Chapter 7
Final Remarks

Jan-Eric Gustafsson and Trude Nilsen

Abstract This book contributes to educational policy, the field of educational effectiveness and practice. In this chapter, the findings from the five studies are summarized and discussed. After a comprehensive examination of the methodological issues related to measurement, causality, analysis, and design, implications for educational practice are proposed.

Keywords Discussion · Instructional quality · Teacher quality · Methodological issues · Contribution

7.1 Overview of the Five Studies

We first begin with a brief synopsis of the findings by chapter (Table 7.1). In summarizing the contributions of individual chapters, we employ the conceptual framework described in Chap. 1 (Fig. 1.1), which is based on the dynamic model of educational effectiveness (Creemers and Kyriakides 2008). This framework outlines the relations between the educational levels, ranging from the national level, via school and class levels to the student level.

As TIMSS 2011 does not explicitly provide detailed information on educational systems at a national level, this level is only implicitly included in the analyses. However, it is evident that the national level has influenced the findings, in that

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
Faculty of Educational Sciences, Centre of Educational Measurement at the University of Oslo (CEMO), Oslo, Norway

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

T. Nilsen and J.-E. Gustafsson (eds.), *Teacher Quality, Instructional Quality and Student Outcomes*, IEA Research for Education 2,
DOI 10.1007/978-3-319-41252-8_7

Table 7.1 A summary of the objective and the findings of each chapter

Chapter	Objective	Results
2	Investigate the relations between instructional quality, teacher quality and achievement in mathematics	Findings from the international model indicated that professional development and preparedness had, on average, the strongest relations with instructional quality and achievement. Teachers' experience influenced instructional quality and students' mathematics achievement. The teachers' attained level and major in math or math education did not matter for instructional quality, but were significantly related to mathematics achievement. Achievement was not influenced by instructional quality. At the educational-system level, findings were mixed, although professional development and preparedness had significant relations to instructional quality and student achievement in a large number of countries
3	Investigate the relations between school climate, instructional quality, and student motivation in mathematics	There was a significant positive relation between instructional quality and achievement motivation in all countries. In a number of countries, instructional quality partially mediated the relation between school climate and achievement motivation. Mediation was most apparent for an orderly school climate, and then for school emphasis on academic success. A safe climate was a mediator in only seven educational systems
4	Investigate the effects of school climate and teacher quality on mathematics achievement using country-level longitudinal analyses	Teachers' attained level of education was found to be quite strongly related to educational achievement. There were also quite substantial relations between student achievement and professional development. Teacher self-efficacy, as assessed by self-reports of preparedness for teaching in different domains, was weakly positively, but insignificantly related to student achievement. The teacher characteristics, years of teaching experience and the major academic discipline studied, had no effect. School emphasis on academic success did not satisfy ideals of unidimensionality, and only items reflecting parental support for student achievement and students' desire to perform well were significantly related to student achievement

(continued)

Table 7.1 (continued)

Chapter	Objective	Results
5	Investigate how instructional quality influences the relation between reading and mathematics achievement	All educational systems revealed a strong positive correlation between reading comprehension and mathematics achievement. Further, a number of countries demonstrated a positive relation between instructional quality and mathematics achievement and between instructional quality and reading achievement. The analysis of the moderation of the relationship between mathematics and reading by instructional quality was inconclusive; moderation was present in only six countries
6	Determine the degree to which instructional quality serves as a protective factor against school bullying victimization	Findings from the international model indicated that higher instructional quality was associated with lower rates of student self-reported bullying victimization. At the educational-system level, findings were mixed. In all systems there was a prevalence of self-reported bullying victimization at the fourth grade. However, girls and students who were more attached to their school tended to report fewer incidences of bullying victimization

most analyses yielded heterogeneous results across educational systems (from here on referred to as countries). Moreover, there was some evidence that countries with similar cultures and educational policies, such as the Nordic or Arabic countries, had similar patterns of results.

At the school level, two chapters examined different aspects of school climate. In line with the conceptual model, the results showed that school climate influenced both teachers' instructional quality and students' educational outcomes. At the classroom level, the results indicated that aspects of teacher quality were associated with instructional quality and student achievement. Instructional quality, as rated by students, had a positive relation with motivation and, in many cases, with achievement. These results are in general agreement with the conceptual model. However, the data revealed huge variations in the strengths of the different relations, something that probably is related to the specific constructs chosen, the indicators selected, and the countries examined. This result emphasizes that great care needs to be applied before generalizations can be made.

