International Comparative Studies in Mathematics: Lessons and Future Directions for Improving Students' Learning

Jinfa Cai, Ida A.C. Mok, Vijay Reddy and Kaye Stacey

Abstract This chapter is based on the Plenary Panel on International Comparative Studies we delivered at the 13th International Congress on Mathematical Education (ICME-13) in 2016. In the past a few decades, international comparative studies have transformed the way we see mathematics education and provide insight for improving student learning in many ways. Out of several possibilities, we selected four lessons we have learned from international comparative studies: (1) examining the dispositions and experiences of mathematically literate students, (2) documenting variation in students' thinking in different cultures, (3) appreciating the varying meanings and functions of common lesson events, and (4) the importance of making global research locally meaningful. Throughout the paper, we point out future directions for research to expand our understanding and build up capacity in international comparative studies.

Keywords International comparative studies · Mathematical thinking · Mathematical literacy · Dispositions · Classroom instruction · Contextual factors · Large-scale studies · Small-scale studies · TIMSS · PISA

This chapter is based on the Plenary Panel at the 13th International Congress on Mathematical Education. In preparation for the Plenary Panel presentation, we

J. Cai (✉)
University of Delaware, Newark, USA
e-mail: jcai@udel.edu

I.A.C. Mok
University of Hong Kong, Pokfulam, Hong Kong

V. Reddy
Human Sciences Research Council, Pretoria, South Africa

K. Stacey
University of Melbourne, Melbourne, Australia

published a topical survey (Cai, Mok, Reddy, & Stacey, 2016) that provides further detail on the issues raised in this chapter. Here, we have summarized four lessons that international comparative studies provide for improving students' learning, and we suggest directions for future work to expand the scope of research and build up capacity in international comparative studies.

In the past several decades many international comparative studies of mathematics have been conducted, first to examine differences in mathematical proficiency and later to examine dispositions among students from different countries and understand the influence of factors such as curriculum, teacher preparation, the nature of classroom instruction, home and school resources, and context, including parental involvement and the organizational structure of education. We use the phrase 'international comparative studies' to refer to studies involving at least two countries (using 'country' loosely to include significant parts of countries), with the intention of making comparisons at the country level. Other names in the literature include cross-national and cross-cultural studies. We include in our definition studies that are small and large, qualitative and quantitative, and initiatives of government or individual researchers. With this definition, we see international comparative studies in mathematics evolving from informal observations to rigorous measurement of the outcomes of schooling, and from the examination of factors that contribute to performance differences to the generation and testing of theories and policies. Current international comparative studies range from small-scale studies involving a few classes with in-depth analyses to large-scale studies like TEDS (M), TIMSS, and PISA that have upwards of half a million participants and multiple measured variables.

International comparative studies in mathematics have provided a large body of knowledge about how students do mathematics in the context of the world's varied educational institutions. In addition, they examine the cultural and educational factors that influence the learning of mathematics and help identify effective aspects of educational practice in homes, classrooms, schools, and school systems. Examining the learning of mathematics in other countries helps researchers, educators, and government policymakers to understand how mathematics is taught by teachers and how it is learned and performed by students in different countries. It also helps them reflect on theories, practices, and organizational support for the teaching and learning of mathematics in their own culture. Stigler, Gallimore, and Hiebert (2000), themselves researchers conducting international studies, explain the value of this research on trends over time and context in a more nuanced way:

> We may be blind to some of the most significant features that characterize teaching in our own culture because we take them for granted as the way things are and ought to be. Cross-cultural comparison is a powerful way to unveil unnoticed but ubiquitous practices. (pp. 86–87)

The highest-profile international comparative studies, such as PISA and TIMSS, have had a significant impact on thinking about education around the world, especially related to the broad characteristics of educational systems and government policy, of which mathematics is just one of several important components. The fundamental purpose of large-scale studies like PISA and TIMSS is to meet

governments' need for objective evidence to monitor educational outcomes, demonstrate possibilities, and assist in developing new policies. There is no sign of a slowing down of international comparative studies either large or small, so the purpose of this paper is to take a step back and reflect on such studies and the lessons we can learn from them.

In this chapter, we discuss four of the many lessons we can learn from international comparative studies for improving students' learning. We chose these four lessons in particular because they represent different styles and strands of work in this area and because they all have the potential to impact students' learning. The first two lessons focus on students' mathematical thinking and achievement. The third lesson focuses on classroom instruction, and the fourth lesson focuses on policy and the effect of contextual factors on learning.

Lesson 1: Promoting Students' Mathematical Literacy

The results of large-scale studies provide many lessons for educational policy related to overall achievement and its links to instruction and student background variables. This section tells just one of the many stories that arise from the PISA 2012 survey: What curriculum, experiences, and dispositions promote mathematical literacy in students? This story shows a side of the PISA survey that is very different from the country rankings that grab newspaper headlines.

Mathematical literacy, the achievement construct measured by PISA, refers to the ability to use mathematical knowledge in situations that are likely to arise in the lives and work of citizens in the modern world. A precise definition is given by the Organisation for Economic Co-operation and Development (OECD, 2013a, p. 25) and discussed by Stacey and Turner (2015a). The 2012 PISA survey examined many aspects of mathematical literacy: the achievement profiles of students across three processes that are involved in exercising mathematical literacy, the learning opportunities that contribute to achievement, in-class experiences and dispositions that influence mathematical literacy, and the effect of classroom experiences with mathematical literacy on more general student attitudes. This section briefly outlines some of the lessons from this work and draws attention to new directions and research questions for mathematics educators.

Country Profiles of the Processes of Mathematical Literacy

Using mathematics to meet a real-world challenge involves three 'processes,' depicted in Fig. 1:

- Formulating situations mathematically (abbreviated to Formulate);
- Employing mathematical concepts, facts, procedures, and reasoning (Employ); and
- Interpreting, applying, and evaluating mathematical outcomes (Interpret).

