

Chapter 2

Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality

Sigrid Blömeke, Rolf Vegar Olsen and Ute Suhl

Abstract This chapter examines how crucial input and process characteristics of schooling are related to cognitive student outcomes. It was hypothesized that teacher quality predicts instructional quality and student achievement, and that instructional quality in turn predicts student achievement. The strengths of these relations may vary across countries, making it impossible to draw universal conclusions. However, similar relational patterns could be evident within regions of the world. These hypotheses were investigated by applying multi-level structural equation modeling to grade four student and teacher data from TIMSS 2011. The sample included 205,515 students from 47 countries nested in 10,059 classrooms. Results revealed that teacher quality was significantly related to instructional quality and student achievement, whereas student achievement was not well predicted by instructional quality. Certain characteristics were more strongly related to each other in some world regions than in others, indicating regional patterns. Participation in professional development activities and teachers' sense of preparedness were, on average, the strongest predictors of instructional quality across all countries. Professional development was of particular relevance in Europe and Western Asian/Arabian countries, whereas preparedness played an important role in instructional quality in South-East Asia and Latin America. The ISCED level of teacher education was on average the strongest predictor of student achievement across all countries; this characteristic mattered most in the Western Asia/Arabia region.

S. Blömeke (✉) · R.V. Olsen
Faculty of Educational Sciences, Centre for Educational Measurement at the University of Oslo (CEMO), Oslo, Norway
e-mail: sigrid.blomeke@cemo.uio.no

R.V. Olsen
e-mail: r.v.olsen@cemo.uio.no

U. Suhl
Institut Für Erziehungswissenschaften, Humboldt-Universität Zu Berlin, Berlin, Germany
e-mail: ute.suhl@staff.hu-berlin.de

Keywords Instructional quality · Teacher quality · Student achievement · Two-level structural equation modeling mediation models · Trends in Mathematics and Science Study (TIMSS) 2011

2.1 Rationale

The framework of the TIMSS study describes policy malleable features at the system, school, classroom and student level that are known to influence selected desired outcomes of education, such as achievement in the core curricular domain of mathematics (Mullis et al. 2009). Without going into details of the multi-stage sampling procedure applied in TIMSS, a distinguishing feature is that it produces a sample of intact classrooms, including their mathematics teacher(s), representing the 4th grade students in the participating countries (Joncas and Foy 2012). In other words, the data set from TIMSS provides a unique opportunity to link responses from students in a classroom with those from their teacher(s) for a large number of world regions, educational cultures and systems (in the following also called “countries”).

It is well known from previous research that classroom matters. First and foremost, teachers matter (for a summary of the state of research see, for example, Kyriakides et al. 2009). Teachers’ experience, teacher education background, beliefs and motivations, as well as their content knowledge, pedagogical content knowledge, and general pedagogical knowledge (actual and perceived), are characteristics that, to varying degrees, have been shown to have effects on student outcomes. Secondly, teaching or instruction matters for student outcomes (for a summary of research see, for example, Seidel and Shavelson 2007). Educational effectiveness studies and qualitatively oriented classroom observational studies seem to converge on some key features of high quality instruction. In short, high quality teaching consists of instructional practices leading to students being dedicated to cognitively active time on task.

However, there are not many studies seeking to model how teacher quality is related to student achievement, and how teacher quality is put into action by what teachers actually do in the classrooms. This research gap applies particularly to international comparative research. Most of the reported studies of these relationships, although valuable (for example Baumert et al. 2010), took place in one country only, and usually in a Western country. Comparative research that tries to extend the findings from these studies to other educational cultures and systems is lacking. The generalizability of the findings is therefore an open question.

From most definitions of learning it follows that learning occurs as a result of an interaction between the individual learner and his or her surroundings. In the school setting these are, such interactions that most often are generally planned and staged by the teacher. Teacher quality should thus matter, but the degree of its influence may vary by depending on teacher quality indicators or among educational systems. Furthermore, although some aspects of teacher quality have been shown to be

directly positively related to student outcomes, they are also resources for the instructional processes in classrooms, and hence teacher quality may be a predictor of instructional quality. As pointed out above, we know for instance that stronger pedagogical content knowledge of mathematics teachers (one possible indicator of teacher quality) is positively related to student achievement in mathematics (Baumert et al. 2010). This may be a direct effect, where teachers influence individual students by diagnosing their (mis)conceptions and addressing these directly, or it may influence the teachers to create classroom conditions for learning where students are cognitively challenged and activated.

In line with this reasoning, we hypothesized that teacher quality is partly mediated by instructional quality. Although the capacity of TIMSS to address this issue is limited because of its design and instruments, the study has collected a lot of information from the teachers about their background and dispositions. The study has also collected rudimentary information, from both the teachers and the students, about the degree to which the classroom is characterized by instructional activities known from other research to be beneficial for student learning.

Against this background, the following research questions led this study:

- (1) *Which teacher characteristics are significantly related to instructional quality?*
- (2) *To what extent do the relations between teacher quality and instructional quality vary by country? Is it possible to identify regions or clusters of countries where similar relational patterns exist?*
- (3) *Is instructional quality significantly related to student achievement? Does this relation vary by country, and, does a pattern exist that applies to countries from larger regions or cultures?*
- (4) *If teacher quality is significantly related to instructional quality and if instructional quality is significantly related to achievement, does instructional quality partially mediate the relation between teacher quality and student outcomes?*

2.2 Theory

2.2.1 *Educational Effectiveness Research as the Point of Reference*

The studies presented in this book are rooted in the tradition of educational effectiveness research (Sammons 2009; Scheerens and Bosker 1997). The analysis in this chapter seeks to establish the structural relationship between aspects of teacher quality, instructional quality and student outcome with the hypotheses that teacher quality matters significantly positively for instructional quality and student outcomes, that instructional quality matters significantly positively for student

outcomes, and that instructional quality partly mediates the influence of teacher quality on student outcomes. Several models for effective schools have been proposed, all of which to some degree include teacher quality and instructional quality. Our model employed a section of the dynamic model proposed by Creemers and Kyriakides (2008). However, this is a “static” model used to analyze cross-sectional data, and thus should accordingly be seen as a pragmatic conceptualization of the relationship between these core concepts of teaching and learning, reflecting the design and data available from the TIMSS study.

Educational effectiveness research (Nordenbo et al. 2008; Scheerens 2013) relates to an explicit notion of input-process-output logic, usually represented by regression models, where an educational outcome, in our case grade four students’ mathematics achievement, is modelled as a function of one or more independent variables, in our case teacher quality and instructional quality. In most of these models one or more intervening concepts are included, in our case instructional quality, to conceptually relate the modelled variables. In other words, this is empirical research that tries to open up the educational system as a “black-box”, where the input is the amount of resources, conditions or other antecedents hypothesized to be related to variation in the outcome. The complexities of studying the degree to which possible inputs affect an outcome involves variables that relate to one or more of the levels in the education system. TIMSS is designed to provide data where these complexities are represented by data at both the student and the class/teacher level.

Scheerens (2013, pp. 10–12) suggested that the lack of a unifying theoretical model for school research may well reflect that “[t]he complexity of educational ‘production’ may be such that different units and levels are addressed by different theories,” and he concluded his systematic review of the theoretical underpinning of educational effectiveness research by stating “[a]s it comes to furthering educational effectiveness research, the piecemeal improvement of conceptual maps and multi-level structural equation models may be at least as important as a continued effort to make studies more theory driven.” This chapter and the other chapters in this book are intended to provide improvements in the conceptual understanding of what characterizes effective instructional practice. By the inclusion of multiple educational systems, these chapters will also contribute to address questions regarding the degree to which educational effectiveness research can provide models and theories which are sensitive also to the wider social, political and cultural context in which education is embedded.

2.2.2 Teacher Quality

Teacher quality (TQ) includes different indicators of teacher qualifications, in particular characteristics of teachers’ educational background, amount of experience in teaching, and participation in professional development (PD), as well as personality characteristics such as teachers’ self-efficacy. A number of previous

studies were able to relate measures of such teacher characteristics to student educational outcomes (see for instance the review by Wayne and Youngs 2003).

