results in Table 2, instructional quality had a significantly positive impact on students' post-instruction mathematics interest ( $\beta = 0.62$ ,  $p < 0.001$ ) and mathematics self-efficacy ( $\beta = 0.35$ ,  $p = 0.04$ ), while social-emotional support had a significantly positive impact only on students' post-instruction mathematics self-efficacy ( $\beta = 0.38$ ,  $p = 0.02$ ). Furthermore, the analysis revealed that the impact of mathematics interest on achievement test scores was insignificant at both the student and teacher levels as well as in pre- and post-tests, but the impacts of self-efficacy at both levels and tests were significant ( $p < 0.001$ ). All these results suggested that the quality of teacher instruction ( $\beta = 0.17$ ,  $p = 0.06$ ) and social-emotional support ( $\beta = 0.15$ ,  $p = 0.09$ ) could considerably affect students' post-instruction mathematics test scores in an indirect way via their mathematics self-efficacy.

<sup>1</sup> Correlation coefficients of 0.00–0.30 indicate a weak relationship; 0.30–0.70, a moderate relationship; and 0.70–1.00, a strong relationship (Ratner, 2009).

**Fig. 3** The path analysis model for students' (post-instruction) learning achievements and specific classroom teaching practices



## 5 Implications and suggestions for further research

To create a supportive learning environment for students, in which they are treated with respect and dignity, and learn to treat others with respect, it is important to set up classrooms conducive to learning. The GTI study, the data of which were analyzed in this in-depth study, focused on teachers' creation of a positive classroom atmosphere, which could support students' learning and development via social-emotional support; in particular, teachers' social-emotional support was evaluated via the observation of teacher–student respect, teachers' encouragement and warmth, and their support for students having difficulties. The results showed that the Shanghai mathematics teachers had a moderate level of this practice and actually received the lowest score among all the participating educational systems, while teachers in the other East Asian system, Japan, had the highest level of social-emotional support. In fact, about one quarter of the Shanghai mathematics classrooms were at a low level, and the ratios for male and female teachers were similar. A further *t*-test confirmed that there was no significant difference between male and female teachers for this practice indicator. However, the teachers who scored 3 or above were all female, indicating great variation among female teachers for social-emotional support, while male teachers showed greater consistency at a lower level. The study further found that quality of instruction was the weakest instructional dimension across all eight participating systems, and Shanghai showed the smallest between-teacher difference.

Regarding the relationship between classroom atmosphere and students' learning, existing studies have not reached consistent conclusions (Brophy, 1999, 2004). Some researchers pointed out that this might be connected to the different conceptualizations and measurements of atmosphere and teacher–student relationships (Klieme et al., 2009) or the occasional inclusion of classroom management in the related theoretical frameworks. Seidel and Shavelson (2007) claimed that, when discussing the impact of a supportive environment and classroom management on learning, it is necessary to treat both as “remote factors,” since they have almost no direct impact on students' performance. This is consistent with the finding of the present analysis that teachers' social-emotional support and classroom management had no significant impact on students' post-instruction mathematics test scores. However, the results of this study indicated that teachers' social-emotional support had a significantly positive impact on students' post-instruction general mathematics self-efficacy. Moreover, due to the significant impact of the latter on students' post-instruction mathematics test scores, teachers' social-emotional support could affect students' performance in the post-instruction mathematics test indirectly via general mathematics self-efficacy. This unexpected result suggests that establishing classrooms rich in social-emotional support can directly help to improve students' self-efficacy and thus support students' learning of subject matter. Considering Shanghai mathematics teachers' moderate performance on this practice indicator, it would be beneficial to improve their ability to create social-emotional supportive classroom environments and respond better to students' academic and emotional needs.

In terms of students' learning outcomes, the GTI study identified three important indicators, namely, mathematics test scores, mathematics interest, and general mathematics self-efficacy. In this study, none of the three types of teaching practices had a significantly direct impact on students' test scores, while teachers' instructional quality and social-emotional support had a significantly positive impact on students' post-instruction mathematics interest and general mathematics self-efficacy after statistically controlling for students' personal characteristics and their pre-instruction performance. In fact, such an impact was also reported in other researchers' work; for instance, researchers found that students' motivation, attitudes toward school, willingness to do homework, and confidence in their learning behaviors are all affected by teachers' attitudes toward teaching, and such an impact is long-term (Ulug et al., 2011). The authors of this study therefore strongly recommended that teachers should provide support to help students learn, since it seems to be vital to create an atmosphere of positive expectations. Such an atmosphere helps motivate students to study hard and maintain such efforts, and it also facilitates the formation of constructive relationships between teachers and students (Tschannen-Moran & Hoy, 2001).