Fig. 1 The PISA 2012/2015
processes of mathematical
literacy in practice (OECD,
2013b)
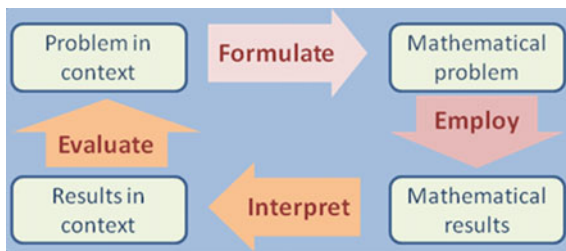
Readers will note the intentional similarity of Fig. 1 to many diagrams depicting
the mathematical modeling cycle. The Formulate process transforms the real-world
challenge into mathematical form by identifying variables and relationships and
making assumptions. The Employ process takes place within the mathematical
world, using the knowledge and skills that form the bulk of school mathematics.
The Interpret process (which, for the purposes of PISA, includes both interpretation
and evaluation of the real-world solution) transforms the mathematical answers
back to the real-world context and judges their real-world adequacy.

PISA 2012 measured the performance of students on each of these three pro-
cesses, revealing, for the first time, interesting country patterns and differences. The
average score for overall mathematical literacy across the OECD was 494, made up
of 492 for Formulate, 493 for Employ, and 497 for Interpret (all standard errors
0.5). Interpret items were the easiest for students, despite the survey design's
intention to select items to measure each process in such a way that the three overall
means would be the same. As with most studies, PISA 2012 showed that boys have
higher mathematics achievement than girls (OECD average gap 11 scale points).
PISA 2012 located the biggest gap between these two groups (OECD average 16
points) to be on the Formulate items. These and other results in this section are
derived from reports from the OECD (2013b, c).

Top-performing countries are generally Asian, and stereotypes might have
predicted their greatest strength to be in routine procedures and hence in the
Employ process. Surprisingly, however, 9 of the 10 top-performing countries'
highest scores were in Formulate. Figure 2 shows this pattern for the
high-achieving country of Japan (mean 536), contrasting with the patterns of rel-
ative scores for the Netherlands and the United Kingdom. Another interesting result
is that the four highest performing countries' lowest scores were in Interpret—the
easiest set of items for the worldwide sample.

Other groups of countries showed consistent but different patterns. The
Netherlands (see Fig. 2), Denmark, and Sweden had their highest scores in both
Formulate and Interpret, the two processes where real-world contexts matter.
Non-Asian English-speaking countries (Canada, Australia, New Zealand, United
Kingdom, United States) were relatively stronger in Interpret only. Nine European
countries scored relatively low in Formulate but higher in both Employ and Interpret.
These newly discovered patterns warrant detailed investigation, especially to
investigate links with curriculum and teaching practices (Stacey & Turner, 2015b).
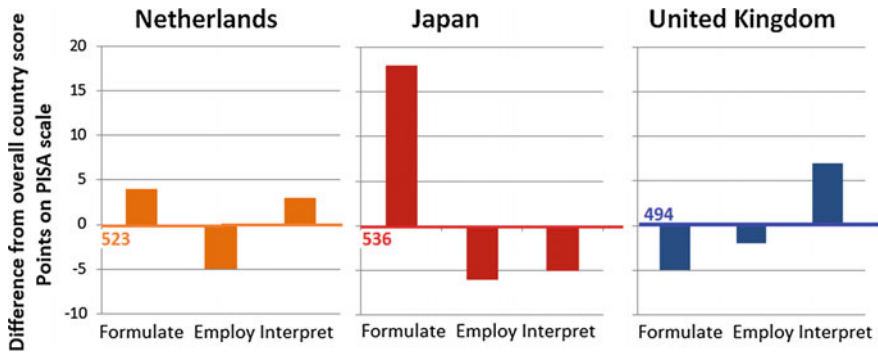
Fig. 2 Sample of PISA 2012 scores on the three processes of mathematical literacy

What Curriculum Experiences Build Mathematical Literacy?

Since PISA's construct of mathematical literacy involves mathematics that is likely to be useful to citizens in all walks of life, it is of interest to know whether a curriculum produces better mathematical literacy outcomes if it is oriented towards abstract mathematics or towards its applications. To answer this question, a sample of PISA students rated how confident they felt about solving a set of mathematics problems and later rated how frequently they had encountered similar problems in class. The sample problems included 'formal' mathematics items lacking any context, such as solving a linear equation or finding the volume of a box, and 'applied' mathematics items, such as using a train timetable and interpreting a newspaper graph. The student ratings were used to create measures of confidence and exposure to applied and formal mathematics[1] (OECD, 2013b, c).

Performance in PISA 2012 was very strongly related to opportunities to learn formal mathematics and secondarily to opportunities to learn applied mathematics. The relationship of PISA performance with exposure to formal mathematics was linear, but quadratic for applied mathematics. The more frequently students are exposed to applied mathematics problems, the better is their PISA performance, but only up to a point—very high exposure is associated with a decline in performance. This may be an outcome of a tendency to place low-performing students in classes with a focus on the 'everyday' applications of mathematics. PISA data reveals this relationship but focused studies are needed to provide a causal explanation.

Japan and the Netherlands, both high-achieving countries, show contrasting patterns of exposure. Students in Japan and other Asian high-performing countries reported low exposure to applied mathematics and high exposure to formal mathematics (OECD, 2013b, c), whereas students in the Netherlands reported high

---

[1]The correct name is "index of experience with pure mathematics," rather than formal mathematics. Confidence is also referred to as self-efficacy. Slightly different constructs in the full reports are conflated here for brevity.

exposure to applied mathematics and low exposure to formal mathematics, perhaps indicating the influence of Realistic Mathematics Education (RME) there. The Netherlands exposure is consistent with the pattern of mathematical process scores shown in Fig. 2, but the Japanese pattern is not. Japanese students perceive an emphasis on formal mathematics but they have nonetheless learned to identify mathematical relationships within real situations and to create appropriate models. How this has happened is an important research question.

