

Chapter 3

Teacher Variables, and Modeling the Relationships Between Teacher Characteristics, Teacher Behaviors, and Student Outcomes



Abstract After a brief discussion of the history of IEA's Trends in International Mathematics and Science Study (TIMSS), and the implications this has had for comparative research into the influence of teachers on student outcomes, this chapter details the key variables that are employed in the remainder of the book. These include several student-level control variables (such as gender, socioeconomic status, and language spoken in the home), teacher-reported measures (experience, education, gender, and time spent on teaching mathematics), and newly developed instruments based on TIMSS responses related to teacher self-efficacy and instructional alignment. All of the analyses in this report make use of TIMSS data, merging several distinct instruments: student assessments, teacher background surveys, and national background surveys. The design of TIMSS has a number of important implications. Relationships between variables should be treated as associations rather than strictly causal, thus a multi-model analytical strategy is needed to test the robustness of statistical results and improve confidence in the reliability of analyses. Relationships were examined across time, across different aggregations of data, and using different statistical procedures, and are discussed in more detail in later chapters. Many features of TIMSS and the topics included in the survey have changed considerably over time and some key variables of interest were not included in earlier versions of the TIMSS, greatly restricting the available sample when analyzing country-level trends (whether as means or regression coefficients). The complex sampling design has important implications for statistical modeling and analysis, and for purposes of analysis, the models in subsequent chapters ignore school effects. Multilevel and classroom-mean models treat each classroom as existing independently, rather than being clustered within schools. Given the acknowledged impact of schools, and the importance of within-school, between-classroom heterogeneity, any conclusions drawn from the analysis should be treated with caution.

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

3.1 The TIMSS Dataset

All of the analyses in this report make use of IEA's Trends in International Mathematics and Science Study (TIMSS) data, merging several distinct instruments: student assessments, teacher background surveys, and national background surveys. The design of TIMSS uses a stratified random sampling to select a representative sample of schools at each grade level of the study, and then to sample intact classrooms within those schools. Separate surveys are conducted at grades four and eight. Since the first cycle of TIMSS in 1995, there have been four subsequent iterations at both grades four and eight (in 2003, 2007, 2011, 2015), as well as an additional survey at grade eight undertaken in 1999.

The design of TIMSS has a number of important implications for researchers. First, despite continuing over time, the TIMSS is not a true longitudinal sample, but rather a series of cross-sectional studies. Because TIMSS does not sample the same group of students, teachers, or schools across time, many of the most common identification strategies employed by researchers to account for unobserved variables and other biases are not available. All of the relationships between variables should therefore be treated as suggested associations rather than strictly causal. To cope with this limitation, we adopted a multi-model analytical strategy to test for the robustness of statistical results. The aim was to improve confidence in the reliability of analyses by examining the stability of relationships across time, across different aggregations of data, and using different statistical procedures. These approaches are discussed in more detail in later chapters.

Second, many features of TIMSS have changed considerably over time. Some of the key variables of interest, such as teacher preparedness to teach mathematics topics, or the mathematics topics expected to be taught according to the national research coordinators, were not included in earlier cycles of TIMSS. The list of topics included in TIMSS has also changed considerably since 1995. For example, although the literature suggests that receipt of professional development could be related to student outcomes (see for example Blank and De Las Alas 2009), the variation in how this construct is incorporated in different cycles of TIMSS is too diverse (in our judgment) to permit its consideration as a variable in multi-year statistical models.¹ Similarly, participating educational systems differ from one cycle of TIMSS to another; they may be present in one cycle of TIMSS but not in another, or may only participate at grade four but not at grade eight. Such variations greatly restrict the available sample when analyzing education-system-level trends (whether means or regression coefficients). For example, Germany did not take part at grade four in 2003, and only participated at grade eight in 1995.

