

Chapter 4

Measuring Teacher Effectiveness Across Time: What Does TIMSS Reveal About Education System Level Trends?



Abstract To understand whether there are identifiable trends in teacher quality and instructional metrics over time, it is important to first recognize that measures of teacher effectiveness may vary by education system. Descriptive results for the teacher quality measures related to mathematics drawn from twenty years of the IEA's Trends in International Mathematics and Science Study (TIMSS) data provide a general picture of the variation in such measures; additional consideration is given to student-level variables. Combining information from multiple cycles of TIMSS, with a specific emphasis on teacher variables, greater focus is given to the post-2003 TIMSS data, largely because these later cycles contain additional teacher-related measures (such as preparedness to teach mathematics topics). Changes in education system means may be due to changes in underlying teacher characteristics. The data suggests that, among the pool of countries that participated in all the TIMSS between 2003 and 2015, there have been increases in the international averages for the number of years of experience that teachers possess, their level of educational qualification to teach mathematics, and classroom time spent on mathematics instruction. However, instructional alignment with national standards has remained surprisingly modest (at around 55–60%) and the international average for instructional alignment declined for grade eight over twenty years of TIMSS. The results suggest that, in many countries, teachers (especially grade four teachers) retain a substantial degree of discretion in what to teach. Variability in teacher behaviors related to opportunity to learn suggests it is highly advisable to consider this factor in studies of teacher effectiveness.

Keywords Opportunity to learn · Teacher education · Teacher experience · Teacher quality · Trends in International Mathematics and Science Study (TIMSS)

4.1 Empirical Approach

In this chapter we present descriptive results for educational systems (or “countries”) participating all waves of the TIMSS between 1995 and 2015, with additional attention paid to those countries that were part of the TIMSS in every cycle since

2003. The focus of this chapter is on student mathematics achievement and its relation to factors associated with teacher quality, which we assess using both teacher characteristics (experience, self-efficacy, formal preparation, gender) and teacher behaviors (time spent on teaching mathematics, content coverage); student characteristics (gender, socioeconomic status, language spoken in the home) are also considered. Our aim was to lay the foundations for the analyses in the later chapters, as well as identify general patterns in teacher quality across TIMSS countries.

Education system (henceforth referred to as country) means were calculated using the IEA Database Analyzer (free to access at www.iea.nl/data), with standard errors generated using a jackknife procedure. Because the TIMSS sampling design recruits a sample of representative classrooms rather than a sample of representative teachers, mean teacher results for each country do not necessarily reflect those of all teachers in a given educational system. Further, teacher-level data are not straightforward means, with each teacher counted equally. Rather, following the TIMSS sampling design, teachers are weighted according to the representativeness of their classrooms, based on the stratified sampling frame in each participating country.

In this descriptive analysis, we first examined all the data across cycles¹ and then by education system. In general, between country differences were observed exhaustively across all variables. Within country variation over time was explored to identify patterns and anomalies. Only countries with three or more years of data were included when examining trends within countries over time. To identify significant differences, confidence intervals were calculated for each country and year combination by using standard errors. In some instances, the data were examined across years only to determine if there were any significant changes over time generally. In this chapter, we report summary frequency distributions for country mean values for each year and grade level (complete results are provided in Appendix A, and should be consulted as source data in the discussion of patterns). Where we discuss general patterns, we first reference all participating countries in a given year, rather than the common set of countries that participated every year (see Table 4.1). Because this could introduce bias (given changes in country participation), we also report mean results from a more limited common pool of countries that participated in every cycle of TIMSS from 2003 to 2015 (see Table 4.2) For most variables, this comprises 18 education systems at grade four and 26 education systems at grade eight (see Table 4.3).

¹This descriptive analysis focuses on data from 2003, 2007, 2011, and 2015. While data from 1995 and 1999 are available in some instances, availability is inconsistent and thus excluded here.