Evidence suggests that the quality of teacher education does have an impact on teachers' educational outcomes in terms of teacher knowledge and skills (Blömeke et al. 2012; Boyd et al. 2009; Tatto et al. 2012); these, in turn, are significantly related to instructional quality and student achievement (Baumert et al. 2010; Hill et al. 2005; Kersting et al. 2012). The degree and major academic disciplines studied can be regarded as indicators of teachers' education, although they are only rough approximations of specific opportunities to learn. In the case of mathematics teachers, a major in mathematics delivers the body of content knowledge necessary to present mathematics to learners in a meaningful way and to connect mathematical ideas and topics to one another, as well as to the learner's prior knowledge and future learning objectives (Wilson et al. 2001; Cochran-Smith and Zeichner 2005). However, knowing the content provides only a foundation for teaching; student achievement is higher if a strong subject-matter background is combined with strong educational credentials (Clotfelter et al. 2007). Correspondingly, teachers' pedagogical content knowledge and content knowledge of mathematics are of great importance for instructional quality and student achievement in mathematics, with the former exerting a greater effect than the latter (Baumert et al. 2010; Blömeke and Delaney 2012). Whether teachers had an education where mathematics or mathematics education were a major focus and the type of degree are proxy variables available in TIMSS. This makes it possible to study how teachers' educational background may affect teaching and students' achievement across countries.

An almost universal characteristic seems to be that teachers do not feel sufficiently prepared for their complex tasks, in particular during the first years on the job (Kee 2012). TIMSS developed three constructs reflecting teachers' preparedness to teach numbers, geometry and data, respectively. The constructs were developed within the context of Bandura's social-cognitive theory, and the measures of teachers' preparedness for teaching may reasonably be assumed to reflect a concept which is similar to teacher self-efficacy (Bandura 1986; Pajares 1996). Self-efficacy beliefs influence thought patterns and emotions, which in turn enable or inhibit actions. Teachers with strong self-efficacy are typically more persistent and make stronger efforts to overcome classroom challenges than others (Tschannen-Moran et al. 1998). TIMSS provides data about teachers' sense of preparedness so that the relation of this dimension of self-efficacy can be examined across countries.

In almost all countries, a variety of professional development activities exist, from very short classes to comprehensive programs (Goldsmith et al. 2014; Guskey 2000). These include school-based programs, and coaching, seminars, or other types of out- and in-service training with the aim of supporting the development of teacher competencies. Overall, meta-analyses support the hypothesis that professional development is positively related to instructional quality and student achievement if the activities meet certain quality characteristics (Timperley et al. 2007). Desimone (2011) classified these quality features into a focus on content,

active learning, coherence, and a certain minimum length of the professional development course to be sustainable and collaborative activities. Collaboration in terms of joint work on cases and practicing under supervision of colleagues seems to be particularly relevant (Boyle et al. 2005). Discussions, reflection and continuous feedback seem to stimulate real changes in beliefs and routines (Goldsmith et al. 2014). TIMSS included several scales that assessed both teachers' participation in formal professional development activities and their involvement in continuous and collaborative professional development activities with colleagues in the school.

2.2.3 *Instructional Quality*

Several studies have established a relationship between measures of instructional quality (InQua) and student achievement, student motivation or other outcomes of schooling. Even though the concept of instructional quality is understood differently by different researchers in the field of educational effectiveness research, there is agreement that it is a multidimensional construct (Baumert et al. 2010; Creemers and Kyriakides 2008). Besides classroom management, three instructional characteristics, namely cognitive activation, clarity of instruction, and a supportive climate, are regarded as essential features (Rakoczy et al. 2010; Decristan et al. 2015). TIMSS includes several measures relating to different aspects of instructional quality, with responses both from teachers and students. For more about the theoretical framework of this construct see Chap. 1.

2.2.4 *Universal, Cultural or Country-Specific Models?*

National specifications of degrees and licenses, foci of programs in terms of majors, amount of in-service training and length and level of teacher education reflect partly overlapping and partly differing visions of the knowledge and skills that teachers are expected to have in a country (Schwille et al. 2013). These specifications of what is required of mathematics teachers before they are allowed to teach mathematics to students at grade four can be assumed to be intentionally developed by national educational policy makers and teacher education institutions (Stark and Lattuca 1997). The same applies to professional development activities provided to teachers or to characteristics regarded as high quality teaching in a country.

In his study of primary school education in England, France, India, Russia, and the United States, Alexander (2001) illustrated the subtle and long-term relationship between culture and pedagogy. Based on videotaped lessons and interviews with teachers, he demonstrated that opportunities to learn provided during schooling reflected a country's educational philosophy transmitted and mediated through the classroom talk between teachers and students. Leung (2006) confirmed similar

cultural differences, specifically with respect to mathematics education in the East and the West. Although mathematics can be regarded as a fairly global construct (Bishop 2004), the curricula of school mathematics, as well as of mathematics teacher education, differ across countries, and are influenced by the context in which they are implemented (Blömeke and Kaiser 2012; Schmidt et al. 1997). With this as a backdrop, it is interesting that a study like TIMSS permits examination of the extent to which the relationship between teacher quality, instructional quality and student achievement can be generalized across the world, or across regions of the world.

2.2.5 Control Variables

Current research indicates that in some countries gender differences in students' mathematics achievement still exist, but that these vary in their direction (Mullis et al. 2012). There is an even stronger relationship between students' socioeconomic background and achievement (Mullis et al. 2012). In order to estimate the relation of teacher quality and instructional quality to mathematics achievement of students at grade four, the background characteristics of students need to be controlled for in the analysis.

2.3 Methods

2.3.1 Sample

This study is based on grade four student and teacher data from the majority of countries participating in TIMSS 2011. Five countries were excluded because there were no data on one or more predictors (Austria, Belgium, Kazakhstan and Russia) or there were very high levels of missing values for most of the variables included in the analysis (Australia). For students with more than one mathematics teacher, data from only one of the teachers was included at random, resulting in a data set with a simple hierarchical structure, where students were nested in one specific class with one specific teacher. The amount of data excluded by this procedure was negligibly small (for details see Chap. 1). The final sample included 205,515 students from 47 countries nested in 10,059 classrooms/teachers with an average classroom size of 20 students. Student sample sizes per country varied between 1423 and 11,228, with the number of classrooms/teachers ranging from 67 to 538, and an average classroom size between 12 and 34 students. The school level was neglected in the analyses to avoid overly complex hierarchical models. Furthermore, the choice of omitting the school level in the analysis is based on the fact that for many countries the classroom and school level cannot be analyzed separately, since only one grade four classroom was drawn per school.

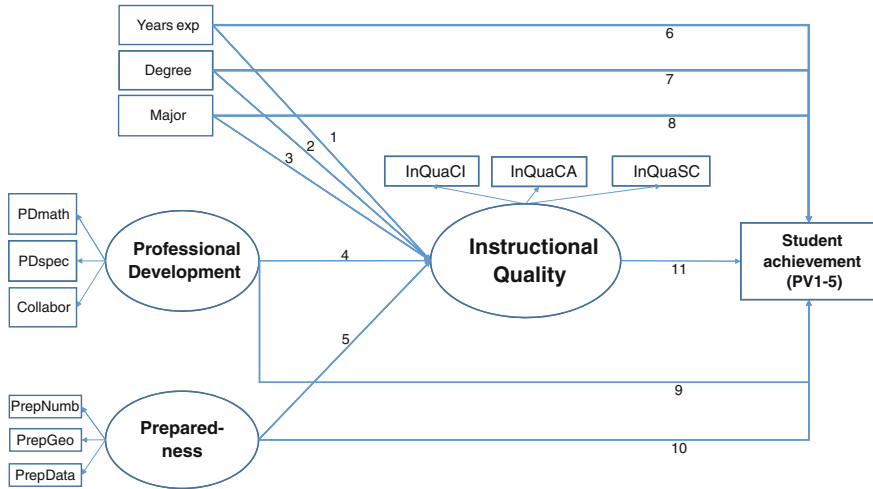


Fig. 2.1 Model of the hypothesized relations of teacher quality (*left hand side* of the figure) in terms of years of teaching experience (*Years exp*), teacher education degree (*Degree*), major focus of teacher education (*Major*), professional development represented by three indicators (*PDmath*, *PDspec* and *Collabor*), and sense of preparedness represented by three indicators (*PrepNumb*, *PrepGeo* and *PrepData*), to instructional quality (*InQuaCI*, *InQuaCA*, and *InQuaSC*), and to student achievement represented by five plausible values (*PV1-5*; *right hand side* of the figure); all abbreviations are explained in Table 2.1, and the numbers linking the relations hypothesized correspond to columns in Table 2.2, where the actual estimates can be found

2.3.2 Variables

A structural model was developed to reflect the hypothesized relations between teacher quality, instructional quality and student achievement (Fig. 2.1). Furthermore, the internationally-pooled descriptives of all variables, including their range across countries were inspected (Table 2.1).¹

Teacher quality measures

Teacher quality is represented by three central dimensions in our model, namely teacher education background, participation in professional development (PD) activities, and teachers’ sense of preparedness. Teacher education background is described by teachers’ years of experience and their formal initial education. These characteristics were included as separate categorical and manifest variables because they do not reflect a joint and theoretically derived latent construct. Instead they represent different and not necessarily related dimensions of teacher quality.