Thirdly, the complex sampling design of TIMSS has important implications for statistical modeling and analysis. Because typically only one intact classroom is selected in a given school, it is very difficult to distinguish classroom- from

¹For example, teachers were asked whether they had received professional development related to critical thinking in 2003 and 2007, but not in 2011 and 2015. By contrast, they were asked about professional development focused on student needs in 2011 and 2015, but not 2003 and 2007.

school-level effects. Therefore, for analytical purposes, we ignored school-level effects in our models. Multilevel and classroom-mean models treat each classroom as existing independently, rather than being clustered within schools. Given the acknowledged impact of schools, and the importance of within-school between-classroom heterogeneity, this amounts to an important qualification to any conclusions drawn from our analysis (Schmidt et al. 2015). Readers should understand that TIMSS samples a random population of classrooms meant to reflect an entire education system's population, rather than a straightforward random sample of all students or teachers. As such, strictly speaking, the associations explored in this study should be interpreted as reflecting representative classrooms, rather than all of an educational system's teachers and students.

The complex sampling design, which stratifies within blocks of schools, giving different weights to different schools (and hence classrooms), means that standard calculation of standard errors (and hence statistical significance) is inadvisable. Instead, TIMSS requires a jackknifing procedure, which produces more accurate standard error estimates but at the cost of much greater computational burdens (especially for multilevel models). It also introduces problems with subdividing types of classrooms, as we discuss in Chap. 7.

The educational systems examined in our analyses are usually referred to as "countries." This is for ease of reading, but it should be noted that there are a number of systems that are not countries as such, but are units with a degree of educational autonomy that have participated in TIMSS following the same standards for sampling and testing.

A final word of caution about the analyses in forthcoming chapters. Cross-country comparisons of teacher characteristics and behaviors inevitably raise questions about the consistency of these concepts across different cultural contexts. This is a particular problem for survey results, since along with difficulties in translation, there may be problems related to culturally-specific interpretations; these are some of the inevitable challenges encountered when conducting comparative research.

3.2 Operationalization of Variables

The variables in this study are drawn from the TIMSS database (see <https://www.iea.nl/data>). Additional details can be found in the TIMSS 2015 user guide (Foy 2017) and other TIMSS user guides and technical reports (<https://timssandpirls.bc.edu/isc/publications.html>). The first three variables that we included in the various statistical models are student control variables that have been associated with student outcomes and/or different learning opportunities, which means that excluding them could result in spurious relationships between teacher quality and student outcomes. These are: student gender, number of books in the home (a common proxy measure of socioeconomic status reflecting both parental education and parental income), and whether the student speaks the language of the test at home (Chudgar and Luschei 2009). The remaining variables are teacher variables that the research literature and

policymakers have to varying degrees treated as measures of teacher effectiveness (as we have already discussed in Chap. 2): namely, education, experience, content knowledge, time on mathematics, and instructional content. We also incorporated an additional teacher-level variable that was included in all iterations of TIMSS, which is teacher gender.

3.2.1 Student Gender (*Stmale*)

The student gender variable (*Stmale*) denotes the gender of each student as indicated by the student on the TIMSS survey. Student gender is a dichotomous variable and is captured by the question: “Are you a girl or boy?” Higher values indicate a male student. Student gender is included for both grades four and eight for each cycle of TIMSS included in this study (2003, 2007, 2011, and 2015).²

3.2.2 Student Language (*Lang*)

The student language variable (*Lang*) denotes the language of each student, as indicated by the frequency of using the language of the test at home. Student language is captured by the question: “How often do you speak [language of test] at home?” The question has four response categories: (1) I always speak [language of test] at home; (2) I almost always speak [language of test] at home; (3) I sometimes speak [language of test] and sometimes speak another language at home; (4) I never speak [language of test] at home. Student language is included for both grades four and eight for each cycle of TIMSS included in this study (2003, 2007, 2011, and 2015).

3.2.3 Student Estimated Number of Books in the Home (*Books*)

The number of books in the home (*Books*) variable denotes the number range of books, not including magazines, newspapers, or school books, in each student’s home as estimated by the student. The variable is captured by the question: “About how many books are there in your home? (Do not count magazines, newspapers, or your school books.)” The question has five response categories: (1) None or very few (0–10 books); (2) Enough to fill one shelf (11–25 books); (3) Enough to fill one bookcase (26–100 books); (4) Enough to fill two bookcases (101–200 books); (5) Enough to

²Readers should note that, in Chap. 6, student gender, as well as several other variables, are coded differently in order to match the Blomeke et al. (2016) operationalization, and hence a different variable name is employed. Please see Chap. 6 for further details.

fill three or more bookcases (more than 200). As part of the TIMSS survey, each of the response categories included an explanatory illustration displaying how each category would likely look. The Books variable is included for both grades four and eight for each cycle of TIMSS included in this study (2003, 2007, 2011, and 2015).