Table 4.1 International means of the key variables considered in this study for all participating TIMSS countries, 1995–2015

Grade	n	Year	Alignment	Books	Lang	Mathprep	Mathtime	Prepared	Exp	Tmale	Performance
4	26	1995		3.23	1.23		236		16.37	0.23	496
4	29	2003	0.56	2.78	1.57	3.53	243	2.52	15.88	0.21	490
4	44	2007	0.57	2.77	1.51	3.59	235	3.38	16.71	0.20	479
4	57	2011	0.51	2.75	1.35	3.66	267	3.33	16.72	0.21	487
4	52	2015	0.54	2.79	1.62	3.85	257	3.40	16.67	0.18	507
8	40	1995		3.32	1.12		217		17.38	0.42	504
8	38	1999		3.02	1.26	3.63	210		16.58	0.40	487
8	51	2003	0.67	2.83	1.55	3.69	211	2.58	15.93	0.42	468
8	57	2007	0.61	2.72	1.58	3.71	201	3.58	15.21	0.43	458
8	50	2011	0.55	2.68	1.64	3.56	230	3.29	15.39	0.42	464
8	44	2015	0.55	2.70	1.66	3.89	226	3.28	15.44	0.40	481

Notes Grade = grade level, n = number of educational systems that participated, Year = year of the TIMSS cycle, Alignment = proportion of mathematics topics reported as covered by teachers compared with national expectations, Books = index (1–5) of number of books in the home, Lang = index (1–4) of testing language spoken in the home, Mathprep = index (1–5) of teacher education to teach mathematics, Mathtime = mean number of minutes spent on mathematics teaching per week, Prepared = index (1–4) of self-efficacy to teach mathematics, Exp = experience teaching in years, Tmale = index of teacher gender (female = 0, male = 1), Performance = mean student TIMSS score in mathematics

Table 4.2 International means for the key variables considered in this study for commonly participating countries (see Table 4.3), 2003–2015

Grade	n	Year	Alignment	Books	Lang	Mathprep	Mathtime	Prepared	Exp	Tmale	Performance
4	18	2003	0.56	2.87	1.51	3.41	242	2.52	15.92	0.21	505
4	18	2007	0.57	2.86	1.49	3.53	245	3.38	16.36	0.21	511
4	18	2011	0.53	2.86	1.35	3.66	278	3.36	17.23	0.20	518
4	18	2015	0.56	2.81	1.59	3.74	264	3.37	17.65	0.19	527
8	26	2003	0.66	2.93	1.54	3.52	212	2.60	15.01	0.44	490
8	26	2007	0.62	2.87	1.55	3.62	205	3.57	14.79	0.42	487
8	26	2011	0.56	2.81	1.58	3.48	226	3.28	15.25	0.42	495
8	26	2015	0.58	2.77	1.60	3.79	220	3.25	15.78	0.40	501

Notes Grade = grade level, n = number of educational systems that participated, Year = year of the TIMSS cycle, Alignment = proportion of mathematics topics reported as covered by teachers compared with national expectations, Books = index (1–5) of number of books in the home, Lang = index (1–4) of testing language spoken in the home, Mathprep = index (1–5) of teacher education to teach mathematics, Mathtime = mean number of minutes spent on mathematics teaching per week, Prepared = index (1–4) of self-efficacy to teach mathematics, Exp = experience teaching in years, Tmale = index of teacher gender (female = 0, male = 1), Performance = mean student TIMSS score in mathematics

Table 4.3 Educational systems participating in all cycles of TIMSS, 2003–2015

Grade 4	Grade 8
Australia	Australia
Ontario (Canada)	Bahrain
Quebec (Canada)	Botswana
Chinese Taipei	Ontario (Canada)
England	Quebec (Canada)
Hungary	Chinese Taipei
Iran	England
Italy	Hong Kong
Japan	Hungary
Lithuania	Iran
Morocco	Israel
Netherlands	Italy
New Zealand	Japan
Norway	Jordan
Russian Federation	South Korea
Singapore	Lebanon
Slovenia	Lithuania
United States	Malaysia
	Morocco
	Norway
	Russian Federation
	Saudi Arabia
	Singapore
	Slovenia
	Sweden
	United States

4.2 Curricular Alignment

4.2.1 Grade Four

The teacher curricular alignment variable (Alignment) responses indicate that alignment between a country’s national expectations of topic coverage and actual instruction ranged between 0.34 and 0.72, where 0.00 reflects no alignment and 1.00 reflects perfect alignment between curriculum and instruction (see Appendix A). This wide range suggests there is variation in curriculum alignment across countries and years, but warrants further exploration. The mean curricular alignment was 0.54, suggesting that on average, teacher instruction was aligned with the

national curriculum just over 50% of the time as an average for all TIMSS countries participating in the cycles from 2003 to 2015. For the subset of countries that participated in each year, the mean remained essentially constant (0.56 in both 2003 and 2015).