¹For country-specific descriptives including information about their distribution in terms of skewness and kurtosis see Appendices A and B; for more details about the item format see the TIMSS data analysis manual (Foy et al. 2013).

Table 2.1 Descriptives of the variables used in the model

Description of item or item parcel	Variable label in Fig. 2.1	Label in the international database ^a	Mean (SD) and [range of means across countries] ^b	Reliability (coefficient alpha) for item parcels	Percentage missing data and [range across countries]
Number of years of experience	Years exp	ATBG01	“More than 20 years” [“Less than 5 years”–“More than 20 years”] ^c		7 [1–22]
Level of formal education completed	Degree	ATBG04	“Finished ISCED level 5A, first degree” [“ISCED 3”–ISCED 5A, second”] ^c		5 [0–21]
Focus on either mathematics OR mathematics education	Major	ATBG05AC ATBG05BA	0.39 (0.40) [0.04–0.97]		7 [0–41]
PD in mathematics instruction (broad activities)	PDmath	ATBM11A to C	0.43 (0.37) [0.11–0.78]	0.79	7 [0–26]
PD in mathematics instruction (specific challenges)	PDspec	ATBM11D to F	0.37 (0.34) [0.13–0.66]	0.65	7 [0–26]
Collaborative school-based PD with peers	Collabor	ATBG10A to ATBG10E	0.64 (0.30) [0.37–0.94]	0.80	4 [0–19]
Preparedness to teach numbers	PrepNumb	ATBM12AA to ATBM12AH	0.93 (0.13) [0.74–0.99]	0.89	7 [0–27]
Preparedness to teach geometry	PrepGeo	ATBM12BA to ATBM12BG	0.90 (0.15) [0.72–0.97]	0.87	7 [0–27]
Preparedness to teach data	PrepData	ATBM12CA to ATBM12CC	0.90 (0.18) [0.70–0.98]	0.92	14 [1–60]
Instructional quality: Clarity of instruction	InQuaCI	ATBG15A and ATBG15C	0.88 (0.15) [0.68–0.96]	– ^d	4 [0–19]

(continued)

Table 2.1 (continued)

Description of item or item parcel	Variable label in Fig. 2.1	Label in the international database ^a	Mean (SD) and [range of means across countries] ^b	Reliability (coefficient alpha) for item parcels	Percentage missing data and [range across countries]
Instructional quality: Cognitive activation	InQuaCA	ATBG15B and ATBG15F	0.73 (0.19) [0.55–0.87]	– ^d	4 [0–19]
Instructional quality: Supportive climate	InQuaSC	ATBG15D and ATBG15E	0.94 (0.12) [0.78–0.99]	– ^d	4 [0–19]
Student achievement: five plausible values		ASMMAT01 to ASMMAT05	500 (100) [248–606]		

International mean values were computed by averaging country means

Note PD = professional development, SD = standard deviation

^aRefers to the labels in the TIMSS 2011 user guide for the international database (Foy et al. 2013)

^bAll scales transformed to a 0–1 scale representing proportion of maximum score for the scale

^cModal category across countries

^dFor parcels with only two items coefficient alpha is not meaningful

The variation between countries for these variables was remarkably large. Across all countries, the modal category of number of years of experience (“By the end of this school year, how many years will you have been teaching altogether?”) was more than 20 years. The Eastern European countries were particularly pronounced in having many teachers with extensive teaching experience, indicating an older teaching force than elsewhere (see Appendix A, Table A.1). But there were also countries in the data set where the largest group of teachers that taught mathematics at grade four had less than 10 years of experience, and, in some countries, less than 5 years of experience. The Arabian countries were most pronounced in having a relatively young teaching force.

Teachers provided information about their degree from teacher education (“What is the highest level of formal education you have completed?”) out of six options from “did not complete ISCED level 3” to “finished ISCED level 5A, second degree or higher”. Across all countries, the modal category was “ISCED level 5A, first degree”, indicating that many countries had a large proportion of teachers with a bachelor degree. But there were also some countries where the largest group of teachers did not have university degrees, but had completed practically-based programs at ISCED level 3. Italy and the African countries were most pronounced in this respect (see Appendix A, Table A.2). In contrast, there were countries where the largest group of teachers held a university degree at least equivalent to a master

degree (“ISCED level 5A, second degree or higher”). The Eastern European countries were most pronounced in this respect.

A dichotomous variable was created by combining teachers’ responses to two questions regarding their specialization in mathematics. This variable identifies teachers with a major in mathematics or in mathematics education (“During your <post-secondary> education, what was your major or main area(s) of study?” and “If your major or main area of study was education, did you have a <specialization> in any of the following?”). On average, slightly fewer than 40 % of all teachers across all countries had a major with a specialization in mathematics. However, in some countries the proportion was below 10 % (for example in some of the Eastern European countries), whereas in other countries the proportion was more than 80 % (for example in several Arabian countries) (see Appendix A, Table A3).

Furthermore, there were measures of teachers’ participation in PD activities. One set of questions asked the teachers whether or not they had participated in PD during the last two years. These questions are represented in the model by two item parcels reflecting either broad PD activities covering, for example, “mathematics content” in general, or reflecting PD activities preparing for specific challenges, for example “integrating information technology into mathematics”. Across all countries, approximately 40 % of the teachers had participated in broad or specific PD activities, respectively. However, the between-country variation was large, from countries having as few as 10 % the teachers taking part in broad or specific PD, to countries where more than two-thirds of the teachers had taken part in one or both forms of PD activities. It is difficult to discern any systematic cultural pattern in these differences (see Appendix A, Table A.4).

In addition, there was a set of questions regarding whether teachers had taken part in collaborative activities representing continuous, collaborative and school-based PD (“How often do you have the following types of interactions with other teachers?”, with “Visit another classroom to learn more about teaching” as an “exemplary” form of interaction). Across all countries, teachers commonly participated in these types of activities two to three times each month. However, in some countries the largest group of teachers participated in collaborative PD daily or almost daily. These questions were included as the third item parcel defining the latent construct of PD.²

The third teacher quality dimension included in the model reflects teachers’ self-efficacy. The indicator used was their self-reported sense of preparedness to teach specific topics in mathematics within the three domains of number, geometric shapes and measures, as well as data display (“How well prepared do you feel you are to teach the following mathematics topics?”, with “Adding and subtracting with

²The TIMSS data set includes an IRT-based construct composed of these items, labelled as Collaborate to Improve Teaching (CIT). For the purpose of being able to interpret the mean and range in country comparisons in the same way as the other two parcels, we therefore opted for a classical mean raw score used as a third item parcel, each representing different aspects of PD. Furthermore, we were able to confirm measurement invariance of the latent construct PD with this indicator.

decimals” included as an exemplary topic). For each domain, teachers were asked to rate these topics on a three-point Likert scale from “Not well prepared” (0) to “Very well prepared” (2). Teachers were also invited to use a “not applicable” response category if the topic was not covered in their curriculum. In our analysis, the items marked as not applicable were treated as missing. To simplify the final model, the three domains were represented as item-parcel indicators of the latent construct of preparedness. Across all countries, the mean of the three item parcels was each time around 1.8 and, thus, close to the maximum category of the Likert scale. This suggests that there was little discrimination evident in the items. The international variation was also more limited within this dimension than in others included in the model. The lowest means were around 1.5 and, thus, straddled the categories “Somewhat prepared” and “Very well prepared”. Interestingly, slightly lower self-efficacy was most evident in Japan and Thailand (see Appendix A, Table A.5).

Instructional quality measures

The measure of InQua applied in this chapter is based on the teacher questionnaire in TIMSS where six questions asked teachers to report how often they perform various activities in this class (“How often do you do the following in teaching this class?”). This measure was preferred over other measures available (see Sect. 2.5) since it has a more explicit relation to three of the four characteristics of high quality instruction (Table 2.1). Teachers were asked to rate these activities on a four-point Likert scale from “Never” (0) to “Every or almost every lesson” (3). These items are represented by three item parcels with two items in each parcel covering different aspects of the latent construct InQua. The first parcel reflected teaching characteristics that were intended to deepening students’ understanding through clear instruction (such as “Use questioning to elicit reasons and explanations”). The second parcel pursued this objective through cognitive activation (through questions such as “Relate the lesson to students’ daily lives”). The final parcel covered a supportive climate (for example “Praise students for good effort”). Across all countries, the indicators for a supportive climate appeared to be widely present, as the mean was close to the maximum of the scale. The mean of the other two parcels was slightly lower. Interestingly, Scandinavian countries had the lowest means on the cognitive-activation item-parcel (see Appendix A, Table A.6). Some international variation existed on all three item parcels.