Students' Disposition Towards Formal and Applied Mathematics

PISA 2012 also provided some important lessons about student dispositions. Dispositions are especially relevant to the current international governmental climate in which the importance of mathematical literacy to economic well-being is widely acknowledged, with many countries aiming to entice students into STEM careers. Figure 3 shows a strong association between students' reporting of high exposure to a task and their confidence in solving it. Figure 3 also illustrates a general finding that confidence is higher for solving formal mathematics problems than applied problems, at each level of exposure. One explanation is that solving applied mathematics problems requires both a good understanding of the underlying abstract structure as well the ability to analyze the real-world situation—in other words, it requires the PISA mathematical processes of Formulate and Interpret, as well as Employ.

Most countries display a gender difference in confidence in mathematics: PISA 2012 located this difference in the applied problems. Figure 4 compares boys' and girls' reported confidence in solving a sample of applied problems (first six column pairs) and formal problems (last two column pairs; OECD, 2013b). The gender
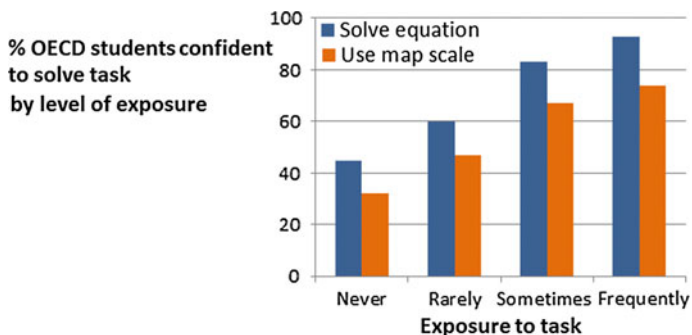

Fig. 3 Percentage of OECD students reporting confidence in solving a formal problem and an applied problem (data from OECD, 2013b)
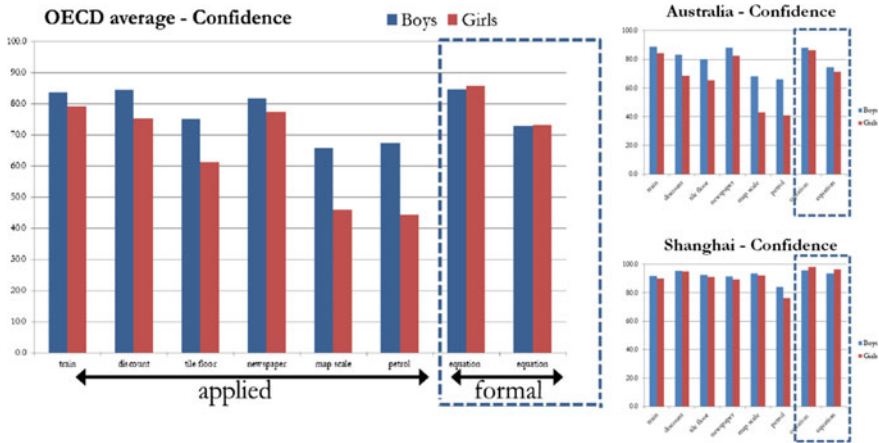
Fig. 4 Confidence of boys and girls in solving eight problems for all OECD countries, Australia, and Shanghai

difference is large for applied problems but is not evident for formal mathematics problems. For example, across OECD countries, 75% of girls reported being confident or very confident when calculating a 30% discount on a TV (second column pair), compared to 84% of boys. The two small graphs on the right side of Fig. 4 show the gender differences for a typical OECD country (Australia) and the lack of gender differences in Shanghai. These gender gaps for applied mathematics problems are likely to have an impact on gender differences in achievement and also on career choices. How can the gender equality of confidence in Shanghai be made a reality everywhere?

PISA 2012 also linked dispositions to exposure to formal and applied mathematics. Overall, students who reported having been more frequently exposed to formal mathematics tasks reported more positive engagement, drive, motivation, and self-beliefs. The same relationship held for applied mathematics tasks, but it became a very strong relationship when controlling for students' achievement. Because of the clear instructional importance, more detailed analyses of the PISA data and further studies are warranted to better understand the links between dispositions, achievement, and exposure to various types of mathematics.

Summary

This section discussed findings from PISA on the curriculum, experiences, and dispositions that promote mathematical literacy in students. These findings illustrate the power of large-scale studies to go well beyond providing country rankings to identify new phenomena worth studying. Better understanding of results such as

these requires both large- and small-scale research, within and between countries, looking at standards, curriculum, teaching, learning, and assessment.

Lesson 2: Understanding Students' Thinking

Over 20 years ago, Bradburn and Gilford (1990) suggested that studies with relatively small, localized samples in a small number of sites can provide useful international comparisons. They can reveal unique findings beyond the scope of large-scale studies and also complement large-scale studies by providing deep understanding about different societies and education systems, thereby enhancing interpretations and implications. Examples of such small-scale studies are Cai (1995, 2000), Cai, Ding, and Wang (2014), Ma (1999), Silver, Leung, and Cai (1995), Song and Ginsburg (1987), and Stevenson et al. (1990).

In Cai et al. (2016), we shared the analysis of two problems to show the value of such in-depth studies. Here, we provide another example from a study by Cai and Hwang (2002), in which they examined Chinese and U.S. sixth graders' mathematical problem solving and problem posing and the relations between them. One pair of tasks is in Fig. 5.

Doorbell (Problem Solving version)

Sally is having a party.
The first time the doorbell rings, 1 guest enters.
The second time the doorbell rings, 3 guests enter.
The third time the doorbell rings, 5 guests enter.
The fourth time the doorbell rings, 7 guests enter.
Keep on going in the same way. On the next ring a group enters that has 2 more persons than the group that entered on the previous ring.

Question 1: How many guests will enter on the 10th ring? Explain how you found your answer.
Question 2: In the space below, write a rule or describe in words how to find the number of guests that entered on each ring.
Question 3: 99 guests entered on one of the rings. What ring was it? Explain or show how you found your answer.

