Chapter 1
Conceptual Framework and Methodology of This Report

Trude Nilsen, Jan-Eric Gustafsson and Sigrid Blömeke

Abstract In this volume, five separate studies examine differing aspects of relations between teacher quality, instructional quality and learning outcomes across countries, taking into account context characteristics such as school climate. The 2007 and 2011 TIMSS (Trends in Mathematics and Science Study) cycles provided the research data. These five studies cover grade four or grade eight students and their teachers, including cognitive or affective-motivational learning outcomes. This introductory chapter describes the overall conceptual framework and the research questions posed by each chapter, and outlines the general design features of TIMSS. Key constructs, and common methodological issues among the five studies are discussed, and this introduction concludes with an overview of all chapters.

Keywords Instructional quality · Teacher quality · Student outcome · Theoretical framework · Trends in Mathematics and Science Study (TIMSS)

1.1 Introduction

Researchers and practitioners have long known that the quality of teachers and the quality of their instruction are key determinants of student learning outcomes (Klieme et al. 2009; Seidel and Shavelson 2007). However, the relationships have

T. Nilsen (✉)
Department of Teacher Education and School Research, University of Oslo, Oslo, Norway
e-mail: trude.nilsen@ils.uio.no

J.-E. Gustafsson
Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden
e-mail: jan-eric.gustafsson@ped.gu.se

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

often been difficult to quantify and understand empirically. Reviews of previous research have pointed to challenges in measuring teacher and instructional quality (Schlesinger and Jentsch 2016; Kunter et al. 2013). Moreover, the impact of student background often swamps the effects of the other variables, rendering them less visible. Finally, due to teacher selection and rules of certification, these variables often vary only little within a school system, making it difficult to identify effects.

Advancements in psychometrics and quantitative methods, along with the establishment of international large-scale assessments (ILSA), offer researchers new opportunities to study relations between teachers, their instruction and learning outcomes (Chapman et al. 2012). For instance, ILSA data provide the opportunity for multi-level analysis, standardized definitions of variables, trend design and representative samples from a large number of educational systems, in the following also called countries. Perhaps the best known ILSAs are the International Association for the Evaluation of Educational Achievement (IEA) Trends in Mathematics and Science Study (TIMSS), and the Organisation for Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA) and Teaching and Learning International Survey (TALIS). Out of these, TIMSS is the only one that provides data on the student, class and school levels. TIMSS therefore provides data well suited for an examination of relations between teacher quality, instructional quality and student outcomes across cohorts, time, and countries from all continents.

Using the world as a global educational laboratory may contribute toward an international understanding of teacher quality and instructional quality, and establish their importance for student learning outcomes across and within countries and over time. This demands research that takes into account: (1) the complexity of educational systems with many hierarchical layers and interwoven relationships (Scheerens and Bosker 1997); (2) the complexity of relationships within each layer with direct and indirect effects; (3) the variation of these relationships across countries; and (4) their development over time. Since it is difficult to take all these complexities into account within one study, combining results from different studies investigating subsets of relations may currently be the best way to make progress.

This book presents five studies which have been undertaken in this spirit. The studies complement each other to address the complexities mentioned above. The studies examined the following research questions:

(1) *Which relations exist between teacher quality, instructional quality and mathematics achievement in grade four across and within countries, and is it possible to identify larger world regions or clusters of countries where similar relational patterns exist?* (Chap. 2)
(2) *Which relations exist between school climate, instructional quality, and achievement motivation in mathematics in grade eight across and within*

countries, and is it possible to identify larger world regions or clusters of countries where similar relational patterns exist? (Chap. 3)

(3) *To what extent can a causal influence of school climate and teacher quality on mathematics achievement in grade eight be identified in country-level longitudinal analyses?* (Chap. 4)

(4) *Which relations exist between instructional quality and reading, and between instructional quality and mathematics achievement in grade four, and to what extent does instructional quality moderate the relations between reading and mathematics achievement?* (Chap. 5)

(5) *Which relations exist between bullying and instructional quality in grade four across countries and within countries?* (Chap. 6)

The last chapter of this book summarizes the results obtained in these five studies and discusses conceptual and methodological challenges, as well as possible improvements in both research and practice. In taking this approach, our aim is to contribute to educational effectiveness research, to educational policy and practice, and to the field of educational measurement.

1.2 Conceptual Framework

Our research is situated within the field of educational effectiveness research, and this field has made great progress over the last three decades. This is partly because certain limitations of previous studies have been amended (Creemers and Kyriakides 2008; Chapman et al. 2012). These limitations included models which could only partially account for the nested nature of data, non-random samples, cross-sectional designs, or non-robust software. However, while there were methodological advances within the field of educational effectiveness, Creemers and Kyriakides (2006, p. 348) argued that there was also a need for "rational models from which researchers can build theory." Over the years, they developed and tested a model for educational effectiveness, which they called the dynamic model of educational effectiveness. This model takes into account the complexity of educational systems, where students are nested within classes that are nested within schools, where variables within and across these levels can be directly and indirectly related, and where changes occur. This model also accounts for a national context level, which refers to the educational system at large, including the educational policy at the regional and/or national level, which should be examined in comparative studies (Kyriakides 2006). The model is well recognized internationally (Sammons 2009).

