

Hochweber, Jan; Vieluf, Svenja

Gender differences in reading achievement and enjoyment of reading. The role of perceived teaching quality

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:

formally and content revised edition of the original source in:

The journal of educational research 111 (2018) 3, S. 268-283



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /
Please use the following URN or DOI for reference:

urn:nbn:de:0111-pedocs-170406

10.25656/01:17040

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-170406>

<https://doi.org/10.25656/01:17040>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

This is an Accepted Manuscript of an article published by Taylor & Francis in The Journal of Educational Research in 2018 (Volume 111, issue 3), available online:

<http://www.tandfonline.com/10.1080/00220671.2016.1253536>

Gender Differences in Reading Achievement and Enjoyment of Reading:
The Role of Perceived Teaching Quality

Jan Hochweber

University of Teacher Education St. Gallen, St. Gallen, Switzerland

Notkerstrasse 27, 9000 St. Gallen, Switzerland

Phone: +41 71 243 96 65 Email: jan.hochweber@phsg.ch

Svenja Vieluf

DIPF | Leibniz Institute for Research and Information in Education

Rostocker Straße 6, 60323 Frankfurt am Main, Germany

Phone: +49 69 24708 222 Email: vieluf@dipf.de

This article may not exactly replicate the final version published. It is not the copy of record.

The final article is available at: <https://doi.org/10.1080/00220671.2016.1253536>

© 2016 Taylor & Francis Group, LLC

Abstract

We examined the extent to which classroom-specific relationships between students' gender and their reading achievement and enjoyment of reading are associated with student-perceived teaching quality. Based on a sample of 10,543 9th-graders from 427 classrooms, multilevel analyses revealed that effective classroom management, adequate pacing, and a strong focus on language competencies were related to a less pronounced increase of girls' advantage in reading achievement during 9th grade. High levels of teacher support and focus on language competencies were related to smaller gender differences in enjoyment of reading at the beginning of 9th grade, though not associated with change of these differences over the school year. Our findings suggest that high teaching quality is not only related to higher reading achievement and reading enjoyment in classrooms as a whole, but may also help to mitigate the increase of gender gaps in reading achievement and motivation commonly observed in secondary school.

Keywords: enjoyment of reading, gender differences, reading achievement, reading motivation, teaching quality

Gender Differences in Reading Achievement and Enjoyment of Reading:
The Role of Perceived Teaching Quality

Gender differences in educational achievement have been a topic in the educational discourse for decades. While previous work was mainly concerned with disadvantages of girls, recently the attention has shifted to the underachieving boy. The “boys’ debate” is fueled by lower attainment and poorer examination results of boys, but also by specific disadvantages, most notably, in reading. Over several decades, elementary school studies found boys to score poorer on reading tests than girls (e.g., Gates, 1961; Masters & Forster, 1997; Pauley, 1951). Results from the large-scale assessment PIRLS, which assessed fourth-graders’ reading competencies in 49 countries—including the United States, Canada, and 23 European countries—suggest that these findings are generalizable across a variety of education systems (Mullis, Martin, Foy, & Drucker, 2012). These gender differences persist into secondary education. For instance, in the PISA 2012 study, 15-year-old girls outperformed boys in reading in all 65 participating countries and economies (OECD, 2014; see also Lietz, 2006). A thorough overview of large-scale assessments demonstrating girls’ advantage in reading throughout their school career is given by Miller and McKenna (2016).

Corresponding with boys’ lower reading achievement, a rich body of research indicates that boys are less interested in and engaged with reading than girls. Boys were found to have less positive attitudes towards academic and recreational reading (Coles & Hall, 2002; McKenna, Conradi, Lawrence, Jang, & Meyer, 2012; Sainsbury & Schagen, 2004), to place less value on reading (Durik, Vida, & Eccles, 2006; Eccles, Wigfield, Harold, & Blumenfeld, 1993), and to have poorer intrinsic motivation to read (Marinak & Gambrell, 2010; McGeown, Goodwin, Henderson, & Wright, 2012). Gender differences in reading motivation already occur in elementary school (Durik et al., 2006; Marinak & Gambrell, 2010) and

persist into middle and high school (Durik et al., 2006; McKenna et al., 2012; Stanat & Kunter, 2002; Swalander & Taube, 2007).