Outcome measure

We selected student achievement in mathematics represented by five plausible values as our outcome measure. The scale was defined by setting the international mean to 500 and the standard deviation to 100. Country means varied between 248 and 606 points, which is a difference of more than 3.5 standard deviations (for more information, see Martin and Mullis 2012).

Control variables

Data about gender and socioeconomic background were gathered through students' self-reports to the questions "Are you a girl or a boy?" and the frequently used proxy measure of home background "About how many books are there in your home?"³

2.3.3 Analysis

The research questions were examined using multi-level structural equation modeling (MLSEM). The intra-class correlation (ICC) for students' achievement in the pooled international data set (ICC = 0.30) and within countries (ICC = 0.07–0.56) were all above the threshold at which multi-level modeling is recommended (Snijders and Bosker 2012).

Item-parcels were used as indicators, as recommended when structural characteristics of the constructs are the focus of interest (Little et al. 2002), as applies in the present investigation, and when sample size is limited in comparison to the number of parameters to be estimated (Bandalos and Finney 2001). The latter also applies to the present investigation given that there are only about 140 to 260 classrooms in most of the countries. By using parcels as indicators for the latent variables, the number of free parameters to be estimated was significantly reduced. The items were combined into parcels based on theoretical expectations confirmed by initial exploratory analysis of sub-dimensions in the latent variables included in the model.

Data analysis was carried out using the software MPlus 7.4. The clustered data structure was taken into account by using a maximum-likelihood estimator with robust sandwich standard errors to protect against being too liberal (Muthén and Muthén 2008–2012). Missing data were handled by using the full-information-maximum-likelihood (FIML) procedure. The model fit was evaluated with the chi-square deviance and a range of fit indices.⁴

³The TIMSS data set includes an index representing socioeconomic background in terms of Home Educational Resources (HER) that includes also other indicators such as parental income, occupation and education level. Nevertheless, we opted for using "books at home" because in contrast to HER this variable has remained unaltered for many cycles and very similar indicators of home background are used in all other international large-scale studies. This makes it easier to compare results with previous research. Moreover, "books at home" has been and still is a powerful predictor of achievement (compared to parents' education, which is part of HER).

⁴The fit indexes were evaluated to the following commonly recommended criteria: a ratio of chi-square deviance and degrees of freedom of <2 indicates a very good model fit, estimates <3 indicate a good fit. Estimates of the comparative fit index (CFI) and the Tucker Lewis index (TLI) >0.95 indicate a very good model fit, and estimates >0.90 indicate a good model fit (Hu and Bentler 1999). Estimates of the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) <0.05 indicate a very good model fit, and estimates <0.08 indicate a good model fit.

Before the final model was run, measurement invariance (MI) across countries was tested for the latent constructs in the model. Comparing constructs and their relations across countries produces meaningful results only if the instruments measure the same construct in all countries (Van de Vijver and Leung 1997). In order to ascertain such equivalence, MI was established using multiple-group confirmatory factor analysis (MG-CFA; Chen 2008). As instructional quality and the teacher constructs were measured at the classroom level, we tested for measurement invariance at the school level. Firstly, configural invariance was examined, which means that in each country the same items had to be associated with the same latent factors. As a second step, we tested for metric invariance, by studying whether the factor loadings were invariant across countries. Invariance of factor loadings enabled us to compare the relationship between latent variables across groups. It was possible to establish metric invariance for all latent constructs included in the present model (see Appendix B).

To examine our research questions, a single-group model was first applied before country-by-country analyses were carried out. In the multi-group model, factor loadings were constrained to be the same for all countries, reflecting the metric invariance criterion referred to above, in order to ensure comparability. Indirect relations at the between-level were estimated by multiplying the coefficients for the respective direct relation. In the single-group model, the two control variables gender and books at home were grand-mean centered on the international mean, whereas all predictors, the mediator InQua and the dependent variable student achievement in mathematics were group-mean centered on the country means. In the multi-group model the control variables were again grand-mean centered (this meant now on the country mean) whereas the predictors, the mediator and the dependent variable remained unaltered. Relations were regarded significant on the within-level if $p < 0.05$, but given the relative small number of units at the between-level as compared to the number of parameters to be estimated, a more liberal decision rule for the significance testing with $p < 0.10$ was applied for this level.

2.4 Results

2.4.1 Model Fit

The fit of the pooled model to the full data set was very good; both with respect to relative and to absolute fit indices (see Table 2.2). Only the ratio of the chi-square deviance to the degrees of freedom was unsatisfactory which is commonly observed with large samples. Within countries, the model fit varied substantially but given the small sample sizes the fit was sufficient on most indices in the majority of countries. Only in nine out of the 47 countries more than two of the applied indices indicated an unsatisfactory model fit. Typically for these cases, the CFI and TLI estimates were below the threshold of 0.90 and the SRMR estimate on the between-level above the 0.08 criterion.

Table 2.2 Sample size, intra-class correlation (ICC) and model fit of the pooled and the country-by-country models

Country	Number of students	Number of classes	\emptyset cluster	ICC math	χ^2	df	χ^2/df	RMSEA	SRMR_BW	CFI	TLI
Pooled model	205,515	10,059	20.4	0.30	232.71	60	3.9	0.00	0.04	0.97	0.97
<i>Western Asian/Arabian countries</i>											
Armenia	4433	205	21.6	0.27	78.24	60	1.3	0.01	0.07	0.94	0.93
Azerbaijan	3574	226	15.8	0.46	110.94	60	1.9	0.02	0.09	0.82	0.77
Bahrain	3656	150	24.4	0.27	100.29	60	1.7	0.01	0.10	0.85	0.81
Georgia	4322	220	19.6	0.36	121.16	60	2.0	0.02	0.08	0.78	0.72
Iran	5271	229	23.0	0.35	75.45	60	1.3	0.01	0.07	0.96	0.95
Kuwait	3401	177	19.2	0.14	93.72	60	1.6	0.01	0.06	0.85	0.81
Oman	9584	389	24.6	0.20	79.92	60	1.3	0.01	0.07	0.96	0.95
Qatar	3485	169	20.6	0.44	107.24	60	1.8	0.02	0.07	0.91	0.88
Saudi Arabia	4202	189	22.2	0.36	54.65	60	0.9	0.00	0.06	1.00	1.02
Turkey	6896	250	27.6	0.38	116.92	60	2.0	0.01	0.08	0.89	0.85
United Arab Emirates	11,228	467	24.0	0.45	107.86	60	1.8	0.01	0.06	0.90	0.87
Yemen	6357	190	33.5	0.42	68.20	60	1.1	0.01	0.08	0.90	0.87
<i>African countries</i>											
Botswana	3392	126	26.9	0.36	99.93	60	1.7	0.01	0.09	0.90	0.87
Morocco	4186	186	22.5	0.51	108.80	60	1.8	0.01	0.09	0.87	0.83
Tunisia	3090	149	20.7	0.25	90.01	60	1.5	0.01	0.09	0.91	0.88
<i>Latin American countries</i>											
Chile	4695	174	27.0	0.38	92.86	60	1.6	0.01	0.09	0.82	0.77
Honduras	2671	120	22.3	0.46	84.45	60	1.4	0.01	0.09	0.87	0.83
<i>European countries</i>											
Croatia	4004	260	15.4	0.12	99.08	60	1.7	0.01	0.08	0.92	0.90
Czech Republic	4363	228	19.1	0.14	87.00	60	1.5	0.01	0.08	0.96	0.95

(continued)

Table 2.2 (continued)