In this book, a conceptual framework (Fig. 1.1) is used that starts with the dynamic model of educational effectiveness (Creemers and Kyriakides 2008) and operationalizes it with respect to the research questions of this report. In line with
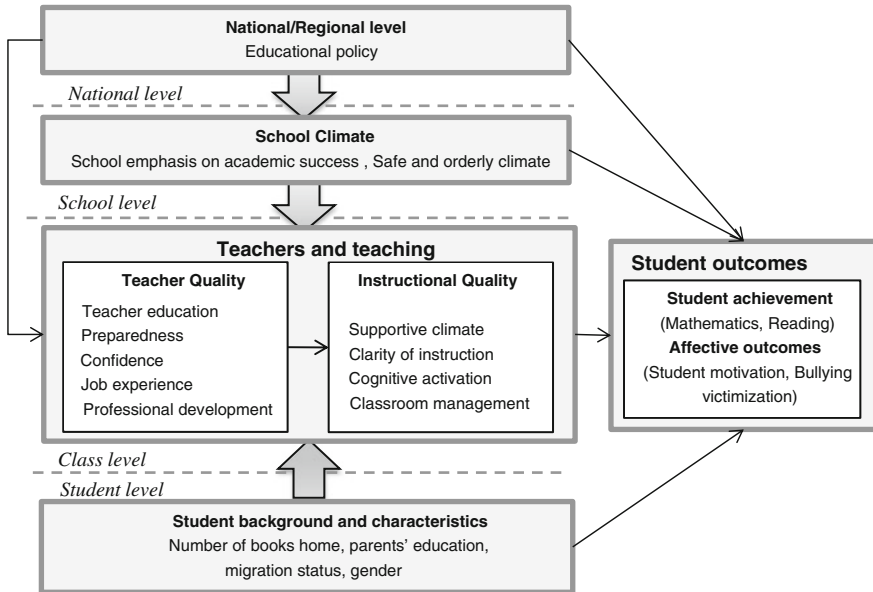
Fig. 1.1 Conceptual framework of determinants of student outcomes examined in this book

Kyriakides et al. (2009) and other studies (for example Baumert et al. 2010; Kane and Cantrell 2010), teacher and teaching variables at the class level are hypothesized to be most important for student learning. The conceptual framework focuses on relations between the national, school, class, and student level. The model shows how the national level is hypothesized to influence the school and teacher levels, as well as student outcomes in the five studies of this report. These relations may be both direct and indirect. Because of differences between educational systems, including different cultural contexts, educational values, educational policies, and structural features of the school system, we hypothesize that the relations of the indicators examined at lower levels, such as schools, classes and students, vary substantially within countries. Based on existing research, we also hypothesize that patterns exist that reflect similarities between *groups* of countries, due to similarities in culture, values, policies or school structure (see for example Blömeke et al. 2013).

School level variables are hypothesized to influence the class and student level (Fig. 1.1). In this book, we examine the school features School emphasis on academic success and Safe and orderly climate. The class level contains two important variables for learning outcomes, namely teacher quality and instructional quality. These constructs are also hypothesized to be interrelated (Fig. 1.1). Finally, in line with existing research (Gustafsson et al. 2013; Hansen and Munk 2012) student characteristics (such as gender and minority status) and home background (for example, parents' education) are hypothesized to be related to student outcomes. Such outcomes may be cognitive or affective.

1.3 Operationalization of School-, Class- and Student-Level Features

This section presents a brief outline of how crucial constructs were operationalized. A detailed presentation is provided in the following chapters.

1.3.1 Teacher Quality

Goe (2007) presented a framework for understanding the key components of teacher quality and their relations to student learning outcomes. According to this framework, teacher quality includes both teacher qualifications and characteristics (*inputs*) that influence teachers' instruction (*process*) and student *outcomes* (e.g., achievement and motivation). In this book, teacher quality is operationalized via qualifications such as teacher education level, job experience and participation in professional development activities, as well as by teacher characteristics such as self-efficacy. The Teacher Education and Development Study in Mathematics (TEDS-M) was the first international large-scale assessment that examined these features, with representative samples from a broad range of countries (see for example Blömeke et al. 2011; Tatto et al. 2012). In mathematics, teacher quality has been shown to be of importance for student achievement in a number of within-country studies (Baumert et al. 2010; Blömeke and Delaney 2014). A substantial research gap exists with respect to non-Western countries and comparative research across countries applying the same kind of instruments. This book intends to narrow this research gap.

1.3.2 Instructional Quality

Instructional quality is a construct that reflects those features of teachers' instructional practices well known to be positively related to student outcomes, both cognitive and affective ones (Decristan et al. 2015; Fauth et al. 2014; Good et al. 2009; Hattie 2009; Klusmann et al. 2008; Seidel and Shavelson 2007). The construct is understood and operationalized differently across the field but its multi-dimensionality was revealed in major research projects originating in both Europe (Baumert et al. 2010; Kunter et al. 2008) and the United States (Ferguson 2010; Kane and Cantrell 2012). As with teacher quality, a research gap exists with respect to non-Western countries and calls for comparative research across countries.