The student level included variables describing student background; although not the focus of the current research, these were used as control variables. As expected, SES, gender, and migration background were related to student outcomes

in many countries, however, the strength and the direction of relations varied. For example, it is not a given that girls are worse at mathematics than boys.

The conceptual model included both cognitive and affective outcomes. The cognitive outcomes (mostly achievement in mathematics, but also in reading) were related to variables at the teacher level (such as instructional quality as rated by students, and aspects of teacher quality) and at the school level (namely aspects of school climate). The affective outcomes (including bullying and motivation) were also related to instructional quality; motivation was also related to school climate.

7.2 Discussion of Substantive Issues

In the following discussion, we follow the structure of our conceptual framework (Chap. 1, Fig. 1.1). We start by discussing the school level, and then proceed with the teacher level and then the student level. Discussions related to the national level are included where required.

Two chapters examined reported school climate within the TIMSS 2011 grade eight data: Scherer and Nilsen (Chap. 3) investigated safe and orderly school climate and school emphasis on academic success (SEAS), while Gustafsson and Nilsen (Chap. 4) examined the role of SEAS in their country-level longitudinal design.

Scherer and Nilsen (Chap. 3) found that the three aspects of school climate were positively related to both perceived instructional quality and motivation in a number of countries. Instructional quality was positively related to motivation, and mediated the influences of an orderly climate in about half of the countries, and the influence of SEAS in about 30 % of the countries.

These results are interesting, given that few studies have established the relations between school climate, instructional quality, and student motivation (Thapa et al. 2013; Wang and Degol 2015). The findings were heterogeneous across countries, but indicated patterns for countries with similar educational systems, cultures and educational policies. For instance, the influence of an orderly climate on motivation was mediated by instructional quality in all English-speaking countries.

Gustafsson and Nilsen (Chap. 4) included the 38 countries that participated in both the 2007 and 2011 cycles of TIMSS, in a longitudinal approach. They found that only the part of SEAS that reflects parental support and students' desire to do well had a positive influence on achievement. SEAS reflects teachers, parents' and students' priorities and ambitions for academic success, The results from this chapter emphasize the importance of the parents' and students' contributions to the academic school climate, and hence extends existing research, which has, for the most part, focused on school leaders and teachers (see for example, Wang and Degol 2015).

Although both these studies accessed different outcome and school climate variables (Scherer and Nilsen focused on students' motivation and a broad range of school climate, whereas Gustafsson and Nilsen focused on student achievement and

SEAS only), their findings both confirm a positive relation between good school climate and educational outcomes in mathematics. This confirms the expectations of our conceptual model and previous reviews (see Thapa et al. 2013; Wang and Degol 2015).

At the class level, teacher quality and instructional quality were the two main constructs. Instructional quality was included in every study (except for Chap. 4, where data was not available), and the studies reported in Chap. 3 (Scherer and Nilsen), Chap. 5 (Nortvedt et al.), and Chap. 6 (Rutkowski and Rutkowski) investigated instructional quality as rated by students, and aggregated these ratings to the classroom level.

Scherer and Nilsen (Chap. 3) found that, for grade 8 students, instructional quality was positively related to all three motivational constructs (intrinsic and extrinsic motivation and self-concept) in 48 of the 50 countries. This finding is a major extension of previous research. Our literature review indicated that most previous studies conducted on relations between instructional quality and motivation were single country studies conducted primarily in Germany or the USA (see for example, Covington 2000; Fauth et al. 2014; Kunter et al. 2013; Stroet et al. 2013; Wang and Eccles 2013). Given that the sample includes countries from all continents, with diverse cultures and educational policies, the findings in Chap. 3 emphasize the need for teachers to support their students by engaging them, and providing clear and comprehensive instruction, in order to promote students intrinsic and extrinsic motivation and self-concept in mathematics.

Nortvedt et al. (Chap. 5) found perceived instructional quality to be positively related to mathematics and reading achievement in 40 % of the countries. Because their construct measured the aspects of instructional quality related to a supportive climate and clarity, this finding is important for two reasons. First, the aspects found to be strongly related to achievement in the existing research were, first and foremost, cognitive activation and classroom management (see Kunter et al. 2013); the findings of Nortvedt et al. emphasize the additional importance of these two other aspects of instructional quality. Second, the bulk of previous research on instructional quality has been conducted in Germany or the USA, and very seldom in developing countries; these findings extend this research to a more international level. Taking these two considerations together, Nortvedt et al.'s study indicates that the cultural context and the educational system play key roles in the aspects of instructional quality that are important for student achievement. For instance, aspects that are important for student achievement in Germany may differ from those important in Oman. In general though, the findings of Nordtvedt et al. are in agreement with existing research and support the idea that quality of instruction matters for student achievement (Baumert et al. 2010; Klieme et al. 2009; Good et al. 2009; Pianta and Hamre 2009; Scherer and Gustafsson 2015; Wayne and Youngs 2003).