Country	Number of students	Number of classes	\emptyset cluster	ICC math	χ^2	df	χ^2/df	RMSEA	SRMR_BW	CFI	TLI
Denmark	3355	187	17.9	0.14	106.95	60	1.8	0.02	0.08	0.91	0.88
Finland	4053	241	16.8	0.12	105.21	60	1.8	0.01	0.08	0.91	0.89
Germany	3102	171	18.1	0.19	107.09	60	1.8	0.02	0.08	0.92	0.90
Hungary	4919	243	20.2	0.25	82.03	60	1.4	0.01	0.06	0.97	0.96
Italy	2360	137	17.2	0.28	63.22	60	1.1	0.01	0.07	0.99	0.98
Lithuania	4540	273	16.6	0.19	147.20	60	2.5	0.02	0.06	0.89	0.86
Netherlands	2038	110	18.5	0.14	104.54	60	1.7	0.02	0.08	0.91	0.88
Norway	2841	184	15.4	0.15	68.67	60	1.1	0.01	0.08	0.98	0.97
Poland	4821	249	19.4	0.10	142.93	60	2.4	0.02	0.09	0.86	0.82
Portugal	3674	221	16.6	0.37	110.29	60	1.8	0.02	0.08	0.87	0.84
Romania	3523	193	18.3	0.38	118.40	60	2.0	0.02	0.09	0.83	0.78
Serbia	4219	214	19.7	0.23	73.20	60	1.2	0.01	0.07	0.96	0.95
Slovak Republic	5331	301	17.7	0.27	115.98	60	1.9	0.01	0.07	0.93	0.91
Slovenia	4350	239	18.2	0.07	90.71	60	1.5	0.01	0.07	0.94	0.93
Spain	3715	181	20.5	0.17	109.22	60	1.8	0.02	0.09	0.91	0.88
Sweden	3554	211	16.8	0.09	82.33	60	1.4	0.01	0.09	0.96	0.95
<i>English-speaking countries</i>											
England	2499	134	18.6	0.22	98.29	60	1.6	0.02	0.08	0.91	0.89
Ireland	4303	215	20.0	0.16	117.52	60	2.0	0.02	0.09	0.85	0.81
New Zealand	5189	418	12.4	0.29	162.90	60	2.7	0.02	0.07	0.81	0.76
Northern Ireland	2909	151	19.3	0.11	78.86	60	1.3	0.01	0.07	0.98	0.97
USA	10,460	538	19.4	0.26	100.13	60	1.7	0.01	0.05	0.96	0.95

(continued)

Table 2.2 (continued)

Country	Number of students	Number of classes	\emptyset cluster	ICC math	χ^2	df	χ^2/df	RMSEA	SRMR_BW	CFI	TLI
<i>South-east Asian countries</i>											
Hong Kong	3504	129	27.2	0.39	112.44	60	1.9	0.02	0.08	0.86	0.81
Japan	3651	156	23.4	0.06	104.69	60	1.7	0.01	0.10	0.94	0.92
Korea	4026	142	28.4	0.04	84.29	60	1.4	0.01	0.09	0.97	0.96
Malaysia	1423	67	21.2	0.09	136.79	60	2.3	0.03	0.11	0.65	0.55
Singapore	6044	338	17.9	0.56	152.34	60	2.5	0.02	0.08	0.84	0.79
Chinese Taipei	4218	154	27.4	0.07	98.71	60	1.7	0.01	0.10	0.95	0.93
Thailand	4084	158	25.8	0.43	92.21	60	1.5	0.01	0.07	0.93	0.91

Note ICC = intra-class correlation, df = degrees of freedom, RMSEA = root mean square error of approximation, SRMR_BW = standardized root mean square residual for the between level, CFI = comparative fit index, TLI = Tucker-Lewis index

2.4.2 *Relation Between Teacher Quality, Instructional Quality and Mathematics Achievement*

The pooled model using the data from all countries reveals that participation in PD activities and teachers' sense of preparedness were the strongest predictors of InQua (see Table 2.2), with relatively large effect sizes given that the directions of relations typically vary across countries. Effect sizes around $\beta = 0.20$ may therefore be a first indication of a widely recognizable, if not universal, pattern. This is supported by the country-by-country results. In almost half of the countries PD activities (23 countries) and preparedness (22) were significantly related to InQua, with moderately strong effect sizes ($\beta = 0.61$ or $\beta = 0.50$ respectively), all of which were uniformly positive. Whereas PD activities were related to InQua particularly in European (11 out of 18) and Western Asian/Arabian (7 out of 12) countries, teachers' sense of preparedness was significantly associated with InQua in South-East Asia (4 out of 7), Latin America (2 out of 2) and the Scandinavian (4 out of 5) countries. The relevance of the predictor preparedness was also evident through its somewhat weaker, but still statistically significant relation to student achievement.

Another predictor that influenced InQua and students' mathematics achievement was teachers' experience. On average, across countries, students with higher mathematics achievement were taught by more experienced teachers, and teachers with more experience also reported higher instructional quality. However, for both of these relationships there were also significant effects in the opposite direction for a number of countries, which contradicts the hypothesized relationship.

Teachers' level of education was not associated with InQua in the pooled data set, but a significant positive relationship was found in nine countries. However, students who were taught by teachers with relatively higher ISCED levels performed somewhat higher in the mathematics achievement test, and this positive relationship was also confirmed for twelve of the countries. This characteristic was most prominent in the Western Asia/Arabia-region, although with moderate effect sizes.

Whether a teacher education program had had a major focus on mathematics or mathematics education did not significantly predict InQua. Still, as with teacher education level, students in classrooms demonstrating stronger mathematics achievement were in the overall international analysis more often taught by a teacher who had majored in one of these fields. Within countries, these relationships were mostly insignificant, but we found also both moderate significant positive and negative coefficients in some countries (Table 2.3).

Across all countries, mathematics achievement of students at grade four was not predicted by InQua, and within countries the predictor had a significant relation to achievement in only three countries. As a result, the mediation effect of InQua was negligible and thus the hypothesized mediation effect of InQua on student achievement is not supported by the data included in this analysis.

Table 2.3 Results for the single-group pooled model and the country-by-country models. (Numbers in column headers refer to relations displayed in Fig. 2.1)

Country	InQua on indicators of TQ										Achievement on TQ and InQua										Indirect effects of TQ × InQua on Student achievement					Level 1 variables		
	1	2	3	4	5	6	7	8	9	10	11	1 × 6	2 × 6	3 × 6	4 × 6	5 × 6	Girl	Books	Books									
Pooled model	0.08	-0.00	-0.00	0.21	0.18	0.08	0.05	0.05	0.00	0.06	0.02	0.00	-0.00	-0.00	0.00	0.00	-0.04	0.21	0.21									
<i>Western Asian/Arabian countries</i>																												
Armenia	-0.23	0.12	-0.19	0.21	0.19	0.09	0.27	0.05	0.10	0.09	-0.10	0.02	-0.01	0.02	-0.02	0.03	0.08	0.08	0.08									
Azerbaijan	0.02	0.49	0.31	-0.12	0.33	0.08	-0.01	-0.13	0.23	-0.06	0.24	0.00	0.12	0.07	-0.03	0.08	0.05	0.11	0.11									
Bahrain	-0.08	0.16	-0.16	0.28	0.19	0.06	0.22	-0.48	0.12	0.11	-0.01	0.00	0.00	0.00	0.00	0.06	0.06	0.05	0.05									
Georgia	0.07	0.07	0.06	-0.16	0.07	-0.13	-0.06	0.03	-0.04	0.17	0.15	0.01	0.01	0.01	-0.02	0.01	0.03	0.16	0.16									
Iran	0.14	0.06	0.16	0.21	0.17	0.23	0.12	0.12	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	0.14	0.14									
Kuwait	0.07	-0.13	0.19	0.43	0.12	0.01	-0.16	-0.08	-0.10	0.07	0.19	0.01	-0.02	0.04	0.08	0.02	0.17	-0.01	-0.01									
Oman	0.02	0.21	-0.19	0.29	0.17	0.11	0.09	0.01	0.13	-0.00	0.29	0.01	0.06	-0.06	0.08	0.05	0.12	0.13	0.13									
Qatar	0.12	0.13	0.16	0.49	0.01	0.26	0.25	-0.26	-0.08	-0.02	0.09	0.01	0.01	0.01	0.04	0.00	0.03	0.02	0.02									
Saudi Arabia	-0.30	-0.17	-0.03	0.61	0.22	0.13	-0.00	0.04	-0.01	0.07	0.23	-0.07	-0.04	-0.01	0.14	0.05	0.06	0.06	0.06									
Turkey	0.06	-0.03	0.07	0.22	0.10	0.45	0.18	0.08	0.17	0.10	0.11	0.01	0.00	0.01	0.02	0.01	-0.03	0.17	0.17									
United Arab Emirates	0.13	-0.06	0.15	0.14	0.22	-0.06	0.05	-0.37	-0.13	0.05	0.15	0.02	-0.01	0.02	0.02	0.03	0.02	0.05	0.05									
Yemen	-0.13	-0.03	0.05	0.29	0.38	-0.08	0.02	0.04	0.17	-0.16	0.28	-0.04	-0.01	0.01	0.08	0.11	0.01	-0.02	-0.02									
<i>African countries</i>																												
Botswana	-0.20	-0.25	0.02	0.32	0.16	0.19	0.19	-0.05	0.41	0.04	-0.08	0.02	0.02	0.00	-0.03	-0.01	0.13	0.00	0.00									
Morocco	0.34	0.18	-0.02	0.08	-0.12	-0.24	0.03	0.18	0.01	0.06	-0.11	-0.04	-0.02	0.00	-0.01	0.01	0.03	0.04	0.04									
Tunisia	0.29	0.53	-0.06	0.09	0.50	0.13	-0.11	0.07	-0.06	0.05	-0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.06									
<i>Latin American countries</i>																												
Chile	0.13	-0.15	-0.10	0.08	0.39	0.04	0.03	0.09	0.25	0.22	-0.11	-0.01	0.02	0.01	-0.01	-0.04	-0.08	0.09	0.09									
Honduras	0.31	-0.24	0.25	0.20	0.48	-0.09	0.20	0.02	-0.12	-0.08	0.17	0.05	-0.04	0.04	0.03	0.08	-0.11	0.01	0.01									