The operationalization of instructional quality used in this book is mainly based on the model of three "global dimensions of classroom process quality" (Klieme et al. 2001; Klieme and Rakoczy 2003; Lipowsky et al. 2009). Klieme and colleagues' model was developed based on data from the German extension to TIMSS

Video and subsequently applied to data from PISA 2000; its dimensions include cognitive activation, supportive climate, and classroom management. This model is similar to studies carried out independently in the USA (Kane and Cantrell 2012; Pianta and Hamre 2009; Reyes et al. 2012).

Cognitive activation refers to teachers' ability to challenge students cognitively, and comprises instructional activities in which students have to evaluate, integrate, and apply knowledge in the context of problem solving (Baumert et al. 2010; Fauth et al. 2014; Klieme et al. 2009). Supportive climate is a dimension that refers to classrooms where teachers provide extra help when needed, listen to and respect students' ideas and questions, and care about and encourage the students (Kane and Cantrell 2012; Klieme et al. 2009). Supportive climate may include clear and comprehensive instruction, clear learning goals, connecting new and old topics, and summarizing at the end of the lesson, but some research shows that supportive climate should be discriminated from clarity of instruction (Kane and Cantrell 2010). We therefore consider clarity of instruction as a fourth dimension of instructional quality.

1.3.3   School Climate

While teacher quality and instructional quality may directly influence students' learning and motivation, school climate creates the foundation for instruction and may hence influence learning both directly and indirectly (Kyriakides et al. 2010; Thapa et al. 2013; Wang and Degol 2015; see Fig. 1.1). In a recent review of school climate across several fields, Wang and Degol (2015) observed that school climate is defined differently across studies, but that certain aspects may be key. There seems to be broad consensus that academic climate and a safe and orderly climate are such key aspects and that they are positively related to learning outcomes (Bryk and Schneider 2002; Hoy et al. 2006; Thapa et al. 2013).

Academic climate focuses on the overall quality of the academic atmosphere; the priority and ambition for learning and success (Hoy et al. 2006; Martin et al. 2013; Nilsen and Gustafsson 2014; Wang and Degol 2015). School emphasis on academic success (SEAS) is therefore examined as an indicator of academic climate in this book. SEAS reflects a school's ambition and priority for learning and success. It has been shown to be related to students' learning in a number of countries (Martin et al. 2013; Nilsen and Gustafsson 2014). A second variable examined in this book is a safe and orderly climate, which refers to the degree of physical and emotional security provided by the school, as well as to an orderly climate with disciplinary practices (Goldstein et al. 2008; Gregory et al. 2012; Wang and Degol 2015). Studies have revealed that this variable is also related to student learning outcomes.

1.3.4 Student Outcomes

Throughout this book, different types of student outcomes are taken into account to address the multidimensionality of educational objectives of schooling. The main emphasis is on student achievement in mathematics at grade four and eight, but reading achievement using the IEA's Progress in Reading and Literacy Study (PIRLS) data, as well as student motivation and bullying victimization are also examined.

Cognitive outcomes in mathematics and reading

In grade four, students are assessed in TIMSS in the domains Number, Geometric Shapes and Measures, and Data Display, and in grade eight in Number, Algebra, Geometry, and Data and Chance. In addition to covering these content domains, the items also cover the cognitive demands Knowing, Applying and Reasoning (Mullis et al. 2012a). According to Niss (2003), mathematical competence "means the ability to understand, judge, do, and use mathematics in a variety of intra- and extra-mathematical contexts and situations in which mathematics plays or could play a role" (p. 6). In other words, students do not just need knowledge in mathematics, but must also be able to apply knowledge and conceptual understanding in different contexts, and to analyze, and reason to solve problems. The TIMSS framework reflects this notion fairly well (Mullis et al. 2012b) and is also in line with a number of other frameworks in mathematics (e.g. Kilpatrick 2014; Schoenfeld and Kilpatrick 2008).

TIMSS does not capture every aspect of mathematical competence. According to Niss (2003), mathematical competence includes eight different competencies that, for instance, involve mathematical theory like using and understanding theorems, communication in mathematics, handling symbols, including manipulating equations, and making use of aids and tools (including information technology). Although there are some items that reflect such aspects, extra-mathematical contexts and students' communication in mathematics are not measured extensively in TIMSS. In contrast, TIMSS does measure to some extent mathematical theory like using and understanding theorems, and students' ability to handle symbols, including manipulating equations (Hole et al. 2015). Moreover, TIMSS is based on the cores of the curricula of all countries participating, and it includes crucial cognitive demands such as knowing, applying and reasoning. Thus, TIMSS measures the key competencies in mathematics described by Niss (2003) to a satisfying degree.