Examining variation in curricular alignment within a country over time provides interesting distinctions between educational systems. In some countries, alignment remained relatively constant between cycles. For example, in England, curricular alignment was 0.58 in 2003, 0.55 in 2007, 0.57 in 2011, and 0.55 in 2015. The overlapping confidence intervals suggest these differences were not significantly different, providing evidence that curriculum alignment in England has remained steady since 2003. In other countries, alignment has changed across cycles. For example, curricular alignment in the United States was 0.68 in 2007, 0.56 in 2011, and 0.62 in 2015 (United States data were not available for this variable prior to 2007).² In this case, confidence intervals do not overlap for any cycle, indicating statistically significant differences in curricular alignment within the United States over the various cycles of TIMSS. Differences in curricular alignment within a country over time may be attributable to policy change, but further research would be needed to verify such strong conclusions. In the United States example, most states adopted the Common Core State Standards in the 2010–2011 school year (see <http://www.corestandards.org/about-the-standards/development-process/>), which may have contributed to the recorded decline in curricular alignment in 2011 (as states and teachers adjusted to the new standards).

4.2.2 *Grade Eight*

In grade eight, there was a larger spread in the values associated with teacher curricular alignment than at grade four, which ranged from 0.25 to 0.88 (see Appendix A). Mean curricular alignment was slightly higher than that at grade four (0.59), meaning that, on average, instruction was aligned with the given national curriculum 59% of the time. For the countries that participated in every cycle of TIMSS from 2003 to 2015, the average curricular alignment declined from 0.66 to 0.58.

Examining within country variation over time indicated that although some countries had constant curricular alignment over the years (i.e., Georgia), others showed gradual improvements in curricular alignment. For example, Australia's curricular alignment has steadily improved since 2003, with curricular alignment measuring 0.54 in 2003, 0.55 in 2007, 0.57 in 2011, and 0.59 in 2015; note that the difference between their curricular alignment in 2003 and 2011 was significant. More often than not, curricular alignment demonstrates a more random pattern of

²Although some countries, like the United States, do not have a national curriculum, in such cases national-level educational administrators (like the United States Department of Education) responded to the TIMSS national curriculum survey, and provided estimates of the national expectations for topic coverage at each grade level.

variation across cycles, alternating between increases and decreases. For example, in the United States, curricular alignment decreased in 2011 as it did at grade four; again this dip may be attributed to short-term policy changes, and the transition to the new Common Core State Standards.

Japan showed the highest average degree of curricular alignment over time; alignment remained consistently above 80% in the 2003, 2007, 2011, and 2015 cycles of TIMSS.

4.3 Teacher Preparation to Teach Mathematics

4.3.1 *Grade Four*

The teacher preparation to teach mathematics (Mathprep) variable for grade four ranged from 1.00 to 4.93 (see Appendix A). This variable was measured by a five-point scale, where 1.00 corresponds with having no formal preparation to teach mathematics and 5.00 corresponds with having specialized preparation in both primary education and mathematics (namely that a given primary teacher would have been trained for content knowledge (CK), pedagogical content knowledge (PCK), and general pedagogical knowledge appropriate for grade four students). A mean value of 3.67 indicates that, on average, across all cycles and countries, teachers majored in primary education and/or mathematics but did not always have both qualifications. For the pool of 18 educational systems that participated in all four cycles of TIMSS from 2003 to 2015, the mean was 3.59 across years, rising steadily from 3.41 to 3.74.