While it is understandable that the most attention is paid to academic performance, this study revealed that the considerably positive impact of teaching practices on students' academic performance is through their general self-efficacy. Therefore, while paying attention to academic performance, it is particularly important to strengthen students' emotional experiences in classrooms to enhance their learning motivation and other non-cognitive performance. Studies have found that students' learning motivation has a positive impact on their learning (see Lai, 2011; Sekhar et al., 2013; Vu et al., 2021). Some researchers even suggested that for junior high school students, motivation may be the factor most strongly correlated with academic performance (Guthrie & Wigfield, 2000). However, it appears that the role of motivational processes has been ignored or poorly articulated in the norms, beliefs and practices related to mathematics teaching, learning and assessment over the past two centuries (Hannua et al., 2016). Moreover, researchers have almost neglected how mathematics motivation was acquired, consolidated and matured related to instruction with few exceptions (Middleton & Spanias, 1999). Consistently, the seminal work by McLeod (1992) claimed that many research studies on teaching have failed to pay adequate attention to emotional issues. The international large-scale PISA and TIMSS studies have conducted useful explorations in this regard. Both studies have consistently revealed that, although there are great variations in the self-reported learning motivation of students from different countries or regions, a positive correlation between motivation and test scores was observed in all the participating countries

or regions (Mullis et al., 2008; OECD, 2010). The present analysis also revealed that students' motivation significantly and positively correlated with their mathematics test scores, although the corresponding direct impact in the later path analysis did not reach a significant level at either the student or teacher level. Moreover, this study showed that the quality of teachers' instruction plays a positive role in enhancing students' motivation. All these results suggest that the impact of teaching practices is multidimensional.

Since the 1980s, the impact of controllable factors (e.g., teachers and opportunities to learn) on students' learning processes has gradually attracted attention. Researchers have tried to investigate a variety of teachers' behaviors, such as teaching procedures, teachers' guidance, feedback, plans, and preparation, to identify factors that can significantly affect students' academic performance. However, the investigation of classroom instructional practices in such research mostly used self-report questionnaires as the main research method, which limited the studies' accuracy in revealing the impact of actual teaching practices. The present study attempted to analyze the empirical data provided by the OECD's GTI international video research project, which was based on video analysis technology, to examine the actual practices in classroom teaching and the direct and indirect impacts of these practices on students' test scores, learning interest, and self-efficacy. In particular, the GTI identified six different types of classroom teaching practices, and further distinguished these practices from the qualitative and quantitative perspectives in order to determine the components and indicators that can be used for video observation and coding, and then formulated domain scores corresponding to the related practices. Overall, such a highly accepted observation method directly evaluates classroom practices by taking both qualitative and quantitative perspectives into account, which, to a certain extent, overcomes the limitations of teachers' subjective self-reports or comprehensive evaluations conducted by third parties, and may also verify the validity of the observation data at the same time.

In this study, Shanghai teachers' in-class social-emotional support and instructional quality had a significant impact on students' self-efficacy, which further had a significantly positive impact on students' mathematics test scores, with the magnitude of the former being larger than that of the latter. This suggests that it is important to improve the level of teachers' social-emotional support and establish supportive classroom environments that can facilitate students' learning. At the same time, this study revealed that Shanghai teachers still have a lot of room for improvement regarding this particular practice. The GTI also surveyed teachers' confidence and attitudes and found that teachers who were more confident and worked more actively were more likely to obtain higher scores for various practice indicators. For Shanghai teachers, those with high self-efficacy tended to

have higher scores for social-emotional support and instructional quality.