In Chap. 5 of this book, reading achievement is included as well as mathematics achievement because reading literacy is regarded to be the foundation of most learning processes and an important ability students need to acquire during schooling. The data come from TIMSS and PIRLS 2011, where reading is defined as "the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers in

school and everyday life, and for enjoyment" (Mullis et al. 2009). This definition has changed over study cycles, but is a good reflection of recent theories of reading literacy (Alexander and Jetton 2000; Ruddell and Unrau 2004; for more details, see Chap. 5).

1.3.5 Student Affective Outcomes

In addition to achievement, a number of studies also include interest, motivation, and self-beliefs as student outcomes (Bandura 1997; Eccles and Wigfield 2002). These constructs reflect students' motivational states (see Chap. 3 for more theory on this). A substantial research gap exists with respect to studies in which school-, teacher- and class-level features are related to affective student outcomes in Western and non-Western countries, as well as with respect to comparative research across countries applying the same set of instruments. This book intends to reduce this research gap.

Given that learning takes place in social settings (i.e., in classrooms and schools), social interaction with peers must also be taken into account in considering student outcomes and their determinants. One of the constructs reflecting the results of such interactions refers to bullying victimization, which is has been shown to be linked with achievement and motivation (Engel et al. 2009; Skues et al. 2005) and has been found to be related to classroom and school factors such as discipline, teacher support, instructional quality and school climate within several countries (Kyriakides et al. 2014; Murray-Harvey and Slee 2010; Richard et al. 2012). This aspect of research is progressed in this book using a comparative approach applied across a large range of countries.

1.4 TIMSS Design

TIMSS is an international large-scale survey of student achievement in mathematics and science. First conducted in 1995, TIMSS assesses students in grade four and eight every fourth year. Most chapters in this book draw on the 2011 TIMSS data, which included over 60 countries. All chapters considered as many countries as possible, but some countries had to be excluded depending on the chapter's research question; for example due to missing data on a crucial variable.

The TIMSS assessments include so-called trend items, meaning that the exact same items are reused in adjacent cycles (for example repeated for both 2007 and 2011; such data are used in Chap. 4 of this report). There are roughly equal numbers of multiple choice and constructed response (open) items. In order to cover the broad range of content and cognitive domains, approximately 200 items were included in the mathematics assessment. To ease the burden of responding to such a large number of items, TIMSS uses a so-called rotating matrix-sampling design

(for more on this, see Martin and Mullis 2012). Hence, students do not all answer the same set of questions/items.

Because each student only responds to a part of the item pool, the TIMSS scaling approach uses multiple imputation methodology to obtain proficiency scores for all students. This method generates multiple imputed scores or plausible values from the estimated ability distributions (Martin and Mullis 2012). In addition, a conditioning process, in which student responses to the items are combined with information about the student's background, is implemented to increase score reliability. Plausible values hence provide consistent estimates of population characteristics. In 1995, the mean mathematics achievement was set to a score of 500, with a standard deviation of 100. After this, all cycles have been calibrated to the same scale as that of 1995 by means of concurrent calibration, using the trend items and data from countries that participated in adjacent cycles (Martin and Mullis 2012).

In addition to assessment in mathematics, students, parents, teachers and school leaders respond to questionnaires with questions pertaining to background and context (Foy et al. 2013).

TIMSS employs a two-stage random sample design, where schools are drawn as a first stage, and then intact classes of students are selected from each of the sampled schools as a second stage. Hence, students are nested within classes, and classes are nested within schools. Students are representative samples of the entire population of students within a country. Teachers are connected to the sample of classes within each country, which does not necessarily mean that TIMSS includes representative samples of teachers. Hence, results concerning teacher variables, such as teachers with high levels of education, reflect representative samples of students whose teachers have high levels of education. Some classes had more than one mathematics teacher. The percentage of students with more than one mathematics teacher was 1.4 % in grade four, and 1.7 % in grade eight. For students with more than one mathematics teacher, data from only one of them was included at random. The amount of data deleted by this procedure was negligibly small.

1.5 Measuring Key Constructs

The rich data from the large number of participating students, teachers, classrooms, schools and educational systems offer great opportunities to explore and compare different solutions to these measurement challenges, and to investigate characteristics of different measurement models. But as issues of validity and reliability of measurement are present in virtually all empirical research, they also provide challenges in secondary analyses of large-scale data such as TIMSS. Typically, few items are available to measure each of the many complex constructs that are central to educational research. Furthermore, since these items need to reflect conceptualizations of constructs in many different cultural and educational contexts, they may not be perfectly relevant as indicators of the theoretical constructs that a particular researcher wants to investigate.

The researchers involved in the different chapters designed measurement approaches to suit their research problems within the common framework and with the data available from TIMSS (see http://timssandpirls.bc.edu/timss2011/international-database.html). Below we present the measurement solutions adopted for the constructs used in more than one chapter.