In general, values for teacher preparation to teach mathematics at grade four remained relatively consistent within countries throughout the testing years, with only slight variations within some countries. For example, Quebec's values for this variable were 3.86, 3.89, 3.95, and 3.95 for the years 2003, 2007, 2011, and 2015, respectively. While these values slightly increased over the years, the differences were not significant, which suggests that there was little change in teacher preparation to teach mathematics at grade four in Quebec. Conversely, as an example an education system that showed significant changes in teacher preparation to teach mathematics at grade four, in Singapore the value went from 3.55 in 2003 to 4.18 in 2015. This significant difference indicates that teachers in Singapore have become better prepared to teach mathematics in grade four. Significant differences in teacher preparation to teach mathematics within countries, as in Singapore, may reflect policy changes that impacted teacher preparation, meriting further exploration.

Notably, Italy had consistently low values for teacher preparation to teach mathematics in grade four, of <1.50 for all four cycles of TIMSS. At the opposite end of the spectrum, in the Netherlands, teacher preparation to teach mathematics values were >4.5 for three out of the four TIMSS cycles considered.

4.3.2 Grade Eight

The findings for teacher preparation to teach mathematics in grade eight were very similar to those for grade four, ranging between 1.00 and 4.89 with a mean of 3.69, which suggests that, on average, grade eight teachers majored in mathematics education and/or mathematics, but did not always have both qualifications (see Appendix A). Again, this TIMSS scale presumes that more exhaustive formal preparation in mathematics content and pedagogy indicates better preparation to teach mathematics. The mean for commonly participating countries (Table 4.3) was 3.60, increasing from 3.52 in 2003 to 3.79 in 2015.

Throughout the testing years, there was more within country variation in teacher preparation to teach mathematics at grade eight than at grade four. Teacher preparation to teach mathematics in grade eight significantly increased between 2003 and 2015 in some countries (such as Quebec, Canada), and significantly decreased in others (for example, Saudi Arabia), or showed no consistent pattern (for example, England). The greater temporal within-country variation in teacher preparation to teach mathematics in grade eight points to possible between grade differences in teacher preparation requirements. For example, perhaps the requirements around teacher preparation to teach mathematics in grade four are more clearly defined than they are for grade eight.

At grade eight, Morocco consistently had low values associated with teacher preparation to teach mathematics (three reporting years with values <2.00), while conversely, Romania had three reporting years with values >4.00 .

Interestingly, at both grades four and eight, the mean values associated with teacher preparation to teach mathematics increased over time. At grade four, the mean value for teacher preparation to teach mathematics was 3.53 in 2003 and 3.85 in 2015. Again, at grade eight, the mean value was 3.69 in 2003 and 3.88 in 2015. Similar results were found when trends were restricted to only those countries that participated in all four cycles of TIMSS between 2003 and 2015 (an increase of >0.3 for both grades). These increases may point to overall improved teacher preparation to teach mathematics at grades four and eight across countries. One possible explanation for this improvement is that teacher preparation became more of a policy priority over time, resulting in more stringent teacher preparation requirements.

4.4 Teacher Time on Mathematics

4.4.1 Grade Four

The large range (from 65.90 to 479.24 min) of teacher reported time spent on teaching mathematics as measured in minutes at grade four (Mathtime) suggests significant variation within counties over time (see Appendix A). For many countries there were differences in time spent on teaching mathematics across the different test cycles.

For example, Australia's average teacher time on mathematics started at 260 min in 2003, remained steady in 2007 at 266 min, increased dramatically to 346 min in 2011, and finally decreased to 306 min in 2015. The changes in 2011 and 2015 were both significant, and may reflect the introduction of a national curriculum between 2008 and 2012. Similarly, time on mathematics in the United States started at 245 min in 2003, and subsequently increased to 289 min in 2007, 343 min in 2011, and 359 min in 2015. These increases were all statistically significant, with the largest jump occurring between 2007 and 2011. As mentioned earlier, this increase in the time teachers spent on teaching mathematics in the United States may be attributed to implementation of the Common Core State Standards. There was a general increase in time spent on mathematics in fourth grade among the pool of commonly participating countries (Table 4.3), rising from 242 min in 2003 to 264 min in 2015 (although this was a decrease from a high of 277 min in 2011).

The wide within-country variation across years may suggest that time spent teaching mathematics in grade four is not often standardized. However, there were a handful of instances where teacher time on mathematics was consistent within countries over years, such as in Singapore where time spent on mathematics varied only between 325 and 329 min over five testing periods.