Doorbell (Problem Posing version)

Text as above with Questions 1, 2, and 3 replaced with the following:

For his student's homework, Mr. Johnson wanted to make up three problems BASED ON THE ABOVE SITUATION: an easy problem, a moderate problem, and a difficult problem. These problems can be solved using the information in the situation. Help Mr. Miller make up three problems and write these problems in the space below.

Fig. 5 Problem-solving and problem-posing versions of the Doorbell task

Problem-Solving Results

The U.S. and Chinese students had almost identical success rates (70%) when they were asked to find the number of guests who entered on the 10th ring (Question 1). However, the success rate for Chinese students (43%) was significantly higher than that of the U.S. students (24%) for Question 3 (ring number for 99 guests). The difference is due to their use of different strategies.

Appropriate solution strategies for Questions 1 and 3 were classified into three types: abstract, semi-abstract, and concrete. An abstract strategy generally followed one of two paths: the number of guests who entered on a particular ring of the doorbell is equal to two times that ring number minus one (i.e., $y = 2n - 1$, where y represents the number of guests and n represents the ring number) or the number of guests is equal to the ring number plus the ring number minus one (i.e., $y = n + [n - 1]$). Students used their rule to answer Question 3 (99 guests).

Students who used a semi-abstract strategy made a number of computation steps to yield a correct answer. Students who used a concrete strategy made a table or a list or noticed that each time the doorbell rang two more guests entered than on the previous ring and sequentially added twos to find an answer.

Of the students with appropriate strategies, 44% of the Chinese students and 1% of the U.S. students used abstract strategies for Question 1. For Question 3, 65% of Chinese students used an abstract strategy, compared to only 11% for the U.S. sample. Most U.S. students (75%) chose concrete strategies, compared to 29% of the Chinese students.

Problem-Posing Results

There were similarities and differences in the kinds of problems generated by the two samples. In general, as students in both samples moved towards generating problems of greater difficulty, they tended to move away from posing problems solely about the given information. By far the least common problem types for both groups were those based on reversed thinking (e.g., find ring number given number of guests, as in Question 3, or find total number of rings for a given total number of guests). Chinese students, however, were much more likely to pose problems involving only the given information. U.S. students posed more extension problems than did Chinese students, and a smaller percentage U.S. students (29%) posed no extension problems compared to Chinese students (41%). Similarly, more U.S. students (31%) than Chinese students (21%) posed only extension problems.

The most frequently generated types of problems differed between the two samples. The most frequently generated problems for U.S. students involved finding the number of guests at a particular ring for the easy and moderate problems, and computing the total number of guests after a specific ring for the difficult problem. In contrast, the most frequently generated problems among Chinese

students were non-extension problems (e.g., How many guests entered on the fourth ring?) for the easy problem, and problems asking for the number of guests entering on a ring beyond the fourth ring for moderate and difficult problems.

Summary

Scores arising from large-scale studies are useful for providing an overall picture of students' performance in mathematics and enable rigorous statistical examination of patterns and relationships among variables, including those which may predict students' learning outcomes. However, scoring on the basis of correctness alone conceals some important aspects of students' performance. The results above demonstrate that different students can use different strategies to obtain the same score. Such important differences in students' mathematical thinking may reflect differences in teachers' beliefs and instructional practices (e.g., Cai et al., 2014; Cai & Wang, 2010). In order to provide the education community with a deeper understanding of the teaching and learning of mathematics, it is essential for international comparative studies to provide in-depth evidence of students' thinking and reasoning, including the qualitative analysis of solution strategies, mathematical errors, mathematical justifications, and representations (Cai, 1995).

Lesson 3: Changing Classroom Instruction

Complementary Roles of the TIMSS Video Study and the Learner's Perspective Study

This section draws upon the work of two studies of teaching practice, the TIMSS Video Study and the Learner's Perspective Study (LPS). By zooming in on these two studies, we discuss what we may learn from international comparative studies concerning classroom instruction. The first TIMSS Video Study took place in 1995 (Stigler & Hiebert, 1999) and the over-arching conclusion, reported in *The Teaching Gap* (Stigler & Hiebert, 1999), was that teaching is a cultural activity. The follow-up TIMSS 1999 Video Study (Mathematics) compared teaching practices in the U.S. with six countries that showed higher performance in TIMSS: Australia, the Czech Republic, Japan, the Netherlands, Switzerland, and Hong Kong (Hiebert et al., 2003). Taking the stance that teaching is a cultural activity, the study aimed to build a picture of what typical teaching looked like in different countries and to give researchers and teachers the opportunity to discover alternative ideas about how mathematics might be taught (Stigler & Hiebert, 2004).

LPS (Clarke, Emanuelsson, Jablonka, & Mok, 2006) was designed to examine the practices of eighth grade mathematics classrooms in an integrated, comprehensive way. The project has now developed into a research community in Australia, China, the Czech Republic, Finland, Germany, Israel, Japan, Korea, New Zealand, Norway, the Philippines, Portugal, Singapore, Slovakia, South Africa, Sweden, the United Kingdom, and the U.S. LPS juxtaposes the observable practices of the classroom and meanings attributed to those practices by teachers and students. Instead of aiming for a representative national sample as the TIMSS Video Study did, LPS aimed to understand what might be made possible by competent teachers, locally recognized as such.

Lesson Structures and Lesson Events

The TIMSS Video Study explored lesson structures via the coding of processes like reviewing, demonstrating the problem for the day, practicing and correcting seatwork, and assigning homework (Stigler & Hiebert, 1999), aiming to present a typical "average" lesson for international comparison. LPS used the coding of the TIMSS Video Study to explore patterns of lesson structures of a sequence of consecutive lessons. The findings indicated that the teachers documented in LPS showed little evidence of a consistent lesson pattern, but instead appeared to vary the structure of their lessons purposefully across a topic sequence.