Since reading plays a key role both for learning in various domains and participation in cultural and societal activities, boys' lower reading achievement and lack of motivation to read is a cause of concern for policy makers and practitioners. Given the vital role schooling plays in children's learning to read, educational processes are likely to have considerable influence on the gender gap in reading, particularly the learning that takes place in the classroom. However, up to now, knowledge about classroom characteristics that are related to the gender gap in reading is limited. While a large deal of debate has evolved around the relevance of context characteristics such as teachers' gender (e.g., Marsh, Martin, & Cheng, 2008) and single-sex schooling (e.g., LePore & Warren, 1997), empirical studies have rather not confirmed their hypothesized influence. One possible reason is that these variables are too detached from student learning, and it might prove more insightful to investigate the processes actually taking place in the classroom.

In this study, we focused on the role of three global factors or "basic dimensions" of teaching quality—structure, cognitive activation, and support—in shaping classroom-specific gender differences with respect to reading achievement and an important aspect of intrinsic reading motivation, enjoyment of reading (Guthrie & Wigfield, 2000). Rather than representing specific teaching practices (e.g., setting up small-group work), these basic dimensions capture the global quality of classroom processes, based on cognitive-constructivist or socio-constructivist concepts of teaching and learning (see Decristan et al., 2015). Practice-oriented guidelines and field reports stress the relevance of quality of teaching for reducing the gender gap in reading (e.g., Alloway, Freebody, Gilbert, & Muspratt, 2002; Cresswell, Rowe, & Withers, 2002), but pertinent empirical research is scarce. By exploring the role of teaching quality, our study contributes to finding starting points to enhance gender

equity in reading. In the following, we briefly address the basic dimensions of teaching quality. We then describe how we suggest these dimensions to be related to the gender gap in reading achievement and enjoyment of reading.

Basic Dimensions of Teaching Quality

In both U.S. and European classrooms, researchers have independently and consistently identified three global factors or basic dimensions of teaching quality. U.S.-based research is strongly rooted in work by Pianta, Hamre, and colleagues in pre-kindergarten, kindergarten, and elementary classrooms. The Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) and its predecessor, the Classroom Observation System (COS; e.g., National Institute of Child Health and Human Development Early Child Care Research Network, 2002), are observational measures capturing the quality of teacher–child interaction in a classroom. The CLASS framework encompasses three major domains to describe the quality of classroom processes: classroom organization, instructional support, and emotional support (e.g., Pianta & Hamre, 2009; Reyes, Brackett, Rivers, White, & Salovey, 2012).

In the European context, Klieme, Pauli, and Reusser (2009) presented a theoretical framework of teaching quality that has been developed in the course of the 1995 TIMSS video study (Klieme, Schümer, & Knoll, 2001) and extended in the video intervention study “Quality of Instruction, Learning, and Mathematical Understanding” (Klieme et al., 2009). Their model assumes three comparable dimensions, (a) classroom management, clarity, and structure, (b) cognitive activation and deep content, and (c) supportive climate (for similar models see Baumert et al., 2010; Kunter et al., 2013; Lipowsky et al., 2009). Complementary to Pianta et al.’s work, research based on this framework has focussed mainly on secondary school classrooms. In this study, we primarily refer to Klieme et al.’s (2009) conceptualization, and will thus give a short introduction.

The first dimension, in the following abbreviated to “structure”, is rooted in research designs from the process-product paradigm, such as “direct instruction” (e.g., Rosenshine & Furst, 1971). Features indicating a high quality are a clear and explicit structure of the lesson, a clear and unequivocal presentation of the content, good classroom management, adequate pacing, close monitoring of student progress, and informative and encouraging feedback. The second dimension, in the following abbreviated to “cognitive activation”, goes back to cognitive-constructivist theories. Instruction is cognitively activating when it uses challenging content and initiates problem solving in complex, authentic situations (Brown, 1994; Stefanou, Perencevich, DiCintio, & Turner, 2004). The third dimension, in the following abbreviated to “support”, is based on self-determination theory. A high level of support implies that students experience autonomy, competence, and social relatedness during lessons (Deci, Ryan, & Williams, 1996; Reeve, 2002).

Teaching Quality and Student Outcomes

There is growing evidence that structure, cognitive activation, and support—or, similarly, classroom organization, instructional support, and emotional support—are related to students’ academic achievement and motivation from preschool and elementary to secondary school.