Countries consistently at the lower end of the grade four time on mathematics spectrum included Chinese Taipei, Norway, and Sweden; countries consistently at the higher end of the grade four time on mathematics spectrum included Portugal, Canada (Quebec), Italy, and the United States. Taken together, these findings suggest wide variations in time on mathematics in grade four.

4.4.2 Grade Eight

While the range (80.60–350.35 min) of teacher reported time spent on teaching mathematics as measured in minutes in grade eight was not quite as large as grade four, it was still large enough to warrant further exploration (see Appendix A). There was a weaker trend among commonly participating TIMSS countries (Table 4.3) between 2003 and 2015, with the average time spent on teaching mathematics increasing from 212 to 220 min, although, as with grade four, the largest averages were found in 2011.

Examining within country variation provides a clearer story about how much time teachers spend teaching mathematics in grade eight. There was still variation in teacher time on mathematics in grade eight, but the variation was not as large as observed for grade four. Recall that Australia's average time on mathematics in grade four ranged from 260 to 346 min over four test cycles. For grade eight, Australia's time spent on teaching mathematics varied between 208 and 220 min over those same four test cycles. Similarly, in the United States, while there were still significant increases in grade eight time spent on mathematics over the test cycles, the increases were not as large as they were in grade four. For example, over four test cycles, the grade four time on mathematics in the United States varied from 245 to 359 min, a sweep of

over 100 min; meanwhile, the grade eight time on mathematics in the United States over the same four test cycles ranged only between 226 and 265 min. These differing degrees of variation within countries for time on mathematics at grades four and eight may suggest there are widespread grade-level differences in how much time teachers spend on teaching mathematics.

Countries consistently at the lower end of the grade eight time on mathematics spectrum included Cyprus, the Netherlands, Japan, and Sweden; countries consistently at the higher end of the grade eight time on mathematics spectrum included Lebanon, Canada (Ontario), Chile, and the United States.

4.5 Teacher Preparedness

4.5.1 *Grade Four*

Teacher feelings of preparedness to teach mathematics (Prepared) were measured on a four-point scale, where higher values indicated teachers felt better prepared to teach mathematics. At grade four, teacher feelings of preparedness ranged from 2.07 to 3.88, with a mean value of 3.24 (see Appendix A). For those countries that participated in each cycle of TIMSS between 2003 and 2015 (Table 4.3), the mean was 3.16 (rising sharply between 2003 and 2007, but remaining stable afterwards). This mean value suggests that generally teachers feel well prepared to teach mathematics.

Some recognizable patterns emerge when examining grade four teacher feelings of preparedness within countries over the different testing years. In many countries, the values associated with grade four teacher feelings of preparedness significantly increased between the 2003 and 2007 test cycles, and then remained steady for the remaining test cycles. For example, the grade four teacher feelings of preparedness in Norway were 2.61 in 2003, 3.69 in 2007, 3.51 in 2011, and 3.78 in 2015. Similar increases between 2003 and 2007 were seen in several other countries, such as Australia, Belgium (Flemish), Canada (Ontario), Canada (Quebec), Hungary, Italy, Morocco, and the United States. This indicates some change around teacher feelings of preparedness occurred between 2003 and 2007. One possible explanation is that TIMSS does not ask about the same mathematics topics in every cycle; there are cycle-to-cycle alterations to the TIMSS framework that mean that the same topics are not asked every time. It is thus possible that differences in mean teacher responses could be partly attributable to the survey instrument rather than the underlying construct.

4.5.2 *Grade Eight*

For grade eight, there was a wider range for teacher feelings of preparedness than seen at grade four, with values from 1.06 to 3.91 and an overall mean value of 2.91 (see

Appendix A). The average of 2.91 may suggest that generally grade eight teachers felt adequately prepared to teach mathematics, but their feelings of preparedness are not as strong as those of the grade four teachers. However, when the pool of education systems was restricted to only those that participated in all cycles of TIMSS between 2003 and 2015, the international mean was 3.17, virtually indistinguishable from that reported for grade four teachers.