1.5.1   Instructional Quality

Instructional quality is a key construct, central to most of the chapters of this volume. As is described above, there is converging evidence from within-country studies that four dimensions (clarity of instruction, cognitive activation, classroom management, and supportive climate) may be needed to adequately measure instructional quality. In TIMSS, both the student and the teacher questionnaires include items covering some of these aspects. However, some construct under-representation exists in both cases. Furthermore, concerns have been raised about the reliability and validity of both teacher and student assessments of instructional quality. Social desirability bias in teachers' assessments is often mentioned as a threat to validity, as is lack of competence and stability in younger students' assessments of instructional quality. So, both approaches may have benefits and limits. Recent research suggests in addition that while a single student's assessment is likely to be unreliable, the aggregated assessments of a classroom of students may be both reliable and valid (Marsh et al. 2012; Scherer and Gustafsson 2015). All chapters where students' ratings were used therefore identified the construct both at the student and the class level (Marsh et al. 2012; Wagner et al. 2015).

Four chapters investigated instructional quality. Blömeke, Olsen and Suhl (Chap. 2, grade four) used teacher data due to the young age of grade four students. They created three indicators of instructional quality (clarity of instruction, cognitive activation, and supportive climate) from six items included in the teacher questionnaire and used these item parcels as indicators of a latent variable representing instructional quality. They were thus able to deal with the inherent multidimensionality of the construct. Scherer and Nilsen (Chap. 3, grade eight) used four items from the student questionnaire aimed to assess clarity of instruction and supportive climate. They employed a two-level confirmatory factor analysis model with latent variables representing perceived instructional quality at the class- and student-levels. Nortvedt, Gustafsson and Lehre (Chap. 5, grade four) used a similar two-level approach to measure class-level instructional quality, but they took advantage of student assessments of both teaching of mathematics and of reading. Rutkowski and Rutkowski (Chap. 6, grade four) also used student assessments of instructional quality in mathematics with four items in the class- and student-level models to represent instructional quality.

Thus there is considerable overlap between the approaches used in the different chapters, but there also are differences both in the actual items included in the

models and in whether teacher or student responses are relied upon. In the last chapter, we discuss this further, and assess the results obtained from the different analyses.

1.5.2 Teacher Quality

As is described in greater detail in the theoretical section and in Chap. 2, teacher quality may analytically be differentiated into teacher qualifications, such as education, experience and professional development, and teacher characteristics, such as motivation and self-efficacy.

Formal qualifications are indicated by the number of years of education, the level of the teaching license, years of teaching experience, major academic discipline studied, and professional development. These features can be assessed with good reliability. However, formal qualifications are sometimes found to be weakly related to measures of instructional quality or student achievement across educational systems or content areas because a major qualification in mathematics in a program on ISCED level 5 may mean something different that in a program on ISCED level 6 or 7, because recruitment to the more advanced program is more selective. This problem has led to attempts to measure teacher efficiency with value-added techniques, an approach that is approximated in this book by combining the variables available from the TIMSS data set in one model. In other lines of research, teacher knowledge and skills, such as pedagogical content knowledge and content knowledge, are measured directly (see Baumert et al. 2010), but this is not possible to implement in large-scale international studies, unless this is the aim of the study, as was the case with the TEDS-M study (Blömeke et al. 2011, 2013).

Two chapters included teacher quality variables. Blömeke, Olsen and Suhl (Chap. 2, grade four) investigated number of years of experience, level of formal education completed, and major (in this book and the TIMSS framework defined as the main academic discipline studied) in either mathematics or mathematics education, professional development in mathematics instruction, with attention to both broad activities and specific challenges, as well as collaborative school-based professional development with peers. They also measured teacher self-efficacy with items asking about preparedness to teach numbers, geometry and data. Gustafsson and Nilsen (Chap. 4, grade eight) investigated number of years of experience, level of formal education completed, whether teachers had a major qualification in mathematics or not, professional development in five different areas, and teacher self-efficacy in teaching number, algebra, geometry and data and chance. Thus, similar variables were investigated, the differences being due to the fact that different grade levels were investigated.

1.5.3 School Climate

School climate is often regarded as a foundation for instructional quality. Scherer and Nilsen (Chap. 3, grade four) investigated empirically whether this is the case or not across a broad range of countries. Gustafsson and Nilsen (Chap. 4, grade eight) asked if there is a causal relation between school climate and achievement. As a well-established measure of academic climate, SEAS was used in both chapters. In addition, Scherer and Nilsen (Chap. 3) created a safety scale from three items and an order scale from two items of the TIMSS student survey.

1.5.4 Socioeconomic Status

In educational research, socioeconomic status (SES) is often used to control for selection bias, but may also be a variable which is of interest in its own right. In the IEA study frameworks, an item asking about number of books at home (Books) has a long tradition as an indicator of SES. In TIMSS 2011, further SES indicators were introduced: parents' highest level of education and level of home study supports, such as students having their own room or internet connection. The TIMSS Home Educational Resources (HER) index (Martin and Mullis 2012) was created from these indicators.

   SES was included as a control variable in the analyses presented in three chapters. Blömeke, Olsen and Suhl (Chap. 2, grade four), and Rutkowski and Rutkowski (Chap. 6, grade four) used Books as an indicator, while Scherer and Nilsen (Chap. 3, grade four) relied on the HER index. A case can be made for both choices. While the HER index has better measurement properties than Books, the latter indicator has remained unaltered for a long time and similar indicators of home background are used in the other international large-scale studies, allowing for easy comparisons with previous research.