3.2.4 Teacher Experience (Exp)

The teacher experience (Exp) variable denotes the total number of years the respondent reported teaching. Teacher experience is captured by the question: By the end of this school year, how many years will you have been teaching altogether? This is an open response question, and the years are reported in whole numbers. Teacher experience is included for both grades four and eight for each cycle of TIMSS included in this study (2003, 2007, 2011, and 2015). In our analyses, this variable is left as a simple linear variable in an effort to explore the general impact of teacher experience. However, as noted in Chap. 2, additional study should consider a non-linear relationship and/or distinguish early career teachers.

3.2.5 Teacher Gender (Tmale)

The teacher gender (Tmale) variable denotes the gender of each teacher as indicated by the teacher on the TIMSS survey. Teacher gender is a dichotomous variable and is captured by the question: “Are you female or male?” Teacher gender is included for both grades four and eight for each cycle of TIMSS included in this study (2003, 2007, 2011, and 2015). Higher values indicate male teachers. The TIMSS survey coded “2” as male and “1” as female. For the purposes of this work, the variable has been recoded as “1” for male teachers and “0” for female teachers.

3.2.6 Teacher Feelings of Preparedness (Prepared)

Teacher feelings of preparedness (Prepared) is an index of teacher reported self-efficacy to teach mathematics. TIMSS includes a series of questions asking teachers the degree to which they feel prepared to teach various mathematics topics. These topics vary across the various cycles of TIMSS, following the structure of topics in that year’s mathematics framework for that grade level. For each topic, teachers are asked “How well prepared do you feel you are to teach the following mathematics topics?” and can respond on a four-point scale: not applicable (1), very well prepared (2), somewhat prepared (3), and not well prepared (4). For the

purposes of this study, we recoded these responses so that 4 = very well prepared, 3 = somewhat prepared, 2 = not very well prepared, and 1 = not applicable. We then averaged the responses across mathematics topics.

3.2.7 Teacher Preparation to Teach Mathematics (Mathprep)

The Mathprep variable denotes mathematics teacher preparation by indicating teachers who majored in a combination of education and mathematics during their post-secondary studies. Mathprep is included for both grades four and eight for each cycle of TIMSS for which component variables of the index were available (2003, 2007, 2011, and 2015). Mathprep was included in the 2011 and 2015 TIMSS cycles, but had to be derived from the 2003 and 2007 data. The *TIMSS 2015 user guide for the international database* (Foy 2017; Supplement 3, p. 11 for grade four and p. 32 for grade eight) provides detailed information about how to create the variable Mathprep (called ATDM05 at grade four and BTDM05 at grade eight in Foy 2017). Following the coding scheme in the user guide (Foy 2017), the Mathprep variables for the 2003 and 2007 cycles were created similarly using “if then” statements to combine two variables for each grade level. For grade four, the first variable denoted the respondent’s post-secondary major or main area of study, categorized as: education = primary/elementary, education = secondary, mathematics, science, language of test, and other. The second variable denoted whether the respondent specialized in mathematics. For grade eight, the first variable denoted whether the respondent majored in mathematics during their post-secondary education. The second variable denoted whether the respondent majored in education with a mathematics specialty during their post-secondary education. The “if then” statements helped indicate 2003 and 2007 TIMSS respondents who majored in a combination of education and mathematics during their post-secondary studies. There are nuanced differences between the creation of Mathprep for grade four and grade eight because of differences in credentials required to teach mathematics in primary and secondary education. Grade four teachers often have a post-secondary major of primary/elementary education, whereas grade eight teachers often have a post-secondary major in their discipline of interest (such as mathematics).