As with grade four, the values associated with grade eight teacher feelings of preparedness significantly increased between the 2003 and 2007 test cycles, and then remained steady for the remaining test cycles. For example, the grade eight teacher feelings of preparedness in the United States were 2.84 in 2003, 3.87 in 2007, 3.66 in 2011, and 3.48 in 2015. Many other countries displayed this pattern when looking at grade eight teacher feelings of preparedness and, again, this pattern may indicate a systemic change in how TIMSS measured this variable. By contrast, at grade eight, Denmark reported high values (>3.66) for teacher feelings of preparedness for three consecutive test cycles.

At both grades four and eight, the mean values associated with teacher feelings of preparedness displayed large increases between 2003 and 2007, and levelled off in subsequent cycles of TIMSS. For example, at grade four the overall mean value for teacher feelings of preparedness was 2.52 in 2003 and 3.38 in 2007. Similarly, at grade eight the overall mean value was 2.58 in 2003 and 3.58 in 2007. It seems unlikely that teacher feelings of preparedness would change so considerably between two consecutive test cycles, suggesting that these increases may also reflect changes in the metric of teacher feelings of preparedness (due to alterations in the TIMSS framework); this would affect all grade levels and education systems.

4.6 Teacher Experience

4.6.1 *Grade Four*

The teacher experience variable captures the total number of years the teacher has been teaching (Exp). At grade four, teacher experience ranged from 7.63 to 27.64 years, with a mean reported value of 16.29 years (see Appendix A). The comparable mean for the restricted sample of eighteen countries was 16.79 years, rising from 15.92 years in 2003 to 17.65 years in 2015, suggesting an increase in teacher experience over time.

When looking at teacher experience within countries over years, there were some countries where the amount of reported teacher experience was consistent and some countries that showed significant variation in reported teacher experience. Teacher experience in the United States was fairly consistent, hovering between a mean of 13 and 14 years of experience across the test cycles from 2003 to 2015. However, the 2015 mean value of 13.13 years was significantly lower than the other values. This slight decrease in mean years of teacher experience may coincide with the economic

crisis the United States experienced in the late 2000s. During this time, many older teachers opted for retirement options, and this would have systematically driven down mean teacher experience as measured in years. Conversely, mean grade four teacher experience appeared quite variable in Ontario, Canada. While the values for Ontario were similar in 2003 and 2007 (13.11 and 13.15 years, respectively), mean teacher experience dropped to 11.52 years in 2011 and then substantially increased to 14.99 years in 2015. Further research is required to better understand such significant variations.

The education systems that consistently reported the lowest number of years of teacher experience at grade four included Kuwait and Singapore; both reported multiple values ≤ 10 years. Those that consistently reported the highest number of years of teacher experience at grade four included Lithuania and Georgia; both reported multiple values > 20 years.

4.6.2 Grade Eight

At grade eight, teacher experience ranged from 5.00 to 26.72 years, with a mean of 15.77 years (see Appendix A). In other words, grade eight teachers who teach mathematics have 15.77 years of teaching experience on average across test cycles, as measured by TIMSS. The comparable mean for the restricted sample of eighteen countries was 15.21 years. The number of years of teacher experience reported by grade eight teachers started at a lower level and ended at a lower level than grade four, suggesting there were between-grade differences in the level of teacher experience. The trends suggested modest change in net experience, increasing from 15.01 years in 2003 to 15.78 years in 2015.

The within-country analysis of grade eight teacher experience also revealed similar findings to grade four; some countries displayed consistency in reported teacher experience at grade eight and some countries displayed variation in reported teacher experience at grade eight. For example, teacher experience in Australia, Italy, and the United States was relatively consistent across cycles at grade eight, while there was more variation in reported teacher experience at grade eight in Chile, Egypt, and Chinese Taipei.

The education systems that consistently reported the lowest number of years of teacher experience at grade eight included Ghana and Botswana; both reported multiple values < 9 years. Those that consistently reported the highest number of years of teacher experience at grade eight included Romania and the Russian Federation; both reported multiple values > 20 years.