Studies conducted in preschools and elementary schools indicate that structure, cognitive activation, and support are positively related to student achievement and to non-cognitive outcomes such as students’ engagement and social skills (e.g., Cadima, Leal, & Burchinal, 2010; Howes et al., 2008; Mashburn et al., 2008; Reyes et al., 2012; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Moreover, some results suggest that structure rather than support predicts achievement, and support is more strongly related to non-cognitive outcomes (Curby et al., 2009a; Howes et al., 2008; Mashburn et al., 2008).

There are, however, also contradictory findings (e.g., Curby, Rimm-Kaufman, & Ponitz, 2009b; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008).

Studies in secondary schools indicate that cognitive activation is especially relevant to student achievement (Baumert et al., 2010; Kunter et al., 2013), while support seems to be more closely associated with student motivation (Klieme, Lipowsky, Rakoczy, & Ratzka, 2006; Kunter et al., 2013). Structure has been found positively related to achievement (Möller, Jonen, Hardy, & Stern, 2002; Pauli, Drollinger-Vetter, Hugener, & Lipowsky, 2008), but may also enhance motivation (Rakoczy, 2006). Previous research has focused primarily on mathematics, but several studies also support the relevance of these dimensions to the language domains (e.g., Rjosk et al., 2014, 2015). Nevertheless, some research indicates that, in comparison to mathematics, structure may be less relevant to students' language-related outcomes (Hochweber, Steinert, & Klieme, 2012; Rjosk et al., 2014).

Teaching Quality and the Gender Gap in Reading

How might differences in teaching quality become relevant to the gender gap in reading? In many if not most cases, this should be due to differences in other student characteristics which are related to gender. Research rooted in the paradigm of aptitude-treatment-interaction (ATI) has emphasized the role of cognitive, conative, and affective student characteristics moderating the impact of treatments. By being different with respect to such characteristics, girls and boys should profit to different extents from certain aspects of instruction, for example, the pacing of the lesson (Bank, Biddle, & Good, 1980). Hence, these aspects should tend to increase or decrease gender differences in reading achievement and reading motivation. This notion is also compatible with the differential sensitivity hypothesis (e.g., Dar & Resh, 1997), stating that compared to high achievers' performance, low achievers' performance is more school- and classroom-dependent, resulting, among others, from a lack of support and stimulation received at home. Actually, parents provide girls with

more opportunities to make literacy experiences already in early childhood (Millard, 1997; Silinskas et al., 2010) and tend to view girls as more competent readers than boys (Eccles, Jacobs, & Harold, 1990; Rätty, Kasanen, & Kärkkäinen, 2006), which might negatively affect boys' self-perceptions and achievement. Thus, boys might be both more advantaged by high-quality teaching and more disadvantaged by low-quality teaching.

The Role of Structure

Features characterizing well-structured instruction, such as high clarity and adequate pacing of the lessons, seem to particularly benefit low-ability students' learning (Möller et al., 2002; Snow & Lohman, 1984). Less clearly structured instruction seems to be more detrimental to low-ability learners, in particular those with low metacognitive learning skills (Veenman & Elshout, 1995). Given that boys are overrepresented in the lower part of the achievement curve and more often show ineffective metacognitive learning strategies (e.g., Zimmerman & Martinez-Pons, 1990), adequately paced and clearly structured instruction might help to reduce the gender gap in reading by particularly promoting boys' reading skills.

Furthermore, effective classroom management (Emmer & Stough, 2001) might be especially important to regulate boys' behavior in classrooms, and thereby contribute to reducing the gender gap in reading achievement. Boys are known to have a less positive attitude to school, to be more distracted and to show more disruptive behavior in the classroom, and to more often break school rules (Buchmann, DiPrete, & McDaniel, 2008; Johnson, McGue, & Iacono, 2005). Moreover, research suggests that boys may be more susceptible to influences of a negative classroom learning climate than girls (Van de gaer, Pustjens, Van Damme, & De Munter, 2006). Since good classroom management may foster student engagement and reduce student misconduct (Weinert & Helmke, 1995), it might help to counter behavioral tendencies that particularly hamper boys' learning in classrooms.