Another viable unit for comparison employed by LPS was the "lesson event," characterized by a combination of form (visual features and social participants) and function, such as intention, action, inferred meaning, and outcome (Clarke et al., 2006; Clarke, Keitel, & Shimizu, 2006). For example, Kikan-Shido (also known as between-desk instruction or seatwork) had a recognizable structural form evident across all classrooms in all countries. However, the findings suggested that the Kikan-Shido lesson events in Shanghai, German, and Japanese lessons had unique emphases:

- Shanghai lessons: correcting errors, encouraging students to think further (Lopez-Real, Mok, Leung, & Marton, 2004)
- German lessons: questioning to stimulate student mathematical thought (Clarke et al., 2006)
- Japanese lessons: eliciting students' mistakes, their puzzlement, and their opposing solutions; pointing out different solutions or difficulties and giving explanations; and making their way of thinking visible to the group (Hino, 2006)

Overall, the findings from the LPS study suggested reasons additional to those identified in the TIMSS Video Study about why the enactment of Japanese lessons differed from other countries (Mok, 2015).

Multiple Accounts of a Teacher's Practice

Another advantage of the LPS data set is that it allows researchers to reconstruct multiple accounts of classroom scenarios by combining data from all of the lesson materials, including videos, student interviews, and teacher interviews, thereby providing the opportunity to study the practice of a particular teacher in a specific cultural system in depth. For example, an explanation has been sought for the "Asian Learner's Paradox," which refers to the seemingly contradictory phenomenon of outstanding student performance in Asian regions but reports of classroom environments being non-conducive to learning, with characteristics such as directive teaching and large classes (Watkins & Biggs, 2001). Mok (2006) analyzed the LPS data of a Shanghai teacher. To illustrate the teacher's skillfulness, a lesson episode about the train-ticket problem is depicted in Fig. 6.

A student, Dora, who first solved the problem mentally, was invited to share her solution with the class. Dora's answer was arithmetic and intuitive in nature, and was immediately followed by the teacher's paraphrasing with an emphasis on the idea of subtraction. Following this, the teacher asked the class to do the problem again using equations, writing the Equations $3x + y = 560$ and $3x + 2y = 640$ and obtaining the answer by subtracting one equation from the other. Mok's (2006) analysis showed that the teacher had created three levels of contrasts to support a deep understanding of the problem. The first level of contrast is between Dora's answer and the teacher's paraphrase, the second level between the arithmetic method and the equation method, and the third level between the equation-solving methods of subtracting equations (elimination) and substitution. Mok argued that the lesson was by no means spontaneous, but rather represented a synthesis based on that experienced teacher's understanding of a pedagogical framework of variation that was well established in his region (Experimenting Group of Teaching Reform in Mathematics in Qingpu County, Shanghai, 1991). The strong teacher guidance in the lesson arose from the teacher's interpretation of student-centeredness, which was different from its interpretation in Western education communities. The teacher saw himself as non-traditional and made use of his understanding of his students in order to create a planned experience for them with

Train Ticket Problem

Xiu-min and his family went to Beijing for a holiday. They booked 3 adult tickets and 1 student ticket, costing a total of 560 dollars. His classmate Xiu-wang, learning this, decided to join Xiu-ming's family for the trip. Consequently, they bought 3 adult tickets and 2 student tickets, costing a total of 640 dollars. Please calculate the cost of 1 adult ticket and the cost of 1 student ticket.

Fig. 6 Train ticket problem

minimal side-tracking (Mok, 2006). The conceptions of this teacher and his performance in the lesson were quite consistent with the findings of another study that compared conceptions of effective teaching between Chinese and U.S. teachers. Cai and Wang (2010) suggested that the constraints of content coverage, teaching pace, and large class size affected teaching flexibility and student-centeredness.

Lessons for the Implementation of Mathematical Tasks

Both the TIMSS Video Study and LPS classified mathematical problems as "using procedure" problems (success requiring only a memorized procedure or algorithm) and "making connections" problems (success requiring the establishment of relationships between ideas, facts, and procedures and engagement in mathematical reasoning). The TIMSS Video Study showed that all of the countries except Japan used more "using procedure" problems than "making connections" problems. In this way, the U.S. was not different from higher-achieving countries in the kinds of problems that teachers *presented to students*. What or where was the difference? The videos of each country revealed some interesting cultural activities. For example, lessons in the Netherlands frequently used calculators and real-world problem scenarios, and Japanese students spent on average a longer time working to develop their own solution procedures for problems that they had not seen before. In all of the high-performing countries except Australia the teachers implemented a higher percentage of "making connections" problems as "making connections" problems than did U.S. teachers. In contrast, U.S. teachers changed "making connections" problems to "using procedures" problems, thereby lowering the cognitive demand of the problems (Roth & Givvin, 2008; Stigler & Hiebert, 2004).

LPS team members have also made some significant achievements in studying the use of mathematical tasks in classroom instruction (Shimizu, Kaur, Huang, & Clarke, 2010). For example, Huang and Cai (2010) found that LPS teachers from the US and China were willing to implement cognitively demanding tasks in their lessons, yet the Chinese teachers were more frequently able to sustain the cognitive demand of the mathematical tasks during implementation. Mesiti and Clarke (2010) analysed the mathematical tasks in the LPS data from China, Japan, and Sweden and concluded that the classroom performance of a task was ultimately a unique synthesis of task, teacher, students, and situation.

Summary

To conclude, the two international comparative studies discussed in this section played complementary roles in contributing to the understanding of classroom instruction. The TIMSS Video Study, building upon the tradition of large-scale surveys of national samples, suggested seeing teaching as a cultural activity. LPS compared mathematics lessons through analysis of lesson events during a sequence

of lessons and included the perspectives of the teacher and the learners. Although teachers in different cultural systems spent time on the same lesson event, they might in fact have been carrying out the activities with different meanings and functions. The attempt to explain the Asian Learner's Paradox is an example of how the investigation of an effective case might take into account many constraints (such as examination orientation, content coverage, teaching pace, and large class size in a specific cultural system) and culturally-rooted clues (such as the teacher's conceptions and beliefs, students' expectations, the locally-implemented pedagogical framework). Lastly, seeking a common language for comparison has a specific implication for understanding effective instruction in different cultures. Both the TIMSS Video Study and LPS have chosen tasks as a theme for comparison. Different kinds of tasks play different roles in the agenda of effective classroom instruction; nonetheless, how the teacher sustains the intended roles of the tasks during implementation is important.