1.6   Challenges in Analyzing the Data

In addition to measuring the intended constructs appropriately, data analysis also presented challenges. Those that were common across the chapters in this book are briefly discussed below.

1.6.1 Causality

Many of the research questions asked in this report concern issues of causality. Basically, two types of causal questions can be identified. The first type concerns

causal effects, or whether a certain factor (for example instructional quality) influences an outcome variable, such as mathematics achievement. If there is a causal relation, increasing instructional quality will cause mathematics achievement to improve. However, TIMSS data are cross-sectional by nature and can mostly only provide correlations between instructional quality and achievement. There is insufficient evidence to conclude that a causal relation exists because third-variable explanations or reversed causality cannot be excluded.

If, for example, students receiving better instructional quality also have higher SES, an alternative explanation could be that the correlation arises because SES is related both to achievement and to instructional quality. If information about SES is available, this hypothesis can be tested by statistically controlling for the effect of SES on the relation between instructional quality and achievement. However, given that there are many unobserved variables that potentially may account for an observed correlation between instructional quality and achievement, it is unlikely that data on all of them exists. Cross-sectional studies therefore cannot rule out the possibility that omitted variables are causing an observed correlation. A way to strengthen causal inference is to use a longitudinal approach (Gustafsson 2013). Gustafsson and Nilsen (Chap. 4) present the idea behind such an approach and apply it to analyses of effects of teacher quality and school climate on mathematics achievement using data from TIMSS 2007 and 2011.

The other type of causal question concerns causal mechanisms, or how sequences of variables influence one another. Reversed causality is a well-known problem in educational research using cross-sectional data in this context. An example would be that the relation between teacher quality and student achievement is negative although longitudinal studies show the opposite. An explanation could be that a country may have taken specific actions to compensate for weak student achievement, perhaps by placing the best teachers in the weakest classes. The correlation between teacher quality and student achievement based on cross-sectional data would then be negative, although, in this case, longitudinal data would reveal that classes with better teachers develop better than other classes provided the starting achievement level is taken into consideration.

Illustrating how sequences of variables may influence one another, is Blömeke, Olsen and Suhl's (Chap. 2) study, which tested the hypothesis that teacher quality influences instructional quality, which in turn influences mathematics achievement. The question is whether instructional quality partly mediates the relation between teacher quality and mathematics achievement. A similar question is asked by Scherer and Nilsen (Chap. 3), who examined relations between school climate, instructional quality, and achievement motivation in mathematics, asking if instructional quality mediates the relation between school climate and achievement motivation. Informed by strong theory, application of structural equation modeling can provide insights into the mechanisms through which causal effects occur. However, this kind of study also assumes that the relations among variables are causal, and that there may be omitted variables that would change the patterns of results if they were introduced to the model.

1.6.2   Multilevel Data

The sampling design of TIMSS generates data where the observations of students are nested within classes that are nested within schools. Analytical techniques for dealing with such multilevel data are available, and the studies reported here have relied on the procedures implemented in Mplus (Muthén and Muthén 1998–2012). Two levels were included in the analyses because there are few educational systems where the sample includes more than one classroom from each school, making it necessary to combine the school- and class-levels into one class level.

1.6.3   Measurement Invariance

Most of the studies presented here took advantage of measurement models with latent variables. While such models offer great possibilities for summarizing several indicators of a construct that is not directly observable while dealing with problems of measurement error, they also offer challenges, because they are based on assumptions that should not be violated. Thus, when data from multiple groups are analyzed, such as different educational systems, the latent variables must have the same meaning across groups. This can be investigated empirically through analyses of measurement invariance of the latent variables across groups.

   To answer the research questions posed by this book, so-called "metric invariance" must be established because relations between variables are to be compared across countries. This is tested through comparing the loadings of the observed indicators on the latent variables to see if they are the same; if that is the case, metric invariance is established, and relations between constructs across countries can be meaningfully compared. To be able to compare means of latent variables across countries, an added requirement would be that the means of the observed indicators, given the latent variable, are invariant across groups ("scalar invariance").

   In the analyses here, the measurement invariance of the latent constructs used was investigated. In only one case was scalar invariance supported by the data (the bullying scale in Chap. 6), but in most cases metric invariance was supported; in exceptions, separate models were fitted for each group.