4.7 Teacher Gender

4.7.1 *Grade Four*

Teacher gender is a dichotomous variable; a response of 0.00 denotes the teacher is female and a response of 1.00 denotes the teacher is male (Tmale). For grade four, the teacher gender variable ranged between 0.00 and 0.80, with a mean of 0.21 (see Appendix A). This mean value is quite informative because an average close to 0.00 indicates teachers tend to be female, a mean of 0.50 indicates male and female teachers are equally represented, and a mean close to 1.00 indicates teachers tend to be male. The grade four mean of 0.21 (0.20 for the restricted sample), which holds over time, indicates that TIMSS respondents were more likely to be female. In general, teaching tends to be a female-dominated profession in many countries (especially at earlier grades), so these results are not surprising.

Between-country findings yielded interesting results related to grade four teacher gender. Most countries' means for this variable were closer to 0.00 across test cycles, indicating that teachers in each of the countries tended to be female. However, a few countries had results that suggested teacher representation was more gender balanced. Denmark, for example, reported teacher gender values of 0.51 in 2007, 0.42 in 2011, and 0.47 in 2015. In Denmark there are almost equal numbers of male and female teachers, which may reflect Denmark's reputation for greater gender equity. In a few countries, the results indicated that teachers were more often male. In Yemen, for example, teacher gender values were 0.74 in 2003, 0.74 in 2007, and 0.78 in 2011; one explanation for this might be that the culture in Yemen supports a more male-dominated workforce, resulting in fewer female employees in general.

In general, the within-country analysis uncovered little variation in grade four teacher gender across years. However, a few countries did see decreases in their teacher gender variable over time. For example, Morocco's values for grade four teacher gender were 0.64 in 2003, 0.50 in 2007, 0.50 in 2011, and 0.37 in 2015. The substantial change from 2003 to 2015 implies teachers were more likely to be males in 2003, but, by 2015, teachers were more likely to be female. One possible explanation for this shift is that, over time, it became more socially acceptable for females to hold teaching positions.

4.7.2 *Grade Eight*

The descriptive statistics for teacher gender in grade eight differ from those for grade four. At grade eight, teacher gender varied between 0.00 and 1.00, with a mean of 0.42 for both the overall sample and the more restricted sample of eighteen countries (see Appendix A). This mean holds over time and implies that teachers were more often female, and also that for grade eight the gender distribution was more equal between males and females than at grade four. This also aligns with our general

conception that teachers in elementary grades are more likely to be female, whereas teachers in upper elementary and high school are more evenly distributed between males and females.

While many countries, such as the United States, had consistent values for grade eight teacher gender over time, others, like Japan, displayed a level of variation. Generally, there appeared to be more within-country variation for grade eight teacher gender than there was at grade four.

For all four cycles of TIMSS that we investigated, Ghana, Morocco, and Japan consistently reported having more male teachers than female teachers, whereas the Russian Federation, Latvia, Lithuania, and Georgia reported that their teachers were almost exclusively female teachers.

4.8 Student Performance

4.8.1 *Grade Four*

At grade four, student mathematics performance (Performance) in TIMSS ranged from a point score of 223 to 618, with a mean of 491 (see Appendix A). Examining the international mean over time suggests slight increases in overall performance over time. For example, at grade four, the overall mean student performance was 490 in 2003; by 2015, the mean had increased to 506. For our more restricted sample of countries, the international mean rose from 505 in 2003 to 527 in 2015.

Cross-national comparison of grade four student performance within countries showed that some countries demonstrated considerably more variation than others, which may merit deeper investigation. For example, in Armenia, the mean grade four student performance was 455.92 in 2003, 499.51 in 2007, and 452.28 in 2011. These back and forth changes suggest a degree of instability in grade four student performance between test cycles in Armenia. Other countries with considerable variation included Qatar, Yemen, and Kuwait, to name a few. At the same time, Australia saw very little variation in their mean grade four student performance over time, with scores of 499 in 2003, 516 in 2007, 516 in 2011, and 517 in 2015. Other countries exhibiting similar consistency include Belgium (Flemish), New Zealand, and Italy.

As has been widely noted in scholarly and popular publications, countries with the highest scores over time are largely located in East Asia, while countries with the lowest scores over time are largely located in West Asia and Africa. As such, geographical and cultural differences may play an important role in student achievement.