The Role of Cognitive Activation

Research on the interplay between specific student characteristics and cognitive activation due to instruction is scarce. However, some studies in science teaching suggest that gender differences can be reduced by instructional strategies associated with enhanced cognitive activity. Chambers and Andre (1997) found female college students to have less initial experience and interest in the topic of electricity than their male counterparts, but to profit relatively more from a cognitively activating text type. In Geier et al.'s study (2008), inquiry-based instruction helped to reduce the disadvantage in science achievement experienced by African-American boys in urban middle schools. Burbules and Linn (1988) found that boys showed better learning progress than girls in making predictions about water displacement, but the difference was less pronounced when girls' lack of background knowledge was compensated by an early opportunity to reflect about the experimental apparatus. Focusing on a process including scaffolded inquiry, reflection, and generalization, White and Frederiksen (1998) found no gender differences in the effects of the scaffolding process, but reported that inquiry learning can reduce the gap between more and less proficient students in their understanding of inquiry processes.

Together, these studies provide some evidence for differential effects of cognitively activating teaching depending on student gender and prior achievement. They did not address reading achievement; but since cognitive activation is thought to be crucial to learning in general (Mayer, 2004) and is related to learning gains in several domains (e.g., Baumert et al., 2010; Rjosk et al., 2014), they may have implications for the gender gap in reading. As boys tend to show lower reading abilities, and even more, lower motivation to read, they might benefit more than girls from cognitively challenging instruction. There are, however, also contradictory findings. In a study by Von Secker and Lissitz (1999), for instance, boys

showed higher initial science achievement than girls but also profited more from an emphasis on critical thinking, whereas teacher-centered instruction was found to reduce the gender gap.

The Role of Support

Teacher support is assumed to foster student motivation by establishing positive teacher-student relationships. Two longitudinal studies indicate that boys might indeed benefit relatively more from supportive teacher-student relationships. Hamre and Pianta (2001) studied the power of teacher-student relationships in the kindergarten to predict academic and behavioral outcomes from lower elementary to middle school. The correlations of early teacher-student relationships with later academic outcomes and disciplinary infractions tended to be higher for boys than for girls. For students in 3rd to 6th grade, Furrer and Skinner (2003) found students' sense of relatedness to be more strongly associated with academic engagement for boys than for girls. A possible explanation is that, with boys being often less academically engaged than girls, "interpersonal ties to the teacher [...] provide them with a bigger motivational boost" (ibid., p. 159). Fostering these ties might be particularly important in reading instruction, to compensate for a lack of support from boys' homes in earlier stages of reading acquisition (Millard, 1997; Silinskas et al., 2010).

If boys profit relatively more from positive relationships with teachers, a higher level of teacher support might decrease the commonly observed gender gap in reading motivation. However, empirical results are not consistent. Several studies found no interactions between gender and variables such as teacher support and sense of belonging when predicting students' academic, motivational, or behavioral outcomes (e.g., Fan, 2011). Others indicated that associations of teacher-student relationships with these outcomes may in fact be stronger for girls (e.g., Goodenow, 1993).

The Present Study

Our study investigated the extent to which student-perceived teaching quality is related to change in the gender gap in reading achievement and enjoyment of reading. We assumed that teaching quality is not only relevant to the general level of achievement and learning motivation in classrooms but may also contribute to an increase or decrease in gender differences. Our analysis was based on a large, nationally representative sample of German 9th-graders whose reading achievement and enjoyment of reading were assessed at two time points at the beginning and the end of the school year.

A total of seven instructional features representing the basic dimensions of teaching quality were used to predict change in the classroom-specific gender gap in reading achievement and enjoyment of reading (see Figure 1). These features were assessed by student ratings aggregated to the classroom level. Several researchers have argued that student ratings can be reasonably reliable and valid measures of students' learning environments, given that theoretical and measurement issues are appropriately taken into account (e.g., Aleamoni, 1999; De Jong & Westerhof, 2001; Marsh, Ginns, Morin, Nagengast, & Martin, 2011). Due to aggregation, idiosyncrasies in individual perceptions tend to be eliminated and more objective measures of the learning environment can be obtained (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009). Nevertheless, it has to be emphasized that the present study is based on individual perceptions of teaching quality rather than more objective measures that might have been gained, in particular, from classroom observations.