(continued)

Table 2.3 (continued)

Country	InQua on indicators of TQ										Achievement on TQ and InQua										Indirect effects of TQ × InQua on Student achievement						Level 1 variables	
	1	2	3	4	5	6	7	8	9	10	11	1 × 6	2 × 6	3 × 6	4 × 6	5 × 6	Girl	Books										
<i>South-east Asian countries</i>																												
Hong Kong	-0.04	-0.05	0.20	0.25	0.33	0.24	0.14	-0.02	-0.08	0.01	0.35	-0.01	-0.02	0.07	0.09	0.12	-0.11	0.15										
Japan	0.17	-0.10	0.01	0.19	0.15	0.04	0.00	0.04	0.08	0.04	-0.02	0.00	0.00	0.00	0.00	0.00	-0.03	0.29										
Malaysia	-0.25	-0.16	0.47	-0.03	0.08	0.04	0.32	0.07	0.18	0.27	-0.12	0.03	0.02	-0.06	0.00	-0.01	-0.05	0.14										
Singapore	-0.01	-0.01	0.11	0.31	0.21	0.00	0.12	0.05	0.11	0.04	-0.01	0.00	0.00	0.00	0.00	0.00	-0.05	0.12										
South Korea	-0.09	-0.08	0.08	0.29	0.06	0.19	0.07	0.25	0.07	0.09	0.07	-0.01	-0.01	0.01	0.02	0.00	-0.05	0.35										
Chinese Taipei	0.04	0.06	0.13	0.18	0.16	0.08	-0.00	0.16	-0.03	-0.05	-0.05	0.00	0.00	-0.01	-0.01	-0.01	-0.00	0.33										
Thailand	0.18	0.04	-0.09	0.05	0.36	-0.07	-0.11	0.06	-0.10	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.11	0.06										
<i>Central and Eastern European countries</i>																												
Croatia	0.03	0.10	-0.11	0.27	0.19	0.13	0.03	-0.12	0.07	-0.02	0.13	0.00	0.01	-0.01	0.04	0.02	-0.10	0.24										
Czech Republic	0.25	-0.02	-0.08	0.24	0.09	-0.11	0.22	0.03	-0.23	-0.10	0.26	0.07	-0.01	-0.02	0.06	0.02	-0.11	0.32										
Hungary	0.02	-0.03	-0.42	0.17	0.08	0.14	0.01	-0.07	0.15	0.01	-0.13	0.00	0.00	0.05	-0.02	-0.01	-0.05	0.37										
Lithuania	-0.08	0.01	-0.10	0.31	0.13	-0.03	0.10	-0.06	-0.19	0.15	0.18	-0.01	0.00	-0.02	0.06	0.02	-0.04	0.23										
Poland	0.01	0.11	0.18	0.25	-0.14	0.09	0.07	-0.01	-0.07	0.04	0.01	0.00	0.00	0.00	0.00	0.00	-0.10	0.32										
Romania	-0.05	-0.19	0.09	0.07	0.09	0.09	0.21	-0.12	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00	-0.04	0.28										
Serbia	-0.16	0.09	-0.00	0.22	0.22	-0.02	0.25	0.05	0.13	0.25	-0.08	0.01	-0.01	0.00	-0.02	-0.02	-0.05	0.22										
Slovak Republic	0.16	-0.01	0.06	0.15	0.10	0.01	0.06	-0.07	0.08	-0.12	0.02	0.00	0.00	0.00	0.00	0.00	-0.08	0.34										
Slovenia	0.25	0.13	0.12	0.39	0.27	0.26	0.13	0.12	-0.09	0.09	-0.25	-0.06	-0.03	-0.03	-0.10	-0.07	-0.09	0.32										
<i>Scandinavian countries</i>																												
Denmark	-0.08	0.13	0.12	0.20	0.19	0.05	0.05	-0.03	-0.03	-0.28	0.08	-0.01	0.01	0.01	0.02	0.02	-0.06	0.31										
Finland	0.01	-0.08	0.03	0.11	0.39	0.03	0.10	0.06	0.13	0.04	0.16	0.00	-0.01	0.00	0.02	0.06	-0.07	0.25										
Norway	-0.01	0.22	-0.07	0.08	0.32	0.10	-0.02	0.10	-0.07	-0.02	0.17	0.00	0.04	-0.01	0.01	0.05	-0.07	0.24										
Sweden	-0.25	-0.09	0.02	0.18	0.22	0.09	0.21	-0.16	-0.05	0.06	0.07	-0.02	-0.01	0.00	0.01	0.02	-0.08	0.32										

(continued)

Table 2.3 (continued)

Country	InQua on indicators of TQ						Achievement on TQ and InQua						Indirect effects of TQ × InQua on Student achievement				Level 1 variables	
	1	2	3	4	5	6	7	8	9	10	11	1 × 6	2 × 6	3 × 6	4 × 6	5 × 6	Girl	Books
<i>Western European countries</i>																		
Germany	0.05	-0.01	-0.19	0.34	0.20	0.13	-0.13	0.27	0.11	-0.12	0.03	0.00	0.00	-0.01	0.01	0.01	-0.09	0.35
Italy	0.19	-0.06	-0.00	0.30	0.00	0.05	0.12	0.04	0.14	0.07	-0.07	-0.01	0.00	0.00	-0.02	0.00	-0.09	0.17
Netherlands	-0.03	0.18	0.12	0.13	-0.03	0.03	0.05	-0.01	-0.21	-0.03	0.02	0.00	0.00	0.00	0.00	0.00	-0.07	0.25
Portugal	-0.09	0.13	-0.01	0.26	0.31	0.09	-0.10	-0.09	-0.21	0.27	0.05	0.00	0.01	0.00	0.01	0.02	-0.08	0.25
Spain	-0.07	-0.06	-0.05	0.54	0.31	0.32	0.09	-0.12	0.31	0.05	-0.01	0.00	0.00	0.00	-0.01	0.00	-0.10	0.25
<i>English-speaking countries</i>																		
England	0.09	-0.04	-0.16	-0.11	0.12	0.31	0.12	-0.12	-0.26	0.14	-0.13	-0.01	0.01	0.02	0.01	-0.02	-0.05	0.35
Ireland	-0.06	0.25	-0.11	0.10	0.10	0.11	-0.02	0.05	-0.05	0.19	-0.27	0.02	-0.07	0.03	-0.03	-0.03	-0.10	0.36
Northern Ireland	-0.19	0.13	0.06	0.20	0.29	0.11	0.00	-0.04	0.18	-0.05	0.13	-0.02	0.02	0.01	0.03	0.04	-0.03	0.39
New Zealand	-0.11	-0.09	0.12	0.06	0.37	-0.01	0.07	-0.08	0.17	0.14	0.02	0.00	0.00	0.00	0.00	0.01	-0.02	0.27
USA	0.05	0.01	-0.09	0.19	0.16	0.03	0.01	0.05	-0.02	0.07	-0.12	-0.01	0.00	0.01	-0.02	-0.02	-0.08	0.26

Note Standardized regression coefficients are reported at the between-level with significant relations ($p < .10$) indicated in bold. Instructional quality predicted by 1 = number of years of teaching experience, 2 = level of formal education (degree), 3 = major focus of teacher education, 4 = professional development, and 5 = preparedness; mathematics achievement of students at grade four aggregated at the classroom level predicted by 6 = number of years of teaching experience, 7 = ISCED level of formal education (teacher education degree), 8 = major, 9 = professional development, and 10 = preparedness; 11 = relation of instructional quality to mathematics achievement of students at grade four aggregated on the classroom level. At level 1, relations of gender and books at home to mathematics achievement of students at grade four were controlled for; here significant relations ($p < .05$) are indicated in bold

The importance of controlling for students' socioeconomic background was demonstrated by the strong relationship between the number of books at home and student achievement. In 39 out of the 47 countries, students who reported more books also had a higher mathematics score. This applied to all European, English-speaking and South-East Asian countries. In contrast, socioeconomic background was not significant in the African countries. Gender differences were evident in 28 countries, particularly in European (17 out of 18) and Latin America (2 out of 2) countries, and these differences unanimously favored boys. In contrast, Western Asian/Arabian (2 out of 12) and African (1 out of 3) countries were much less affected by gender inequalities, and when these were present in these countries, the differences favored girls.