1.7   Overview of Chapters

Chapter 2 examines the relations between teacher quality, instructional quality and mathematics achievement. Chapter 3 investigates the relations between school climate, instructional quality and student motivation in mathematics. Chapters 2 and 3 conducted cross-sectional secondary analysis of TIMSS 2011 data, using the

Table 1.1 Overview of the chapters

Chapter	Objective	Data and sample	Method of analysis
1	Describe conceptual framework and methodological challenges of the book	–	
2	Investigate relations between instructional quality, teacher quality and student achievement	TIMSS 2011, grade 4	Multi-group, multilevel (students and classes) SEM, mediation model
3	Investigate the relations between school climate, instructional quality and student motivation in mathematics	TIMSS 2011, grade 8	Multi-group, multi-level (students and classes) SEM, mediation models
4	Investigate the influence of teacher quality and school climate on achievement	TIMSS 2007 and 2011, grade 8	Longitudinal analyses of within-country change, difference in differences
5	Investigate if instructional quality can weaken the relation between reading and mathematics achievement	TIMSS and PIRLS 2011, grade 4	Multilevel (students and classes) SEM, random slopes models
6	Determine the degree to which instructional quality serves as a protective factor against school bullying victimization	TIMSS 2011, grade 4	Zero-inflated Poisson regression
7	Summary, discussion and concluding remarks	–	

Note: *SEM* structural equation modelling. IEA TIMSS and PIRLS 2011 data are available at http://timssandpirls.bc.edu/

grade four data set in Chap. 2 and the grade eight data set in Chap. 3, applying multi-group multilevel structural equation modeling (MG-MSEM). Chapter 4 investigates a similar research question to Chap. 3, taking advantage of TIMSS 2007 and 2011 data that are longitudinal at the country-level (Gustafsson 2013). Chapter 5 goes deeper into mathematics education, and investigates the role instructional quality plays in the relation between reading and mathematics achievement in grade four by drawing on both TIMSS 2011 and PIRLS 2011 data. In Chap. 6, instructional quality is investigated in the context of bullying experienced in grade four. Finally, in Chap. 7, we summarize the findings of the five studies, discussing both their contribution to the state of research, and limitations and further research needs (Table 1.1).

References

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 285–310). Mahwah, NJ: Lawrence Erlbaum Associates.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: WH Freeman and Co.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180.

Blömeke, S., & Delaney, S. (2014). Assessment of teacher knowledge across countries: A review of the state of research. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn* (pp. 541–585). Dordrecht: Springer.

Blömeke, S., Suhl, U., & Döhrmann, M. (2013). Assessing strengths and weaknesses of teacher knowledge in Asia, Eastern Europe and Western countries: Differential item functioning in TEDS-M. *International Journal of Science and Mathematics Education, 11*, 795–817.

Blömeke, S., Suhl, U., & Kaiser, G. (2011). Teacher education effectiveness: Quality and equity of future primary teachers' mathematics and mathematics pedagogical content knowledge. *Journal of Teacher Education, 62*, 154–171.

Bryk, A. S., & Schneider, B. L. (2002). *Trust in schools: A core resource for improvement*. New York: Russell Sage Foundation Publications.

Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D., & Sammons, P. (Eds.). (2012). *School effectiveness and improvement research, policy and practice: Challenging the orthodoxy?* Abingdon, Oxon: Routledge.

Creemers, B. P., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366.

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Abingdon, Oxon: Routledge.

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., & Hardy, I. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding? *American Educational Research Journal, 52*, 1133–1159.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132.

Engel, L. C., Rutkowski, D., & Rutkowski, L. (2009). The harsher side of globalisation: Violent conflict and academic achievement. *Globalisation, Societies and Education, 7*(4), 433–456.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9.

Ferguson, R. (2010). *Student perceptions of teaching effectiveness*. Retrieved from http://www.gse.harvard.edu/ncte/news/Using_Student_Perceptions_Ferguson.pdf.

Foy, P., Arora, A., & Stanco, G. A. (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: TIMSS & PIRLS International Study Center and International Association for the Evaluation of Educational Achievement.

Goe, L. (2007). The link between teacher quality and student outcomes: A research synthesis. *National comprehensive center for teacher quality*.

Goldstein, S. E., Young, A., & Boyd, C. (2008). Relational aggression at school: Associations with school safety and social climate. *Journal of Youth and Adolescence, 37*(6), 641–654.

Good, T. L., Wiley, C. R., & Florez, I. R. (Eds.). (2009). Effective teaching: An emerging synthesis. *International handbook of research on teachers and teaching* (pp. 803–816). New York: Springer.

Gregory, A., Cornell, D., & Fan, X. (2012). Teacher safety and authoritative school climate in high schools. *American Journal of Education, 118*(4), 401–425.

Gustafsson, J.-E. (2013). Causal inference in educational effectiveness research: A comparison of three methods to investigate effects of homework on student achievement 1. *School Effectiveness and School Improvement, 24*(3), 275–295.

Gustafsson, J. E., Hansen, K. Y., & Rosèn, M. (2013). Effects of home background on student achievement in reading, mathematics, and science at the fourth grade. In M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—implications for early learning*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Hansen, K., & Munk, I. (2012). Exploring the measurement profiles of socioeconomic background indicators and their differences in reading achievement: A two-level latent class analysis. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 5*, 49–77.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. USA: Routledge.

Hole, A., Onstad, T., Grønmo, L. S., Nilsen, T., Nortvedt, G. A., & Braeken, J. (2015). Investigating mathematical theory needed to solve TIMSS and PISA mathematics test items. *Paper presented at the 6th IEA International Research Conference, 24–26 June 2015*, Cape Town, South Africa. Retrieved from http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2015/IRC-2015_Program.pdf.