Four variables indicating the quality dimension *structure* were included: classroom management, structuredness of instruction, clarity of instruction, and pacing. Effective classroom management fosters student engagement and reduces student misconduct, and should thus counter behavioral tendencies detrimental to learning that are more commonly found among boys. High structuredness and clarity of instruction as well as adequate pacing

should be particularly supportive to low-ability students' learning, and thus benefit boys as well. Hence, we predicted that these instructional features are related to a decrease in the gender gap in reading achievement typically found in favor of the girls. As in the language domains structure has been found associated with student achievement rather than student motivation, we did not assume relationships with the gender gap in enjoyment of reading.

One indicator for the quality dimension *cognitive activation*—teachers' focus on students' language-related competencies—was included. It refers to the targeted attention a teacher pays to language-related aspects like writing a text or correctly using spelling and punctuation rules. A teacher who places a special focus on language during instruction is assumed to set clear and high demands and to offer cognitively challenging learning opportunities (Rjosk et al., 2014). Studies in science teaching suggest that a high level of cognitive activation might particularly benefit students with lower achievement and motivation in the subject. Thus, we hypothesized that a high level of cognitive activation should particularly benefit boys, and consequently, decrease the gender gap in reading achievement. As cognitive activation has been found related to student achievement rather than student motivation, we did not assume relationships with gender differences in enjoyment of reading.

Finally, we included two indicators of the quality dimension *support*, teacher support of their students and teacher orientation towards their students. Both variables should foster positive student-teacher relationships. Although previous research is not completely conclusive, two studies (Furrer & Skinner, 2003; Hamre & Pianta, 2001) suggest that boys may profit more than girls from supportive relationships to their teachers. This seems particularly plausible in reading, since teachers might thereby compensate for a lack of support from boys' homes. Accordingly, given that girls generally report more enjoyment of reading than boys, higher levels of teacher support and teacher orientation towards their

students should be associated with a decrease in the gender gap in reading enjoyment. Since the quality dimension support has been found related to student motivation rather than student achievement, we did not assume relationships with the gender gap in reading achievement.

Method

Sample

This study was based on data from DESI, a longitudinal study of 9th-grade students' language competencies and language instruction in Germany (DESI-Konsortium, 2008). German and English as a foreign language competencies were assessed at the beginning (T1) and the end (T2) of school year 2003/2004. A total of 219 schools were selected randomly from all 16 German federal states, and two Grade 9 classrooms within each school were sampled randomly. Participation in the study was compulsory for all students in these classrooms. Sampling weights were constructed to account for unequal probability of selection, and all statistics reported in this article were computed based on these weights.

Our sample consisted of 10,543 students, attending 427 classrooms. 50.7% were female; the average age at T1 was 14.9 years ($SD = 0.74$). 81% of the students reported they had learned German as first language, 13% learned another language than German, and 6% learned German and another language. Class sizes ranged from 5 to 36 students ($M = 24.7$, $SD = 5.3$). Of students' teachers, 61.8% were female; 7.5% were under 30, 25.3% between 30 and 40, 29.7% between 41 and 50, 32.5% between 51 and 60, and 5.0% above 60 years old. Teachers' average experience in teaching German language courses was 19.3 years ($SD = 11.3$).

Variables

All except two variables were based on the reading achievement test and the student questionnaire administered at T1 and T2. One control variable (teacher gender) was assessed in the teacher questionnaire at T2, and another control variable (socioeconomic status) was

assessed in the parent questionnaire at T2. All scale items are reported in Appendix A1.

Descriptives for all study variables are presented in Appendix A2.

Student level. Two outcomes (reading achievement and enjoyment of reading) and five predictor variables (gender and four control variables) were included at the student level.

Reading achievement. Reading achievement was assessed at T1 and T2 by a test covering the content prescribed in the German federal states' curricula. It consisted of four fictional texts (e.g., short stories) and four non-fictional texts (e.g., newspaper articles), and comprised 38 items. They were distributed over four booklets, of which each included 18 to 20 items. A rotated booklet design was used to ensure that students did not answer the same items at T1 and T2. Longitudinal scaling was done based on a multidimensional Rasch model (e.g., Briggs & Wilson, 2003), where each time point was represented as one dimension. In our analyses, we used weighted-likelihood estimates (WLE). The reliability estimated from independent plausible value draws (EAP/PV reliability; e.g., OECD, 2009) was .70 across both measurement occasions.