2.5 Discussion

TIMSS data provide a unique opportunity to link student outcomes with teacher and instructional characteristics because they collect data from intact classrooms. The good fit of our model to the data within countries and across countries can be regarded as evidence that the model was well specified and that important teacher predictors of student achievement were selected. However, it seems to be important to distinguish between predictors that can be characterized as being more proximal or distal, respectively, to instructional quality or student achievement. Initial teacher education may have happened decades ago in case of experienced teachers, and programs may have been very different at that time compared to current teacher education programs (Wang et al. 2003). Teachers' initial education is in this manner an example of a teacher characteristic which, at least for a large group of teachers, is distal to the other variables included in the model, and moreover, likely confounded with other omitted variables. Taken together this makes it difficult to identify a systematic relationship between features of mathematics teacher education and instructional quality or student achievement.

Professional development activities taken during the past 2 years and teachers' self-efficacy are, in contrast, much more closely related to what happens currently in classrooms. The analysis presented demonstrates that teachers' participation in PD activities and their self-efficacy are both significantly associated with grade four students' mathematics achievement, both in the pooled international model and within a high number of countries. This finding therefore extends research-based knowledge by providing evidence for the generalizability of the influences of self-efficacy (Bandura 1986) and PD (Timperley et al. 2007) across widely different educational contexts.

However, for all other variables in the model, a large variation between the countries was observed and universal relationships with instructional quality and students' achievement were generally not observed. Teachers teach in a context of structures, policies and expectations. Scheerens (2007) separated these conditions into entities that were more or less "given" antecedents (such as population

characteristics or general valuation of education and teachers) and conditions that were more malleable by policy (such as level and type of decentralization or accountability arrangements). These differences in conditions may affect both the between-country and the within-country variability in teacher quality and instructional quality, and also the relationships between these concepts and students' learning outcomes. The TEDS-M study showed that in some countries teacher education is nationally standardized, while in other systems teacher education can be highly decentralized (Ingvarson et al. 2013). Furthermore, in some countries, teachers are trusted by both the public and their employers, who grant them more or less full autonomy in how they implement the curriculum and the instruction. In other countries, teachers will be firmly placed in a hierarchical system, with less freedom to influence the curriculum and instruction, in the extreme case with prescribed and detailed lesson plans.

Correspondingly, for all variables of teacher quality included in this chapter, we observed a noticeably large variation across countries. One potential consequence of such variation is that, in systems where teachers are fully autonomous individuals with responsibility for developing and implementing instruction, a relatively large within-country variation in instructional quality is possible, while systems characterized by teachers being provided with more or less prescribed lesson plans would likely have fewer degrees of freedom for some of the components typically included in instructional quality. In our models, the observed differences in direct relations of several variables describing teacher quality to instructional quality may be a reflection of this wider "ecology" of teaching. Taken together, this variation illustrates how international studies may use systematic differences in conditions and policies for teaching in order to at least provide examples of how alternative policies work in other settings, although, of course, such interpretations should be done with care since the wider cultural context of education represents a range of potentially very influential omitted variables.

In relation to this, it is also worth discussing how the educational system caters for specialized or generalized teachers of mathematics at grade four. It is reasonable to assume that in more or less all countries teachers in secondary schools will have a specialization in one or a few subjects. However, in primary schools, at least in the first years, there will be a larger between-country variation in the degree to which teachers have a general versus a specialized teacher education. Teachers with general qualifications will by default have a broader background with less in depth subject knowledge. This is a variation at the system level, which to a large degree was observed for the two proxy measures of teachers' educational background.

2.6 Limitations of the Study

One limitation of our study was its reliance on cross-sectional data. In order to study the effect of teacher and instructional quality on student achievement, and not the least, in order to study the possible mediation of teacher qualities by

instructional quality, use of data from experimental or longitudinal designs would be preferred. Follow-up studies with improved designs are urgently needed. Since the international studies are repeated at regular intervals, it should be possible to have repeated measures at country level in later surveys.

However, this would imply measures remain unaltered, which we would not recommend given another limitation of our study; the unsatisfactory quality of some of the measures used. This is primarily an issue regarding the measure of instructional quality used in this analysis. This measure was based on items in the general part of the teacher questionnaire. Consequently, the questions did not include explicit references to the subject of mathematics. In several countries, a teacher of grade four mathematics will also be teaching the same class other subjects. It may be that some of the teachers responded to this list of questions without having mathematics instruction in mind, which may cause validity problems (Schlesinger and Jentsch 2016).

There were other related measures which could have been used, and which are used in the analyses in other chapters in this book. A set of questions in the mathematics specific part of the teacher questionnaire also asked teachers to report their instructional activities in mathematics. However, these questions reflect surface characteristics of teaching practices, and did not correspond to the theoretical framework of instructional quality applied in this book, which is based on current research on instructional quality. A measure based on students' responses could also have been used. However, given the low age of the students in grade four, we opted to rely on the teachers' reports. Improvements in the instructional quality measures to better include recent research in this area (in particular the work done by the Klieme group; see for example Decristan et al. 2015; Rakoczy et al. 2010) seem to be urgently needed.

A third feature of our analysis that may be regarded as a minor limitation is, given the limited sample size of teachers and classrooms in many countries, item parcels were applied instead of single items. This leads to some loss of information. Given that the reliability of most parcels was reasonably high, the grouping of items into parcels can be assumed to represent a minor reduction of information with only small consequences for the analysis.⁵ However, given that there are potentially differential relationships between the three indicators and student achievement across countries and within countries, the research questions of this paper may also merit reinvestigation at the item- or indicator level.

There were other dimensions in the TIMSS questionnaire gauging teacher characteristics that were found to be of relevance for students' achievement. These measures were omitted from this analysis for several reasons. Firstly, for some of them it was not possible to confirm metric measurement invariance (this applied, for

⁵This argument is not directly applicable to the parcels representing the three theoretically-based aspects of InQua, since they consist of two variables only. The total internal consistency for the manifest variable using all six variables as compared to the variable using the three item parcels is only a fraction higher, 0.65 as compared to 0.61, which demonstrates that the parcels function almost equivalent to the single items.

example, to teacher motivation) and, secondly, their inclusion would have introduced a risk of multicollinearity. In addition, as a two-level multi-group analysis framework was applied, keeping the model simple was a necessary priority. It should be noted that the final choice of indicators of teacher qualities in our model did not fully match the dimensions cited most often in contemporary teacher effectiveness studies. For example, TIMSS did not include measures of the teachers' actual knowledge and skills to teach mathematics (see for example, Blömeke et al. 2012; Tatto et al. 2012).

2.7 Conclusions and Recommendations

The results of the present study clearly support the relevance of teacher quality for instructional quality and for educational outcomes. Instructional quality and mathematics achievement were significantly related to several teacher characteristics selected on the basis of contemporary research and, their availability within the TIMSS 2011 data. Patterns emerged across countries and cultures, both with respect to the absolute level of some constructs and the relations between teacher quality, instructional quality and outcomes. Some characteristics were more regionally relevant. However, although the model fits the data from the majority of countries, the structural relations represented by this model do not provide a universal model.

The lack of a universally applicable model is obvious: significant research is needed to clarify the generalizability of these results. One particular topic for research concerns the relevance of initial teacher education, which several times was found to be non-significant, replicating previous findings from other cross-sectional surveys (see for instance, Nordenbo et al. 2008). This could be related to the fact that teacher education has changed profoundly in many countries over the last decades (Wang et al. 2003; Darling-Hammond and Lieberman 2012). It is reasonable to assume that characteristics of students recruited into the profession have changed over time. Access to teacher education may historically have been more selective and restricted to students with relatively higher marks from secondary education. Also, the demand and provision of deep mathematical knowledge in the teacher education may have changed as teacher education has been reformed at specific points in time. Teacher experience and formal qualifications as measured in TIMSS are therefore likely confounded with other characteristics not included in our model. Distinguishing between age cohorts would provide important information, but this was not feasible with the current data set given the already rather small sample size. One solution for future surveys could be to include larger samples of teachers and classrooms in countries where changes in some of these confounding characteristics can be described and included in the model from other sources.