Lesson 4: Making Global Research Locally Meaningful— TIMSS in South Africa

This lesson illuminates how a country can find its own voice in using international comparative studies to extend to analyses that are meaningful for the local agenda. South Africa is characterised as a country with high levels of poverty, inequality, and unemployment. These characteristics have an impact on the quality of education and become both determinants and outcomes of the level of development of the country.

As expected in unequal societies, there are high levels of variation between schools. While many countries focus on interventions inside classrooms to improve subject matter knowledge and achievement scores, low-income countries have to focus on two challenges. On the one hand they have to focus on what happens inside the classroom to improve teachers' and students' mathematical knowledge. On the other hand they must identify the effects of the many contextual factors and conditions that influence educational achievement. In this section we share experiences of using the TIMSS achievement data sets and information on South Africa to inform educational policy.

Mathematics Achievement Trends Over 20 Years

Participation in TIMSS 1995 provided the first indicative estimate of national mathematics and science achievement for South Africa. This was followed by the widely publicized results for TIMSS 1999, which lamented the low South African scores and the rank order which placed South Africa last in the set of 38

participating countries. This international comparison catalysed a debate about educational performance in South Africa and involved many sectors of society—politicians, policymakers, academics, teachers, and the public. Newspaper headlines in South Africa asserted, for example, that 'South African pupils are the dunces of Africa' (Sunday Times, 16 June 2000) or that South African students were the 'Bottom of the class in maths' (Sunday Times, 14 October 2001). Low mathematics performance and country rank were repeated again in TIMSS 2003. The newspaper headlines and reaction from politicians and policymakers echoed those following TIMSS 1999, but the challenge for research was to embark on deeper analysis and extend the story to one which could provide policy directions.

An important but overlooked finding from the TIMSS analysis was the range of performance between the 5th and 95th percentiles of performance. Of all the countries participating in TIMSS 2003, South Africa had the widest range of scores between these two percentiles. This wide range led to the characterization that there were two systems of education in the country and that the performance scores in TIMSS were reflective of wide disparities in society and in schools.

The story of South African performance cannot be told through a single national score but through appropriate disaggregation. The disaggregation of the achievement scores revealed a strong correlation between socioeconomic status and achievement scores. Africans, who were most disadvantaged by the apartheid policies, had the lowest performance. African schools are located in areas where most Africans live and these areas have high levels of poverty and unemployment.

South Africa's participation in TIMSS 2011 provided an opportunity to measure the changes in educational performance since 1995. TIMSS was the only study that provided a scientifically rigorous methodology to measure trends over the previous 20 years. Analysis of the four rounds of TIMSS participation showed that the average national mathematics score remained the same over the years 1995, 1999, and 2003 (Reddy, Van der Berg, Janse van Rensburg, & Taylor, 2012). In contrast, from 2003 to 2011 the national average mathematics score increased by 63 points (see Fig. 7). The increases over the last two cycles of TIMSS can be translated to say that overall student performance, though still low, has improved by one and a

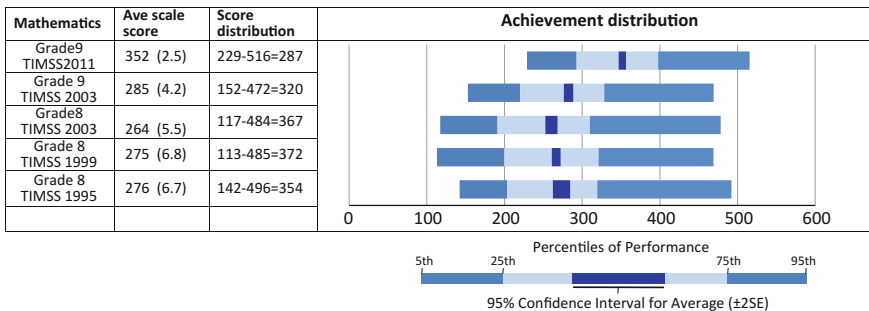| Mathematics	Ave scale score	Score distribution	Achievement distribution
Grade9 TIMSS2011	352 (2.5)	229-516=287	
Grade 9 TIMSS 2003	285 (4.2)	152-472=320	
Grade8 TIMSS 2003	264 (5.5)	117-484=367	
Grade 8 TIMSS 1999	275 (6.8)	113-485=372	
Grade 8 TIMSS 1995	276 (6.7)	142-496=354	

Fig. 7 Trends in mathematics achievement for TIMSS 1995, 1999, 2003, and 2011

half grade levels. In 2011, the range of mathematics scores decreased, suggesting that the country is progressing (albeit slowly) towards more equitable educational outcomes.

Contextual Factors Influencing Educational Achievement

We need to go beyond the achievement scores to investigate the factors that influence mathematical performance. The results of our analyses confirmed the effects of home and school socioeconomic factors. As expected, students who speak the language of the test at home are more likely to achieve higher scores than those who do not.

We explored the effects of two contemporary South African factors on achievement—gender and school violence—and found new complexities in the schooling experience of South African boys and girls. On average, across South Africa, gender differences in mathematics scores were small or non-existent. We also probed students about their attitudes towards mathematics and found that mathematics mattered to both boys and girls. A particularly worrisome finding was the level of indifference among boys about their education. Boys were found to have lower aspirations about their academic careers, showed less interest in mathematics, and engaged less often with an adult regarding their school work. The link between negative attitudes and weak performance was stronger for boys than for girls.