Rutkowski and Rutkowski (Chap. 6) identified instructional quality to be negatively related to bullying internationally, in other words, higher instructional quality was associated with reduced levels of bullying. They observed significant negative relations at both the student and class level; however more educational

systems (23) exhibited significant negative relations at the student level than at the classroom level (10). At the student level, the measure refers to students' individual perceptions of instructional quality. Hence, more significant findings at the student level could reflect students' overall perceptions and attitudes to schooling. Students who are not bullied may tend to be more positive in their ratings.

Very few studies have investigated relations between bullying and instructional quality, although Kyriakides et al. (2014) examined this in five countries (Belgium, Cyprus, England, Greece, and The Netherlands) and found that the risk of bullying is reduced when students are provided with opportunities to learn and high quality teaching. Rutkowski and Rutkowski extend the limited research there is in the field by including 48 countries with diverse cultures, and by emphasizing the need for supportive teachers with clear instruction to reduce bullying.

Blömeke et al. (Chap. 2) found that instructional quality as rated by teachers was significantly related to teacher quality, and teacher quality was also related to student achievement. While previous studies have addressed sub-questions of this general relation (for example, Baumert et al. 2010; Blömeke and Delaney 2012; Goe 2007; Wayne and Youngs 2003), this is the first time a comprehensive model has been applied to almost 50 countries using a broad set of specific indicators. Specifically, in a large number of countries, Blömeke et al. identified positive relations between teachers' experience, attained level of education, and major in math or math education, and students' mathematics achievement, with countries from the same region revealing similar relational patterns. In contrast to their hypothesis, Blömeke et al. did not find significant relations between instructional quality, as rated by teachers, and achievement.

Gustafsson and Nilsen (Chap. 4) found that teachers' attained level of education and professional development had positive effects on mathematics achievement in grade eight. No study has investigated teacher quality with a longitudinal approach and with so many countries (38) before, thus these findings are important, and also support previous research (see Timperley et al. 2007). In contrast to Chap. 2, the study in Chap. 4 found that fewer aspects of teacher quality had a significant influence on achievement. Whether this was because the study in Chap. 2 investigated grade four students and Chap. 4 investigated grade eight students, or because Chap. 4 had a longitudinal approach while Chap. 2 pursued a cross-sectional approach, is difficult to disentangle, and calls for further research.

Student characteristics were related to both cognitive and affective outcomes. Blömeke et al. (Chap. 2) demonstrated a strong relation between SES and achievement, as did the other chapters and previous research (Hansen and Munk 2012). Gender differences in mathematics achievement were found in 28 countries, particularly in Europe and Latin America, unanimously in favor of boys. In Western Asian/Arabian and African countries, gender inequality was less prevalent, and, when present, the differences favored girls. This pattern may be distinguished in the international TIMSS reports (see for example, Mullis et al. 2012), although these reports present descriptive statistics; the gender patterns identified by Blömeke et al. thus extend previous research.

Rutkowski and Rutkowski (Chap. 6) found that bullying rates were generally lower for girls and for students who spoke the language of the test. Moreover, student attachment to the school was predictive of lower bullying rates in 28 out of 48 educational systems. These findings contribute in a major way to policy and to the field of bullying; the large number of countries included in Rutkowski and Rutkowski's analysis considerably extend previous research in this area (Rutkowski et al. 2013).

7.3 Methodological Issues

The complex structure of the large-scale data sets used in this report gave rise to some methodological issues, which point to the need for further research and development.

7.3.1 Measurement

The constructs investigated here are complex and challenging to measure. However, the existence of relations between different aspects of teacher quality on the one hand and student achievement on the other hand suggests that measurement of these variables has been reasonably successful. In contrast, the relative paucity and inconsistency of relations between measures of instructional quality and other variables indicate problems with the ways in which this construct has been operationalized in TIMSS.

There is long-standing controversy over whether teacher or student ratings should be used to assess instructional quality (Desimone et al. 2009; Marsh et al. 2012; Scherer et al. 2016; Schlesinger and Jentsch 2016; Wagner et al. 2015). Our results do not resolve the controversy given that they indicate that there are problems with both approaches. However, as was noted by Blömeke et al. (Chap. 2), teacher and student ratings may capture different aspects of instructional quality, which implies that they will not be highly correlated and that both may be needed for the construct of instructional quality to be adequately captured.