Enjoyment of reading. Enjoyment of reading was measured at T1 and T2 by a three-item scale with a 4-point Likert response format from the PISA 2000 assessment (e.g., "Because reading is fun, I wouldn't want to give it up"). Students scoring high on this scale get easily absorbed in reading and spend time outside school on reading activities. The internal consistency (Cronbach's alpha) was .85 at T1 and .88 at T2.

Gender. Student gender was assessed in the tracking form and was available for all students (0 = *male* and 1 = *female*).

Control variables. Three important potential confounding variables were included as control variables. *Socioeconomic status* (SES) of students' families was measured by the International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom, de Graaf, Treiman, & de Leeuw, 1992). Parents' occupation was assessed in the parent questionnaire.

ISCO codes (ILO, 1990) were assigned to the responses, which were mapped to the ISEI. The highest ISEI of both parents (HISEI) was included as an indicator of student SES. *Basic cognitive abilities* were assessed by the Figure Analogies subscale of the German version of the Cognitive Ability Test (Thorndike & Hagen, 1993), which is highly related to aspects of general intelligence and represents a parsimonious measure of basic cognitive abilities. Finally, we included students' *Language use at home*. Two dummy variables were created, one for students who used a language *other than German* and one for students who both used *German and another language*, with a use of *only German* as the reference category.

Classroom level. Seven scales assessing aspects of teaching quality and three control variables were included as predictors.

Teaching quality. Seven scales with a 4-point Likert response format were selected from the student questionnaire. Each was aggregated to the classroom level to capture students' shared perception of instruction. *Classroom management* was measured by two items, assessing the extent to which the teacher is competently managing students' behavior during lessons (e.g., "My German teacher always knows exactly what is happening in the classroom"; $\alpha = .78$). *Structuredness* was measured by three items, assessing the extent to which teachers explicate the goals and structure of the lesson (e.g., "My German teacher points out the relevant aspects of a lesson"; $\alpha = .79$). *Clarity* was measured by three items (e.g., "The assignments in the German language classrooms are clear and comprehensible to me"; $\alpha = .78$). High scores indicated that teachers express themselves clearly in their presentation and concerning their expectations towards students' work. *Pacing* was measured by four items (e.g., "In German language lessons I often do not get my work done"; $\alpha = .70$). High scores indicated that the teacher progresses very quickly, interrupts the students frequently, and imposes time pressure on them to get their work done, that is, indicated *inadequate* pacing. The scale was recoded to ensure that high scores expressed *adequate*

spacing. Also, one item had to be excluded from the original spacing scale because of a low item-scale correlation (“When my German teacher asks me a question, s/he gives me enough time to reflect upon it”). *Focus on language competencies* was measured by ten items, assessing the importance of language-related competencies during instruction (e.g., “How important is it for your German teacher that you can write grammatically correctly?”; $\alpha = .89$). *Teacher support* was measured by four items, assessing the extent to which the teacher supports students by actively aiding them in the learning process (e.g., “My German teacher gives me advice on how to improve”; $\alpha = .84$). *Student orientation* was measured by six items (e.g., “If somebody has a good idea, my German teacher acknowledges it”; $\alpha = .87$). High scores indicated that students may express their ideas and opinions, and the teacher takes them into account when making instructional decisions.

Control variables. Since gender composition and single-sex schooling have been hypothesized to influence the gender gap, we controlled for the *Proportion of girls*, and *Teacher gender*, as reported in the teacher questionnaire, coded 0 = *male* and 1 = *female*. Additionally, we controlled for *school type*. German secondary school types differ substantially with respect to student composition, pedagogical tradition, and so forth, so that erroneous conclusions regarding other variables’ effects might result when not taking school type into account. Three school types seemed important to consider: *Gymnasium*, the most academically demanding school type; *Realschule*, the intermediate level school type; and schools offering the least academically demanding track, *Hauptschule*. The school types offering this track—stand-alone *Hauptschule*, *Mittelschule* (catering for *Realschule* and *Hauptschule* students), and comprehensive schools—are comparable with respect to their students’ average language competencies (Steinert, Hartig, & Klieme, 2008), and were thus combined into a single category. Descriptives for the three remaining school type categories are presented in Appendix A3. As would be expected, there were statistically significant

differences with respect to the outcome variables and several of the teaching quality variables. To represent school type in our models, we created weighted effect code variables, one representing *Realschule* and one *Gymnasium*, with *Hauptschule* as the reference category. Their values were selected such that the means and intercepts at the classroom level refer to an average classroom with respect to all classrooms in the entire sample.