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal, 43*(3), 425–446.

Kane, T., & Cantrell, S. (Eds.). (2010). Learning about teaching: Initial findings from the measures of effective teaching project. *MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle, WA*. Retrieved from https://docs.gatesfoundation.org/Documents/preliminary-findings-research-paper.pdf.

Kane, T., & Cantrell, S. (2012). Gathering feedback for teaching. Combining high-quality observations with student surveys and achievement gains. *MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle, WA*. Retrieved from http://files.eric.ed.gov/fulltext/ED540960.pdf.

Kilpatrick, J. (2014). Competency frameworks in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 85–87). Dordrecht: Springer.

Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, & K.J. Tillmann (Eds.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 334–359). Opladen, Germany: Leske & Budrich.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). New York: Waxmann Publishing Co.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung im internationalen Vergleich. In E. Klieme & J. Baumert (Eds.), *TIMSS: Impulse für Schule und Unterricht* (pp. 43–47). Bonn: Bundesministerium für Bildung und Forschung.

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology, 100*(3), 702.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805–820. doi:10.1037/a0032583.

Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction, 18*(5), 468–482.

Kyriakides, L. (2006). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data. *Educational Research and Evaluation, 12*(6), 513–534.

Kyriakides, L., Creemers, B., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12–23.

Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal, 36*(5), 807–830. doi:10.1080/01411920903165603.

Kyriakides, L., Creemers, B., Muijs, D., Rekers-Mombarg, L., Papastylianou, D., Van Petegem, P., & Pearson, D. (2014). Using the dynamic model of educational effectiveness to design strategies and actions to face bullying. *School Effectiveness and School Improvement, 25*(1), 83–104. doi:10.1080/09243453.2013.771686.

Lipowsky, F., Rakoczy, K., Pauli, Ch., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean theorem. *Learning and instruction, 19*, 527–537.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124.

Martin, M. O., Foy, P., Mullis, I. V. S., & O'Dwyer, L. M. (2013). Effective schools in reading, mathematics, and science at fourth grade. In M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade-Implications for early learning* (pp. 109–178). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

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., Martin, M. O., Foy, P., & Arora, A. (2012a). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2012b). *TIMSS 2011 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murray-Harvey, R., & Slee, P. T. (2010). School and home relationships and their impact on school bullying. *School Psychology International, 31*(3), 271–295.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (Seventh edition). Los Angeles, CA: Muthén & Muthén.

Nilsen, T., & Gustafsson, J. E. (2014). School emphasis on academic success: Exploring changes in science performance in Norway between 2007 and 2011 employing two-level SEM. *Educational Research and Evaluation, 20*(4), 308–327.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: the Danish KOM project. *Paper presented at the Proceedings of the 3rd Mediterranean Conference on*

Mathematical Education. Atenas: Hellenic Mathematical Society. Retrieved from http://www.math.chalmers.se/Math/Grundutb/CTH/mve375/1213/docs/KOMkompetenser.pdf.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109–119.

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*, 700–712.

Richard, J. F., Schneider, B. H., & Mallet, P. (2012). Revisiting the whole-school approach to bullying: Really looking at the whole school. *School Psychology International, 33*(3), 263–284.

Ruddell, R. B., & Unrau, N. J. (Eds.). (2004). *Theoretical models and processes of reading* (5th ed.). Newark, DE: International Reading Association.

Sammons, P. (2009). The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement, 20*(1), 123–129. doi:10.1080/09243450802664321.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Scherer, R., & Gustafsson, J. E. (2015). Student assessment of teaching as a source of information about aspects of teaching quality in multiple subject domains: An application of multilevel bifactor structural equation modeling. *Frontiers in Psychology, 6*, 1550. doi:10.3389/fpsyg.2015.01550.

Schlesinger, L., & Jentsch, A. (2016). Theoretical and methodological challenges in measuring instructional quality in mathematics education using classroom observations. *ZDM*, 1–12. doi:10.1007/s11858-016-0765-0.

Schoenfeld, A. H., & Kilpatrick, J. (2008). Toward a theory of proficiency in teaching mathematics. In D. Tirosh & T. Wood (Eds.), *International handbook of mathematics teacher education* (Vol. 2, pp. 321–354). Rotterdam: Sense Publishers.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499.

Skues, J. L., Cunningham, E. G., & Pokharel, T. (2005). The influence of bullying behaviours on sense of school connectedness, motivation and self-esteem. *Australian Journal of Guidance and Counselling, 15*(1), 17–26.

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83*(3), 357–385.

Tatto, M. T., Peck, R., Schwille, J., Bankov, K., Senk, S. L., Rodriguez, M., & Rowley, G. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-MM)*. Amsterdam: The International Association for the Evaluation of Academic Achievement. Retrieved from http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/TEDS-M_International_Report.pdf.

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, 3*, 52. doi:10.1037/edu0000075.

Wang, M. -T., & Degol, J. L. (2015). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, 1–38. doi:10.1007/s10648-015-9319-1.