The new variable resulted in five categories for grade four: (1) Respondent did not have formal education beyond upper-secondary; (2) Respondent majored in something other than primary education and/or mathematics; (3) Respondent majored in mathematics but did not major in primary education; (4) Respondent majored in primary education but did not major or specialize in mathematics; and (5) Respondent majored in primary education and majored or specialized in mathematics. Similarly, Mathprep resulted in five categories for grade eight: (1) Respondent did not have formal education beyond upper-secondary; (2) Respondent majored in something other than mathematics and/or mathematics education; (3) Respondent majored in mathematics education but did not major in mathematics; (4) Respondent

majored in mathematics but did not major in mathematics education; and (5) Respondent majored in mathematics and mathematics education.

After creating the Mathprep variable for 2003 and 2007, the 2011 and 2015 Mathprep variable was reverse coded to match the numeric schema for Mathprep in 2003 and 2007. This also allowed for those who majored in primary education/mathematics education and also majored in mathematics to be associated with the highest value (5).

3.2.8 Time Spent on Teaching Mathematics (Mathtime)

The Mathtime variable denotes the amount of time in minutes the respondent reported teaching mathematics to their given class. Mathtime is included for grades four and eight for all years in which the variable was available (2003, 2007, 2011, and 2015). Mathtime was included in TIMSS for years 2003, 2007, and 2015, but in 2011, for both grades four and eight, the Mathtime variable was captured by two variables: one reporting hours (ATBM01A) and the other reporting minutes (ATBM01B). To have consistency across years, we converted the “hours” variable into minutes by multiplying it by sixty. Then, we added the now-converted “hours” variable to the “minutes” variable, resulting in one measure of time spent on teaching mathematics as measured by minutes. This conversion was necessary to have a consistent measure of time spent on teaching mathematics across grade levels and years.

3.2.9 Teacher Curricular Alignment (Alignment)

This variable is an index measuring the degree to which teachers are instructing students in the topics that are intended to be taught at that grade level according to national curricular expectations. The variable was therefore constructed by combining data from two surveys: the TIMSS teacher background survey and the related national background survey. TIMSS teacher respondents were presented with a list of mathematics topics related to the topics contained in the TIMSS framework for that cycle, and asked to mark whether the topic was: (1) mostly taught before this year, (2) mostly taught this year, or (3) not yet taught or just introduced. These responses were then recoded so that “mostly taught” is coded as “1” and the other two responses as zeros. Next, national and/or provincial administrative leaders in education systems participating in the TIMSS were presented with an identical list of topics and asked to respond at what grade level that topic should be taught by teachers. These responses were recoded as “1” for “0”, with “1” indicating that the topic should be taught at the relevant grade level (grades four and eight). The entire matrix of topics in that grade level was then compared with teacher responses. Cells were coded as “1” if a given teacher followed national curricular expectations, in other words, if they did teach a topic that should be taught, or did not teach a topic that shouldn’t be taught. By contrast, a failure to match with national curricular

expectations about a particular topic earned a coding of “0.” The percentage of alignment was then calculated, with 1 indicating perfect alignment and 0 indicating perfect non-alignment.

3.3 Methods

As mentioned previously, the chapters that follow use a variety of statistical methods, reflecting the study’s purposefully multi-model approach. The analytic strategies are discussed briefly here, and are addressed in more detail in the appropriate chapters. Unless mentioned specifically, all analyses employ the procedure for generating jackknifed standard errors and (when considering student outcomes) the calculation and combination of all plausible values, as outlined in the various TIMSS technical reports (see <https://timssandpirls.bc.edu/isc/publications.html>) and the IEA Database Analyzer (see <https://www.iea.nl/data>).

Chapter 4 presents international averages of country-level means (including participating sub-units) for the predictors used throughout the remaining chapters, including teacher experience, teacher preparedness, teacher education (preparation to teach mathematics), teacher time spent on mathematics, and teacher curricular alignment. Variable means are calculated for each educational system participating in each cycle of TIMSS, and confidence intervals are used to identify statistically significant differences for the same country in different iterations of TIMSS.