Instructional quality as reported by teachers or students may be affected by response-style bias caused by, for instance, cultural differences (Wagner et al. 2015). In Chap. 2, Blömeke et al. discussed further potential problems with the measurement of instructional quality by teacher reports.

Instructional quality as reported by students included the important aspects of teachers' clarity and support in learning. However, it is also desirable to capture the other two central aspects of instructional quality, namely cognitive activation and classroom management (Fauth et al. 2014; Klieme et al. 2009). TIMSS includes items in the teacher questionnaire that measure an orderly atmosphere in the school, but this refers to the school level and not to the classroom level. Including more

items and also capturing these two additional aspects of instructional quality would contribute to conceptual breadth and provide more information about whether or not different aspects of instructional quality relate differently to student outcomes.

We would like to reiterate the recommendations made by Blömeke et al. (Chap. 2) concerning development of improved measures of instructional quality. These 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 the qualities rather than the frequency of various activities.

In the studies that employed student ratings of instructional quality, the class-level relations with achievement had differing magnitude and signs. Such differences may be due to the influence of response styles, and it is important that further research investigates these issues more closely. The TIMSS and PIRLS 2011 database would be excellently suited as a basis for initial attempts to deal with these issues, because it includes both student and teacher ratings of instructional quality in three different subjects.

While the need to broaden the measures is most clearly felt for instructional quality, teacher quality also lacks sufficient variables to measure the full breadth of this construct. Teacher quality should, for instance, also include teachers' beliefs (see Goe 2007), but TIMSS does not include such measures.

7.3.2 Causal Inference

It is a well-known limitation of TIMSS that each cycle only collects cross-sectional data. With such data, it is essential not to interpret correlations as expressing causal relations. Gustafsson and Nilsen (Chap. 4) included data from two TIMSS cycles (2007 and 2011) and used analytical methods that provide better support for causal inference than data from one time-point only, because the analysis removed the effect of the omitted variables that are fixed characteristics of the educational systems. While it compensates for many omitted variables, this longitudinal approach does not protect against effects of omitted time-varying characteristics. It also assumes that the estimated causal effect has the same magnitude in all countries. The limitations of both the cross-sectional and the longitudinal techniques invite questions surrounding the agreement of results across these two approaches.

The cross-sectional study of grade four students that Blömeke et al. (Chap. 2) undertook comes closest to Gustafsson and Nilsen's longitudinal study of grade eight students. Several teacher quality indicators were available in both studies, and

with these we can make some comparisons between the pooled models in both the cross-sectional and longitudinal analyses.

For the highest level of formal education completed, there was a significant relation with achievement in the cross-sectional analysis, and this was also the case in the longitudinal analysis when separate analyses were made for OECD and non-OECD countries. A major qualification in mathematics had a significant relation with achievement in the cross-sectional analysis, but not in the longitudinal analysis, and the same pattern was found for number of years of teaching experience. Professional development showed no relation with achievement in the cross-sectional model, but was linked in the longitudinal one. Conversely, preparedness (or self-efficacy) had a significant relation with achievement in the cross-sectional analysis, but not in the longitudinal analysis.

Thus, in some cases, the findings of the two studies overlapped, and, in other cases, not. This calls for further research. The longitudinal analysis estimates a common effect for all educational systems, thus if there are differences between educational systems in the strength and sign of effects this may cause the effects of positive and negative relations to cancel each other out. This could partly explain the lack of significant findings in Chap. 4.

We know from many published studies that there is a positive relation between SEAS and achievement at both student and class levels (see for example, Nilsen and Gustafsson 2014; Martin et al. 2013). In the longitudinal study, there was a significant effect of SEAS on achievement in a latent variable model with SEAS defined by five items. However, when separate analyses were performed for each item, the relation was found to be due to one item asking about parental support, and to one item asking about students' desire to learn. These results suggest that parents and students are more important for the relation between achievement and SEAS than are school factors. The results also suggest that the SEAS construct is multi-dimensional, which needs to be further explored.

The country-level longitudinal approach is a simple way of strengthening the credibility of causal statements based on ILSA data (Gustafsson 2013). As more countries participate in adjacent cycles, the approach becomes more powerful. However, the approach also requires that items are maintained unchanged over cycles; it is thus a great pity to see that few of the items included in TIMSS 1995 are still in use. It is therefore essential that, as questionnaires are changed to improve the measurement of constructs, new items are added while, at the same time, old items are kept.