The impact of teachers' emotional health on the effectiveness of classroom teaching is inevitable. Since teaching is a demanding profession, it requires great emotional and physical resilience. For the Shanghai teachers who participated in the GTI, the higher the degrees of anger or anxiety they reported, the lower the scores they received for classroom management, social-emotional support, and instructional quality. Therefore, failure to adequately and appropriately meet teachers' occupational health needs will undoubtedly have a negative impact on teachers' classroom practices as well as their students' learning outcomes.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Backes, B., & Hansen, M. (2015). Teach for America impact estimates on nontested student outcomes (Working Paper 146). Washington, DC: National Center for Analysis of Longitudinal in Education Research. Retrieved on 2021-03-18 from <http://www.caldercenter.org/sites/default/files/WP%20146.pdf>. Accessed 18 Mar 2021.
- Baleghizadeh, S., & Shakouri, M. (2014). The effect of gender and teaching experience on Iranian ESP instructors' teaching styles. *Journal of Education and Human Development*, 3(2), 979–989.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Bell, C., Castellano, K. & Klieme, E. (2020a). Classroom management. In OECD (Ed.), *OECD Global Teaching InSights: A video study of teaching* (pp. 48–80). OECD.
- Bell, C., Klieme, E. & Castellano, K. (2020b). Social-emotional support. In: OECD (Ed.), *OECD Global Teaching InSights: A video study of teaching* (pp. 81–113). OECD.
- Bell, C., Klieme, E., & Praetorius, A.-K. (2020c). Chapter 2: Conceptualising teaching quality into six domains for the study. In OECD (Ed.), *OECD Global Teaching Insights: Technical report*. OECD. <https://www.oecd.org/edu/school/GTI-TechReport-Chapter2.pdf>. Accessed 31 Dec 2021.
- Bell, C., Qi, Y., Witherspoon, M., Howell, H., & Barragan, M. (2020d). The TALIS video study observation system. OECD. [http://www.oecd.org/education/school/TALIS\\_Video\\_Study\\_Observation\\_System.pdf](http://www.oecd.org/education/school/TALIS_Video_Study_Observation_System.pdf). Accessed 31 Dec 2021.
- Bell, C., Schweig, J., Castellano, K., Klieme, E. & Stecher, B. (2020e). Instruction. In OECD (Ed.), *OECD Global Teaching InSights: A video study of teaching* (pp. 114–245). OECD.
- Bell, C., Qi, Y., Witherspoon, M., Howell, H., & Barragan, M. (2018). Annex B: TALIS video observation codes: Indicators. In OECD (Ed.), *OECD Global Teaching Insights: Technical report*. OECD. <https://www.oecd.org/education/school/GTI-TechReport-AnnexB1.pdf>. Accessed 18 Aug 2021.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43, 972–1059.
- Brophy, J. (1999). *Teaching. Educational practices series 1*. UNESCO, International Academy of Education.
- Brophy, J. (2004). *Motivating students to learn* (2nd ed.). Erlbaum Associates.
- Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying mathematics instruction through different lenses: Setting the ground for understanding instructional quality more comprehensively. *ZDM-Mathematics Education*, 50(3), 355–366.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, 126, 1593–1660.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104, 2593–2632.
- Cho, J., & Baek, W. (2019). Identifying factors affecting the quality of teaching in basic science education: Physics, biological sciences, mathematics, and chemistry. *Sustainability*, 11(14), 1–18.
- Clarke, D. (2006). The LPS research design. In D. Clarke, J. Emanuelsson, E. Jablonka, & I. A. C. Mok (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 237–251). Sense.
- Clarke, D., Emanuelsson, J., Jablonka, E., & Mok, I. A. C. (2006). The Learner's Perspective Study and international comparisons of classroom practice. In D. Clarke, J. Emanuelsson, E. Jablonka, & I. A. C. Mok (Eds.), *Making connections: Comparing mathematics classrooms around the world* (pp. 1–22). Sense.
- Clarke, D., Keitel, C., & Shimizu, Y. (Eds.). (2006). *Mathematics classrooms in twelve countries: The insider's perspective*. Sense.
- Cohen, D. K. (2011). *Teaching and its predicaments*. Harvard University Press.
- Doan, S., & Mihaly, K. (2021). Chapter 23: Regression analysis. In OECD (Ed.), *Global Teaching Insights: Technical report*. OECD. Retrieved on 2021-08-18 from <https://www.oecd.org/education/school/GTI-TechReport-Chapter23.pdf>. Accessed 8 Aug 2021.
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners. The role of non-cognitive factors in shaping school performance: A critical literature review*. University of Chicago Consortium on Chicago School Reform.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93–143). Springer.
- Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11, 125–149.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research (Vol. III)* (pp. 403–422). Erlbaum.
- Hannula, M., Martino, P. D., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanim, E., Lutovac, S., Kaasila, R., Middleton, J.,