We have chosen to focus on cognitive outcomes in this chapter, given that other chapters in this book cover student motivation or bullying as outcomes. It is important to recall that outcomes of education are multi-dimensional and that

cognitive and motivational variables are both important. Evidence suggests that motives are often positively related to cognitive learning outcomes and that motivation supports cognitive learning long term (Benware and Deci 1984; Grolnick and Ryan 1987). Reducing schooling to cognitive outcomes would therefore be a shortcoming. In further studies of how teacher quality and instructional quality relates to outcomes, it would therefore be relevant to include also students' motivation and interest as dependent variables in one and the same model.

Another major recommendation for future studies based on our experience with analyzing the complex relationship between teacher quality, instructional quality and student outcomes, is that future surveys need invest in the development of improved measures of instructional quality. A long-standing controversy exists whether teacher or student ratings describe instructional quality more reliably and/or more validly (Desimone 2011; Schlesinger and Jentsch 2016; Wagner et al. 2015). Current research understanding suggests that the correlation between these two approaches is only moderate and that their relation with student achievement differs. This may reflect not only that students and teachers perceptions differ, but also that the measures represent slightly different aspects of the instructional activities taking place in the classroom. In general, we would therefore recommend that measures of instructional quality, in line with the current practice in the IEA studies, include both types of sources to develop measures of the quality of the instructional activities.

However, the current measures in both the teacher and the student questionnaires fail to fully represent the depth and breadth of the concept of instructional quality. The three core aspects in the measure of InQua that we applied (clarity of instruction, cognitive activation, and supportive climate) are represented by two items only. Each of these aspects represents separate and relatively broad and many-faceted constructs by themselves, which should be reflected in future studies. Furthermore, classroom management is a vital dimension of instructional quality not included in the generic teacher questionnaire. And not the least, as discussed already, the construct used in this chapter is based on generic questions, while it would provide more fidelity to the analysis if a measure specific to the quality of the mathematics lessons had been applied. In future surveys, priority should rather be given to the improvement of context sensitive measures of instructional quality. Frequency of different specific activities may not represent an ideal way to assess the quality with which these activities are carried through. Some actions probably occur relatively often in high quality teaching (for instance, summarizing at the end of the lecture), while others would probably need to be used less often in order to represent an optimal quality (for instance, working on problems with no obvious solution). In summary, new improved measures of InQua should:

- (1) reflect both students' and teachers' experiences,
- (2) have a broader scope, including the four core components, clarity of instruction, cognitive activation, classroom management, and supportive climate,

- (3) cover each of these aspects in depth by including separate, but related, constructs,
- (4) be subject-specific rather than generic, and
- (5) include scales aimed at capturing qualities of various activities.

Open Access This chapter is distributed 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, duplication, 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 work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

- Alexander, R. (2001). *Culture and pedagogy: International comparisons in primary education*. West Sussex: Wiley-Blackwell.
- Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: New developments and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*, 133–180.
- Benware, C., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, *21*, 755–765.
- Bishop, A. (2004). Mathematics education in its cultural context. In T. P. Carpenter, J. A. Dossey, & J. L. Koehler (Eds.), *Classics in mathematics education research*. National Council of Teachers of Mathematics: Reston, VA.
- Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM – The International Journal on Mathematics Education*, *44*, 223–247.
- Blömeke, S., & Kaiser, G. (2012). Homogeneity or heterogeneity? Profiles of opportunities to learn in primary teacher education and their relationship to cultural context and outcomes. *ZDM*, *44*, 249–264.
- Blömeke, S., Suhl, U., Kaiser, G., & Döhrmann, M. (2012). Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teaching and Teacher Education*, *28*, 44–55.
- Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*, 416–440.
- Boyle, B., Lamprianou, I., & Boyle, T. (2005). A longitudinal study of teacher change: What makes professional development effective? Report of the second year of the study. *School Effectiveness and School Improvements*, *16*, 1–27.

- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*(5), 1005–1018.
- Clotfelter, Ch., Ladd, H., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26*, 673–82.
- Cochran-Smith, M., & Zeichner, K. M. (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, NJ: Erlbaum.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Darling-Hammond, D. & Lieberman, A. (Eds.). (2012). *Teacher education around the world: changing policies and practices*. Teacher Quality and School Development Series. Abingdon, Oxon: Routledge.
- Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., & Fauth, B., et al. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding? *American Educational Research Journal, 52*, 1133–1159.
- Desimone, L. M. (2011). A primer on effective professional development. *Phi Delta Kappan, 92*, 68–71.
- Foy, P., Arora, A., & Stanco, G. M. (Eds.). (2013). *TIMSS 2011 user guide for the international database. Supplement 1: International version of the TIMSS 2011 background and curriculum questionnaires*. Chestnut Hill, MA/Amsterdam: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College & International Association for the Evaluation of Educational Achievement (IEA).
- Goldsmith, L., Doerr, H., & Lewis, C. (2014). Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education, 17*, 5–36.
- Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology, 52*, 890–898.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1–55.
- Ingvarson, L., Schwillie, J., Tatto, M. T., Rowley, G., Peck, R., & Senk, S. L. (2013). *An analysis of teacher education context, structure, and quality-assurance arrangements in TEDS-M countries: Findings from the IEA teacher education and development study in mathematics (TEDS-M)*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Joncas, M., & Foy, P. (2012). Sample design in TIMSS and PIRLS. In I. V. S. Mullis & M. O. Martin (Eds.), *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Kee, A. N. (2012). Feelings of preparedness among alternatively certified teachers: What is the role of program features? *Journal of Teacher Education, 63*, 23–38.
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*, 568–589.
- Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*, 12–23.
- Leung, F. K. S. (2006). Mathematics education in East Asia and the West: Does culture matter? In F. Leung, K.-D. Graf, & F. Lopez-Real (Eds.), *Mathematics education in different cultural traditions: A comparative study of East Asia and the West, 13th ICMI study*. New York: Springer.

- Little, T., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel? Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173.
- Martin, M. O., & Mullis, I. V. S. (Eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center Lynch School of Education, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschof, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center Lynch School of Education, Boston College.
- Muthén, L. K., & Muthén, B. O. (2008–2012). *Mplus user's guide* (7th Edn.). Los Angeles, CA: Muthén & Muthén.
- Nordenbo, S. E., Sjøgaard Larsen, M., Tiftikiçi, N., Wendt, R. E., & Østergaard, S. (2008). *Teacher competences and pupil learning in pre-school and school: A systematic review carried out for the ministry of education and research, Oslo*. Copenhagen: Danish Clearinghouse for Educational Research, School of Education, University of Aarhus.
- Pajares, F. (1996). Self-efficacy beliefs in achievement settings. *Review of Educational Research*, 66, 543–578.
- Rakoczy, K., Klieme, E., Lipowsky, F., & Drollinger-Vetter, B. (2010). Strukturierung, kognitive Aktivität und Leistungsentwicklung im Mathematikunterricht. *Unterrichtswissenschaft: Zeitschrift für Lernforschung*, 38, 229–246.
- Sammons, P. (2009). The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement*, 20, 123–129.
- Scheerens, J. (2007). *Conceptual framework for the PISA 2009 background questionnaires*. Twente, The Netherlands: University of Twente.
- Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School effectiveness and school improvement*, 24, 1–38.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM*. doi:10.1007/s11858-016-0765-0.
- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics*. Dordrecht: Kluwer.
- Schwille, J., Ingvarson, L., & Holdgreve-Resendez, R. (Eds.). (2013). *TEDS-M encyclopedia: A guide to teacher education context, structure, and quality assurance in 17 countries. Findings from the IEA teacher education and development study in mathematics (TEDS-M)*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77, 454–499.
- Snijders, T., & Bosker, R. (2012). *Multilevel analysis: An introduction to basic and applied multilevel analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Stark, J. S., & Lattuca, L. R. (1997). *Shaping the college curriculum: Academic plans in action*. Boston, CT: Allyn and Bacon.
- Tatto, M. T., Schwille, J., Senk, Sh, Rodriguez, M., Bankov, K., & Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics: First findings*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Wellington: Ministry of Education.

- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202–248.
- Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2015). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*. Retrieved from <http://psycnet.apa.org/psycinfo/2015-43248-001/>.
- Wang, A. H., Coleman, A. B., Coley, R. J., & Phelps, R. P. (2003). *Preparing teachers around the world (Policy Information Report)*. Princeton, NJ: ETS.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89–122.
- Wilson, S., Floden, R. & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations: Research report prepared for the US Department of Education*. Washington, DC: University of Washington Center for the Study of Teaching and Policy, Seattle.