7.3.3 Design

TIMSS is designed so that only one class per selected school is typically included in the sample; this causes the school and class levels to be confounded. This is unfortunate given that there are both theoretical and empirical reasons to expect that school- and class-level factors and processes are differentially related to

achievement (Yang Hansen et al. 2014). Furthermore, software is now available to allow powerful analyses of such three-level data. We therefore recommend sampling of two classes from each school, when possible.

7.4 Implications

This book can contribute to educational policy, the field of educational effectiveness and practice. Educational policy may benefit from the study findings that point to the importance of teacher quality, and especially teacher education and professional development for high instructional quality and for student achievement in mathematics. Instructional quality was also found to be related to school climate and to student motivation in mathematics. Hence, providing first and foremost an orderly school climate, but also a climate where teachers, students and parents collectively prioritize success and learning, may create the foundations for high instructional quality and boost student motivation in mathematics. This finding is extremely important in addressing the international challenges related to the decline of students' participation in STEM-related studies and careers (OECD 2014). Moreover, the results identify the potential importance of instructional quality in reducing bullying.

The studies found that there were large cultural diversities and heterogeneous findings across the educational systems with respect to the relation between teacher quality, instructional quality and student achievement. Nevertheless, patterns could be identified within groups of countries, confirming previous research that identified countries clustering (Olsen 2006; Olsen et al. 2005). Further research in this area could result in policy-relevant differentiation of knowledge for different categories of educational systems.

Our findings extend existing research on the importance of school climate by: (1) including a wide range of countries across all continents, (2) including three aspects of school climate in the same study (SEAS, safety, and order), and (3) identifying relations with student motivation and instructional quality.

Our work also contributes by applying advanced methodology in the context of international large-scale surveys. Some of the methods used in this book are new and were not previously applied in the field of educational research (such as SEM for longitudinal country-level analyses). The results highlight the integral challenges with some methods (such as using random slopes on the class level), and suggest the need for further methodological research.

7.5 Concluding Remarks

We have investigated countries from all over the world and performed both cross-sectional and trend analyses, while incorporating school and student home background contexts. The studies demonstrated the importance of teacher quality, school climate, and instructional quality for educational outcomes, and although there is not yet a coherent and international understanding of these relations, this research demonstrates progress and the value of international large-scale surveys. ILSAs view the world as a global educational laboratory, providing golden opportunities to investigate questions important to educational policy, research and practice.

References

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. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM, 44*, 223–247.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51*(1), 171–200. doi:10.1146/annurev.psych.51.1.171.

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

Desimone, L., Smith, T., & Frisvold, D. (2009). Survey measures of classroom instruction: Comparing student and teacher reports. *Educational Policy, 24*, 267–329.

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.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.gtlcenter.org/sites/default/files/docs/LinkBetweenTQandStudentOutcomes.pdf.

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

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.

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.

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, NY: Waxmann Publicing Co.

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.

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.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., 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. doi:10.1080/00461520.2012.670488.

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.

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

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.

OECD. (2014). *Education at a Glance 2014: OECD Indicators*. Paris: OECD Publishing. doi:10.1787/eag-2014-en.

Olsen, R. V. (2006). A Nordic profile of mathematics achievement: Myth or reality? *Northern lights on PISA 2003: A reflection from the Nordic countries*, 33–45.

Olsen, R. V., Kjærnsli, M., & Lie, S. (2005). *Similarities and differences in countries' profiles of scientific literacy in PISA 2003*.

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

Rutkowski, D. J., Rutkowski, L., & Wild, J. (2013). Predictors of school violence internationally: The importance of immigrant status and other factors. Paper presented at the 5th IEA International Research Conference, Singapore. Retreived from http://www.iea.nl/irc-2013.html

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.

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology, 7*, 110. doi:10.3389/fpsyg.2016.00110.

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.

Stroet, K., Opdenakker, M.-C., & Minnaert, A. (2013). Effects of need supportive teaching on early adolescents' motivation and engagement: A review of the literature. *Educational Research Review, 9*, 65–87. doi:10.1016/j.edurev.2012.11.003.

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.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Wellington: Ministry of Education.

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://dx.doi.org/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.

Wang, M.-T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28*, 12–23. doi:10.1016/j.learninstruc.2013.04.002.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89–122.

Yang Hansen, K., Gustafsson, J.-E., & Rosén, M. (2014). School performance differences and policy variations in Finland, Norway and Sweden. In *Northern lights on TIMSS and PIRLS 2011*. Denmark: Nordic Council of Ministers.