4.8.2 Grade Eight

At grade eight, the range in student performance scores over time and across countries varied between 264 and 621 score points, with an international mean of 475 (see Appendix A). The examination of the mean over time yields similar findings to that for grade four, with the mean slightly increasing from 468 in 2003 to 481 in 2015, and from 490 in 2003 to 501 in 2015 for our more restricted sample (Table 4.3). In other words, overall TIMSS performance has improved over the test cycles that we considered in our analysis.

In general, there was greater variation in the within-country analysis of student performance at grade eight than at grade four. For example, changes in Chile's grade eight student performance scores were substantial, with a score of 387 in 2003 rising to a score of 427 in 2015. This increase aligns with the overall international increase in TIMSS performance over time. Meanwhile, Malaysia saw a drop in grade eight student performance, their mean score being 508 in 2003 and 465 in 2015. Australia was one of the few countries whose student performance scores were consistent over time at grade eight.

As we found for grade four, countries with the highest grade eight student performance over time were largely located in East Asia, while countries with the lowest grade eight student performance over time were often located in West Asia and Africa.

4.9 Books in the Home

The number of books in the home (Books) is a control variable that serves as a proxy variable to indicate student socioeconomic status. In the TIMSS survey, students are asked to estimate the number of books in their home, with responses placed on a 1 to 5 scale. A larger value denotes more books in the home, which generally corresponds to higher socioeconomic status. Values for books in home looked similar across grades four and eight, ranging between 1.61 and 4.04 at grade four, and 1.84 and 4.31 at grade eight (Appendix A). The international mean across cycles for books in home was 2.82 at grade four and 2.81 at grade eight. International means were quite similar for our more restricted sample of 18 countries, being 2.85 at both grade four and grade eight. Within-country variation across cycles was minimal, indicating that there was little change in socioeconomic conditions across the TIMSS administrations. In general, countries with higher values for books in home were the wealthier countries, whereas countries with lower values for books in home were less wealthy countries.

4.10 Student Language

Student language (Lang) is a control variable that captures the alignment between the language the test is delivered in and how often that same language is spoken in the student's home. This variable was measured on a four-point scale. A lower value indicates more overlap between the language of the test and language spoken at home, for example, a value of 1.00 denotes that the student always speaks the language of the test at home, whereas a value of 4.00 means that the student never speaks the language of the test at home. Values for student language at grade four ranged between 1.03 and 3.08, with a mean of 1.47 (or 1.49 in our more restricted sample of 18 countries; Table 4.3). The distribution of student language indicated that, in most instances, the language of the test was always or almost always spoken at home. Grade eight results were quite similar, with a mean of 1.52 (1.57 for the restricted sample of commonly participating countries; Table 4.3).

4.11 Conclusions

With respect to Research Question 1, “*Are there identifiable trends in teacher quality and instructional metrics over time?*”, the results to this chapter indicate that while trends vary from country to country, a focus on a common pool of countries (Table 4.3) suggested substantial change in teacher quality metrics. Specifically, at both grades four and eight, there were broad-based increases in teacher education and teacher experience. There was also an increase in time spent on mathematics in grade four over the cycles of TIMSS that we investigated, and a smaller increase at grade eight. By contrast, teacher self-reported preparedness to teach mathematics has been largely stable since 2007. Alignment of instructional content with national expectations has also stayed at a relatively consistent level at grade four, but alignment has declined since 2003 at grade eight.

The most striking finding of this chapter was the very modest degree of alignment between teacher instructional content and national expectations of content coverage in mathematics. Among those educational systems reporting alignment in all four TIMSS cycles that we investigated, the international mean was only 0.55 at grade four and 0.60 at grade eight. At grade four, only two educational systems, Hong Kong and Korea, exhibited instructional alignment $\geq 70\%$ over the last four cycles of TIMSS. High alignment rates were much more common at grade eight. It was further quite surprising to find that the United States, which has no official national curriculum, scored relatively highly for alignment. Overall, the results suggest that, in many countries, teachers (especially grade four teachers) maintain substantial discretion in what to teach, a conclusion bolstered by the similarly strong variation

in time spent on mathematics. Compared with the much more stable teacher characteristics (experience, feelings of preparedness, and level of formal education) reported in this chapter, the reported variability in teacher behaviors related to opportunity to learn demonstrates the importance of incorporating these factors into studies of effectiveness.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial 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 license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license 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.