- Jansen, A. L., & Goldin, G. (2016). Attitudes, beliefs, motivation, and identity in mathematics education: An overview of the field and future directions. In: Attitudes, beliefs, motivation and identity in mathematics education (pp. 1–35). Springer.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100, 267–271.
- Hiebert, J., Gallimore, R., Garnier, H., Bogard Givvin, K., Hollingsworth, H., Jacobs, J., Chui, A. M. Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. (2003). *Highlights from the TIMSS 1999 video study of eighth-grade mathematics teaching*. National Center for Education Statistics.
- Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal*, 55, 1076–1112.
- Hill, H. C., Rowan, B., & Loewenberg Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(1), 371–406.
- Husén, T. (Ed.). (1967). *International study of achievement in mathematics: A comparison of twelve countries Vols I-II*. Almqvist & Wiksell Wiley.
- Jackson, C. K. (2012). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9<sup>th</sup> grade teachers in North Carolina* (NBER Working Paper No. 18624). National Bureau for Economic Research.
- Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education*, 83, 135–159.
- Klieme, E. (2020). Curriculum mapping. In OECD (Ed.), *Global Teaching InSights. Technical report*. OECD.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects on teaching and learning Swizz and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.
- Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics*, 64, 560–572.
- König, J., Blömeke, S., Jentsch, A., Schlesinger, L., Nehls, C., & Kaiser, G. (2021). Teaching and learning in the lower secondary mathematics classroom: Analyzing the links between pedagogical competence, instructional quality, and students' mathematics achievements. *Educational Studies in Mathematics*, 107, 189–212.
- Kraft, M. A., & Grace, S. (2016). Teaching for tomorrow's economy? Teacher effects on complex cognitive skills and social-emotional competencies (Working paper). Brown University. Retrieved on 2021–03–18 from [http://scholar.harvard.edu/files/mkraft/files/teaching\\_for\\_tomorrows\\_economy\\_-\\_final\\_public.pdf](http://scholar.harvard.edu/files/mkraft/files/teaching_for_tomorrows_economy_-_final_public.pdf). Accessed 18 Mar 2021.
- Ladd, H. F., & Sorensen, L. C. (2015). *Returns to teacher experience: Student achievement and motivation in middle school* (Working Paper No. 112). National Center for Analysis of Longitudinal in Education Research. Retrieved on 2021–03–18 from [http://www.caldercenter.org/sites/default/files/WP%20112%20Update\\_0.pdf](http://www.caldercenter.org/sites/default/files/WP%20112%20Update_0.pdf). Accessed 18 Mar 2021.
- Lai, E. R. (2011). *Motivation: A literature review research report*. Pearson Research Report. [http://images.pearsonassessments.com/images/tmrs/Motivation\\_Review\\_final.pdf](http://images.pearsonassessments.com/images/tmrs/Motivation_Review_final.pdf). Accessed 18 Mar 2020.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. Yale University Press.
- McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., & Cooney, T. J. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. STIPES.
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 597–622). Macmillan.
- Middleton, J. A., & Spanias, P. A. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education*, 30(1), 65–88.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). TIMSS 2007 international mathematics report. Retrieved on 2020–10–03 from [http://timssandpirls.bc.edu/TIMSS2007/PDF/TIMSS2007\\_InternationalMathematicsReport.pdf](http://timssandpirls.bc.edu/TIMSS2007/PDF/TIMSS2007_InternationalMathematicsReport.pdf). Accessed 3 Oct 2020.
- Opfer, D., Bell, C., Klieme, E., McCaffrey, D., Schweig, J. & Stecher, B. (2020). Understanding and measuring mathematics teaching practice. In OECD (Ed.), *OECD Global Teaching InSights: A video study of teaching* (pp. 33–47). OECD.
- Organization for Economic Cooperation and Development (OECD) (2010). PISA 2009 results: What students know and can do—Student performance in reading, mathematics and science (Vol. 1). Retrieved on 2018–12–16 from <https://doi.org/10.1787/9789264091450-en>.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109–119.
- Praetorius, A.-K., & Charalambous, C. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM-Mathematics Education*, 50(3), 535–553.
- Praetorius, A.-K., Fischer, J. & Klieme, E. (2020a). Questionnaire development. In: OECD (Ed.), *Global Teaching InSights Technical report*. OECD.
- Praetorius, A.-K., Fischer, J., & Klieme, E. (2020b). Teacher and student questionnaire development. In: Global Teaching InSights: Technical report. Section II: Instrument development. OECD.
- Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM-Mathematics Education*, 50(3), 407–426.
- Ratner, B. (2009). The correlation coefficient: Its values range between +1/-1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*, 17, 139–142.
- Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, 35, 852–882.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling metaanalysis results. *Review of Educational Research*, 77(4), 454–499.
- Sekhar, C., Patwardhan, M., & Singh, R. (2013). A literature review on motivation. *Global Business Perspectives*, 1, 471–487.
- Shah, S. R., & Udgaonkar, S. (2018). Influence of gender and age of teachers on teaching: Students' perspective. *International Journal of Current Microbiology and Applied Science*, 7(1), 2436–2441.
- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap*. Free Press.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805.
- Ulug, M., Ozden, M. S., & Eryilmaz, A. (2011). The effects of teachers' attitudes on students' personality and performance. *Procedia Social and Behavioral Sciences*, 30, 738–742.
- Vu, T. V., Magis-Weinberg, L., Jansen, B., van Atteveldt, N., Janssen, T., Lee, L., van der Maas, H., Raijmakers, M., Sachisthal, M., & Meeter, M. (2021). Motivation-achievement cycles in learning: A literature review and research agenda. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-021-09616-7>