Analysis

Multilevel modeling. We estimated a series of multilevel models (Hox, 2010; Raudenbush & Bryk, 2002) with students (Level 1) nested in classrooms (Level 2). The models were built up in two major steps. In Step 1, we specified a multivariate multilevel regression model with reading achievement and enjoyment of reading at T1 and T2 as outcomes. Student gender and the control variables (SES, basic cognitive abilities, language use at home) were entered as student-level predictors. The gender effects on reading achievement and enjoyment of reading at both time points were allowed to vary between the classrooms, that is, we specified a random slope for each of these effects. At classroom level, four random intercepts reflected between-classroom differences in average reading achievement and enjoyment of reading at T1 and T2, respectively. Four random slopes reflected between-classroom differences in the gender–achievement relationship and the gender–enjoyment relationship at T1 and T2, respectively.

To analyze change in the classroom-specific gender gaps during 9th grade, latent change models were imposed on the classroom-level random effects. Models for latent change (Steyer, Partchev, & Shanahan, 2000) are based on repeatedly measured constructs, where each measurement is represented by a latent variable. Additional latent variables are then specified to express the relationships between these variables in terms of initial status and change. In a typical pretest-posttest design, an initial status factor captures differences between persons at time point 1, and a change factor captures differences between persons in

change between time point 1 and 2. In our analyses, latent change models were specified at the *classroom level* to capture differences between classrooms in initial status and change concerning the reading achievement level, the enjoyment of reading level, the gender–achievement relationship, and the gender–enjoyment relationship. The resulting model served as a baseline model for the next step.

In Step 2, classroom-level predictors were added. Figure 2 illustrates this type of model. First, the instructional features were entered one by one, controlling for proportion of girls, teacher gender, and school type. Each predictor was used to predict all initial status and change variables. Second, the statistically significant predictors of change in the gender gaps were entered into a model simultaneously. We decided for this “step-up” strategy, since several instructional features were noticeably correlated (cf. Appendix A2).

Estimation and testing. All models were estimated in Mplus 7 (Muthén & Muthén, 1998-2012) using robust maximum likelihood estimation. For significance testing of the variance components and of the directed hypotheses formulated above, one-sided Wald tests were used (Hox, 2010). Otherwise, two-sided Wald tests were used.

Standardization. To facilitate the interpretation of the results, all continuous variables were standardized. Student-level predictors and outcomes at T1 (reading achievement, enjoyment of reading, SES, basic cognitive abilities) were standardized based on their overall mean and variance in the sample. Outcomes at T2 were standardized based on their mean and variance at T1. Classroom-level predictors (instructional features, proportion of girls) were standardized based on their mean and variance across the classrooms. Categorical predictors were not standardized. Due to standardizing the continuous student-level predictors, between-classroom differences in student characteristics were taken into account when estimating the effects on classrooms’ initial status and change of reading achievement and enjoyment of reading (Raudenbush & Bryk, 2002).

Missing data. Missing data occurred for two reasons. First, a matrix design had been implemented for the student questionnaire. Of the variables in this study, the focus on language competencies items had been included in only one of two questionnaire versions. Consequently, the values on these items were missing by design for about 50% of the students. Since the questionnaires were distributed randomly, no bias due to missing data had to be expected. Second, data was missing since some students did not take part in the reading assessment, and some students and teachers did not fill in any or all questionnaire items. As a result, there was missing data on all variables except student gender, proportion of girls, and school type.

Since traditional missing data approaches (e.g., listwise deletion) may produce biased results, we applied multiple imputation (MI; Schafer & Graham, 2002). In MI, each missing value is replaced by a set of predicted values. MI results can be biased if the hierarchical data structure is not taken into account (Van Buuren, 2010). For this reason, we used the MI facility in Mplus which enables imputation based on unrestricted (H1 imputation) or specifically restricted (H0 imputation) multilevel models (Asparouhov & Muthén, 2010). Since H1 imputation in Mplus does not cover models with random slopes, we used H0 imputation and specified a multivariate twolevel model with random slopes containing all analysis variables. Recent research has shown that relatively small numbers of imputations (e.g., five) may lead to serious power falloff, and it was recommended to use many more imputations (Graham, Olchowski, & Gilreath, 2007). While 100 or more imputations may be optimal, we decided to use 40 imputations which seemed an acceptable trade-off given the considerable increase in computation time.