Chapter 5 presents a number of statistical methods that we used to examine the relationship between teacher characteristics and behaviors and student outcomes. We used a multi-model approach to test the degree to which relationships remained robust across countries, within countries, and across different periods, and we assessed the stability of estimates using different statistical techniques. As a first step, we compared the results for (1) single-level linear regressions (ignoring classroom-level clusters) of individual student outcomes as predicted by teacher variables; (2) two-level linear regression models that cluster students within classrooms using SAS PROC MIXED maximum likelihood statistical software (Singer 1998); and (3) a single-level linear regression of classroom means of student variables with teacher variables. The comparison of these analyses enables the frequency of statistically significant associations between purported measures of teacher effectiveness and student achievement in mathematics to be determined. Additional analyses in Chap. 5 include an examination of the stability of multilevel model regression coefficients within each country across the different cycles of TIMSS, and an exploratory fixed effects regression analysis of changes in country-level means.

Whereas the models in Chap. 5 treat each of the teacher-level variables as independent variables, Chap. 6 introduces a model that uses teacher behaviors (time on mathematics and alignment with national curriculum standards) as mediating variables for instructional quality. We used a structural equation modeling (SEM) approach to analyzing the results for each country in a single cycle of TIMSS (namely 2011), permitting the inclusion of additional variables like teacher professional

development to estimate a latent construct of “instructional quality.” A multilevel model clustering students within classrooms using jackknifed standard errors and five plausible values was applied to each educational system. Comparison of the results in Chaps. 5 and 6 demonstrates that a wide range of statistical techniques can be used to assess whether there are temporally and cross-nationally robust associations between measures of teacher effectiveness and student achievement in mathematics.

Chapter 7 departs from the focus on mean student outcomes to consider the distributional effects of teacher effectiveness. First within-country equity was measured by standard deviation of pooled country-level student achievement. Country-level fixed effect analysis was used to assess the relationship between teacher effectiveness measures and student variation (measured by standard deviations). Second, the relationship between within-classroom variation in student outcomes and teacher quality was analyzed using averaged classroom-level single-level linear regressions. Finally, differences in teacher effectiveness between higher (top quartile) and low (bottom quartile) SES classrooms, as measured by using the average number of books in the home as a proxy for SES, were examined using Welch’s t-tests (which are not sensitive to sample size). This last analysis used an alternative to the jackknifed standard error approach (designed for the entire sample of classrooms) because it examined a sub-sample that is vulnerable to Type I error.

3.4 Conclusions

In the research presented in this volume, we use a number of teacher- and student-level variables and empirical approaches to examine the relationship between measures of teacher effectiveness and student outcomes. It should be emphasized that issues with cross-national comparability and the lack of truly representative student and teacher samples imposes some limits on these variables and statistical methods. International large-sample student-level longitudinal data, such as TIMSS, lacks some of the teacher measures now commonly used in the US-based studies, making comparisons with existing research difficult. A number of potentially important control variables (not to mention a means of distinguishing school effects from teacher effects) are unavailable in TIMSS. As a consequence, it is very difficult to construct more robust causal estimates of teacher effects in multiple countries across time. The aim of this study was more modest: to use multiple statistical models to explore whether there is a consistent pattern between measures of teacher quality (both characteristics and behavior) and student outcomes across time and space, and assess whether the ambiguous findings of US-based research were replicated on an international scale.

References

- Blank, R. K., & De Las Alas, N. (2009). *The effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders*. Report prepared under a grant to the Council of Chief State School Officers from the National Science Foundation, Grant #REC-0635409. Washington, DC: Council of Chief State School Officers. Retrieved from <https://files.eric.ed.gov/fulltext/ED544700.pdf>.
- Blomeke, S., Olsen, R., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilson & J. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes*, IEA Research for Education (Vol. 2, pp. 21–50). Cham, Switzerland: Springer. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-41252-8_2.
- Chudgar, A., & Luschei, T. (2009). National income, income inequality, and the importance of schools: A hierarchical cross-national comparison. *American Educational Research Journal*, 46(3), 626–658.
- Foy, P. (2017). TIMSS 2015 user guide for the international database. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_UserGuide.pdf.
- Schmidt, W., Burroughs, N., Zoido, P., & Houang, R. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Education Researcher*, 44(4), 371–386.
- Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23(4), 323–355.

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.