The second factor we explored was the extent of violence in South African schools and its effect on mathematics achievement. Although concerns about school safety are increasing internationally, violence in schools is considered more serious in South Africa than elsewhere. The degree of school safety largely depends on the type of school that learners attend. We found that children attending public schools experienced more frequent threats of violence than children attending independent schools. The socioeconomic status of students is an indicator for potential exposure to acts of violence, with higher chances of being bullied regularly for students from poor families. There is a higher frequency of bullying for boys than for girls who attend schools with similar characteristics. Schools where there are fewer discipline or safety problems achieve better results, but this relationship is dependent on the size of the school.

Student Progression and Pathways Through Secondary School

In addition to concerns about low mathematics achievement, there is also concern about progression through secondary schools. We analysed the pathways and performances in mathematics of secondary school students in South Africa using a

Table 1 Educational pathways of students in the South African Youth Panel Study

Smooth	Staggered	Stuck	Stopped
Neat, year-on-year grade progression through school	Learners in school for all 4 years of SAYPS, who make some grade progress but have at least one episode of grade repetition	Learners in school for all 4 years of SAYPS, but stuck in grade 9 or 10 for three or more periods	Individuals who leave school before final data collection and do not return

panel-like data set of Grade 8 students who participated in TIMSS 2003 and were tracked to Grade 12 examination data sets. Firstly, students who began with similar Grade 8 mathematics scores had different educational outcomes 4 years later. Secondly, in middle class schools, Grade 8 mathematics scores were a good indicator of who would pass the exit level examination in Grade 12, but this relationship was not as strong in schools for poorer students. Thirdly, there was a stronger association between TIMSS Grade 8 mathematics scores and subject choice of secondary school mathematics in middle class schools than in poorer schools. Fourthly, there was a strong correlation between mathematics performance at Grade 8 and the exit level examination. Overall, this study adds to the body of evidence that suggests that to improve educational outcomes, the policy priority should be to build foundational knowledge and skills in numeracy.

To extend our understanding of the pathways and transitions followed by South African youth, the longitudinal South African Youth Panel Study (SAYPS) was initiated, with the first annual data collection wave in 2011. SAYPS followed Grade 9 learners who participated in TIMSS 2011 for 4 years to explore their educational transitions. We found that students followed one of four educational pathways (Table 1) through secondary school.

Almost half of the sample (47%) followed the smooth pathway while 39% followed a staggered pathway and 14% were either stuck or stopped. There is a predictable story of 'advantage begetting advantage' for students who experience a smooth pathway: With higher than average TIMSS scores and better-educated parents, these students come from homes with more books and have positive attitudes about school. Our analyses show that it is possible to succeed academically despite disadvantage: Just over 43% of the smooth group come from non-fee-paying schools for poorer students. We will study this group further to understand their pathways to success.

Future Directions for Learning from International Comparative Studies

Over the last 3 decades, international comparative studies have transformed the way we see mathematics learning and teaching and the four lessons above have illustrated

that there is still much to learn. Looking to the future, we believe it is important to extend international comparative studies to deepen our understanding of previous findings as well as to build the capacity of researchers to implement them.

Improving Our Understanding of the Outcomes of Large-Scale Studies

Because large-scale studies are generally supported by governments with the intention of assisting in policy development, it is important that the outcomes of these studies are understood as deeply as possible. This often requires further research, sometimes within and sometimes between countries. The case study of South Africa provided an excellent example of how further research within one country using trend data can make the results of an international comparative study more useful for local policy development, as well as contribute to the knowledge base for similar countries. For an example where research within countries and between countries may be useful, let us return to PISA's three mathematical processes of Formulate, Employ, and Interpret and the observation in Lesson 1 that groups of countries (such as the high-performing Asian countries or English-speaking Western countries) exhibit different patterns of (relative) perfomance on the three processes. In-depth analysis of the large-scale data can identify such subtle but important differences; however, we need a range of additional studies to explain the findings. Such studies may, for example, examine the construction of the PISA instruments for anomalies, or conduct local or international comparative studies of students' problem-solving processes and/or curriculum experiences. Large-scale studies are very expensive; we need to use the data they provide towards maximum benefit in understanding why students perform as they do.

Investigating New Questions Through Small-Scale Studies

In-depth, small-scale international comparative studies can provide unique opportunities for us to understand students' mathematical thinking. The more information teachers have about what students know and think, the more opportunities they can create for student success. Teachers' knowledge of students' thinking has a substantial impact on their classroom instruction and hence upon students' learning. Thus, small-scale comparison studies provide insights on students' learning and understanding in the context of different cultural systems and at an enactment level of the teaching practice and students' learning. These insights are important to policymakers, researchers, educators, and teachers.

Small-scale international comparative studies can also start to explore many urgent and important research questions. For example, is there really a creativity gap between students in Asian and Western countries and if so, why? Future international comparative studies and international collaboration should answer these questions empirically, building up to large-scale studies.

Building the Capacity of Researchers

There are distinct advantages for individual researchers to collaborate on in-depth, small-scale international comparative studies, because relatively modest resources are required. LPS is an excellent example of a long-standing collaboration which has capitalized on shared interests with a fluid structure within which many people can work together. The recent rapid increase of international comparative studies on curriculum is another example (Lloyd, Cai, & Tarr, 2017). Many individual researchers chose to focus on certain aspects of curriculum as they conducted comparative analyses across nations. These studies provided new insights into the content and design of mathematics textbooks and generated key questions about relationships between written curricular materials and students' opportunities to learn. Another avenue for individual researchers is to engage in secondary analyses of large-scale international comparative studies, which generally make a great deal of data publically available. This work can also be done with few resources and the often severe time constraints that apply to individual or beginning researchers.

Our overall message is that international comparative studies can provide a wealth of information for mathematics education researchers and policymakers. The mathematics education community has a unique capacity to contribute to an in-depth understanding of both national and international findings, and hence to assist us all in learning the right lessons from international comparisons.

References

Bradburn, M. B., & Gilford, D. M. (1990). *A framework and principles for international comparative studies in education*. Washington, DC: National Academic Press.

Cai, J. (1995). A cognitive analysis of US and Chinese students' mathematical performance on tasks involving computation, simple problem solving, and complex problem solving. *Journal for Research in Mathematics Education Monograph Series 7*. Reston, VA: National Council of Teachers of Mathematics.

Cai, J. (2000). Mathematical thinking involved in US and Chinese students' solving process-constrained and process-open problems. *Mathematical Thinking and Learning, 2*, 309–340.

Cai, J., Ding, M., & Wang, T. (2014). How do exemplary Chinese and US mathematics teachers view instructional coherence? *Educational Studies in Mathematics, 85*(2), 265–280.

Cai, J., & Hwang, S. (2002). Generalized and generative thinking in US and Chinese students' mathematical problem solving and problem posing. *Journal of Mathematical Behavior, 21*(4), 401–421.

Cai, J., Mok, I., Reddy, V., & Stacey, K. (2016). *International comparative studies in mathematics: Lessons for improving students' learning*. New York, NY: Springer.

Cai, J., & Wang, T. (2010). Conceptions of effective mathematics teaching within a cultural context: Perspectives of teachers from China and the United States. *Journal of Mathematics Teacher Education, 13*(3), 265–287.

Clarke, D., Emanuelsson, J., Jablonka, E., & Mok, I. A. C. (Eds.). (2006). *Making connections: Comparing mathematics classrooms around the world.* Rotterdam, The Netherlands: Sense.

Clarke, D., Keitel, C., & Shimizu, Y. (Eds.). (2006). *Mathematics classrooms in 12 countries: The insiders' perspective.* Rotterdam, The Netherlands: Sense.

Experimenting Group of Teaching Reform in Mathematics in Qingpu County, Shanghai. (1991). *Xuehui jiaoxue [Learning to teach]*. Beijing, China: People Education Publishers.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study.* Washington, DC: US Department of Education, National Center for Education Statistics.

Hino, K. (2006). The role of seatwork in three Japanese classrooms. In D. Clarke, C. Keitel, & Y. Shimizu (Eds.), *Mathematics classrooms in 12 countries: The insiders' perspective* (pp. 59–74). Rotterdam, The Netherlands: Sense.

Huang, R., & Cai, J. (2010). Implementing mathematical tasks in USA and Chinese Classrooms. In Y. Shimizu, B. Kaur, R. Huang, & D. Clarke (Eds.), *Mathematical tasks in classrooms around the world* (pp. 145–164). Rotterdam, The Netherlands: Sense.

Lloyd, G. M., Cai, J., & Tarr, J. E. (2017). Issues in curriculum studies: Evidence-based insights and future directions. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 824–852). Reston, VA: National Council of Teachers of Mathematics.

Lopez-Real, F. J., Mok, I. A. C., Leung, F. K. S., & Marton, F. (2004). Identifying a pattern of teaching: An analysis of a Shanghai teacher's lessons. In L. Fan, N.-Y. Wong, J. Cai, & S. Li (Eds.), *How Chinese learn mathematics: Perspectives from insiders* (pp. 382–412). Singapore: World Scientific Publishing Co.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Hillsdale, NJ: Erlbaum.

Mesiti, C., & Clarke, D. (2010). A functional analysis of mathematical tasks in China, Japan, Sweden, Australia and the USA: Voice and Agency. In Y. Shimizu, B. Kaur, R. Huang, & D. Clarke (Eds.), *Mathematical tasks in classrooms around the world* (pp. 145–164). Rotterdam, The Netherlands: Sense.

Mok, I. A. C. (2006). Shedding light on the East Asian Learner Paradox: Reconstructing student-centeredness in a Shanghai classroom. *Asia Pacific Journal of Education, 26*(2), 131–142.

Mok, I. A. C. (2015). Research on mathematics classroom practice: An international perspective. In J. C. Sung (Ed.), *Selected regular lectures from the 12th International Congress on Mathematical Education* (pp. 589–606). New York, NY: Springer.

Organisation for Economic Co-operation and Development. (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2013b). *PISA 2012 results: What students know and can do* (Vol. 1). Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2013c). *PISA 2012 results: Ready to learn students' engagement, drive and self-beliefs* (Vol. III). Paris, France: OECD Publishing.

Reddy, V., Van der Berg, S., Janse van Rensburg, D., & Taylor, S. (2012). Educational outcomes: Pathways and performance in South African high schools. *South African Journal of Science*, *108*(3/4).

Roth, K., & Givvin, K. B. (2008). Implications for math and science instruction from the TIMSS 1999 Video Study. *Principal Leadership, 8*(9), 22–27.

Shimizu, Y., Kaur, B., Huang, R., & Clarke, D. (Eds.), (2010). *Mathematical tasks in classrooms around the world*. Rotterdam, The Netherlands: Sense.

Silver, E. A., Leung, S. S., & Cai, J. (1995). Generating multiple solutions for a problem: A comparison of the responses of US and Japanese students. *Educational Studies in Mathematics, 28*(1), 35–54.

Song, M. J., & Ginsburg, H. P. (1987). The development of informal and formal mathematical thinking in Korean and US children. *Child Development, 58,* 1286–1296.

Stacey, K., & Turner, R. (2015a). *Assessing mathematical literacy: The PISA experience*. Heidelberg, Germany: Springer.

Stacey, K., & Turner, R. (2015b). PISA's reporting of mathematical processes. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of 39th psychology of mathematics education conference* (Vol. 4, pp. 201–208). Hobart, Australia: PME.

Stevenson, H. W., Lee, S., Chen, C., Lummis, M., Stigler, J. W., Liu, F., et al. (1990). Mathematics achievement of children in China and the United States. *Child Development, 61,* 1053–1066.

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

Stigler, J. W., & Hiebert, J. (1999). *"The teaching gap." Best ideas from the world's teachers for improving education in the classroom.* New York, NY: The Free Press.

Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership, 61* (5), 12–17.

Watkins, D. A., & Biggs, J. B. (Eds.). (2001). *Teaching the Chinese learner*. Hong Kong: Comparative Education Research Centre, The University of Hong Kong.