Results

Gender Differences in Reading Achievement and Enjoyment of Reading in 9th Grade

The results from the multilevel models are presented in Table 1. Since our hypotheses refer to relationships at the classroom level, only the classroom-level results are shown. In Model M0, gender, SES, basic cognitive abilities, and the language at home variables were entered as student-level predictors, with all gender effects specified as random. At classroom level, we specified four latent change models to capture the classrooms' initial status and change in reading achievement level, enjoyment of reading level, the gender–achievement relationship, and the gender–enjoyment relationship. This model is equivalent to the model illustrated in Figure 2, except that the classroom-level predictors were not yet included. The intercepts for reading achievement and enjoyment of reading refer to these variables' average levels across all classrooms at the beginning of 9th grade and the average change in these levels across all classrooms during 9th grade. The intercepts for the gender–achievement and gender–enjoyment relationships refer to the average gender differences in reading achievement and enjoyment of reading across all classrooms at the beginning of 9th grade and the average change of the gender differences during 9th grade. In the following, we focus on the gender-related results (lower part of Table 1).

At the beginning of 9th grade, girls reported a higher enjoyment of reading than boys ($B = 0.560$), but the gender difference in reading achievement ($B = 0.024$) was not statistically significant. In the course of 9th grade, the gender gaps in reading achievement and enjoyment of reading increased ($B = 0.097$ and $B = 0.046$, respectively), in either case in favor of the girls. Statistically significant variances ($p < .05$; not shown in Table 1) indicated unexplained between-classroom variation in the initial status and change of reading achievement level, enjoyment of reading level, and the gender–enjoyment relationship. Significant variation was

also found in the change of the gender–achievement relationship during 9th grade, but not in the gender–achievement relationship at the beginning of 9th grade.

Predicting the Gender Differences in Reading Achievement and Enjoyment of Reading

Next, the instructional features and the classroom-level control variables were specified as predictors of the initial status and change variables. In each of the Models M1 through M7, one teaching quality variable was entered along with the control variables.

As shown in the upper part of Table 1, pacing, focus on language competencies, and highest school type were related positively, and lower school type was related negatively to the classrooms' reading achievement level at the beginning of 9th grade. Classroom management, clarity of instruction, pacing, and focus on language competencies were related positively to the change in reading achievement during 9th grade. Finally, the teacher being female was associated with a somewhat better development of reading competencies. All teaching quality variables except classroom management were positively associated with the classrooms' initial enjoyment of reading level. Attending the highest school type was positively, attending the lower school type was negatively related to the enjoyment of reading levels at the beginning of 9th grade. No predictor was significantly associated with change of the enjoyment of reading levels during 9th grade.

We expected the scales representing the quality dimensions structure and cognitive activation to be associated with a smaller change of the reading achievement gap in favor of the girls. As shown in the lower part of Table 1, teachers' classroom management ($B = -0.034$), pacing ($B = -0.044$), and focus on language competencies ($B = -0.046$) negatively predicted change in the gender–achievement relationship during 9th grade, consistent with our hypothesis. None of the predictors was related to the gender–achievement relationship at the beginning of 9th grade. Furthermore, we expected the variables representing the quality dimension support to be associated with a smaller change of the enjoyment of reading

differences in favor of the girls. Contrary to our hypothesis, neither teacher support nor student orientation predicted change in the gender–enjoyment relationship. However, teacher support ($B = -0.043$) and focus on language competencies ($B = -0.059$) were associated with a smaller gender gap in enjoyment of reading at the beginning of 9th grade.

Finally, the three instructional features predicting change in the gender–achievement relationship during 9th grade—classroom management, pacing, and focus on language competencies—were entered jointly as predictors into a model, again controlling for proportion of girls, teacher gender, and school type. In our data, the instructional features were noticeably intercorrelated (cf. Appendix A2), even across the three quality dimensions. Although the constructs were still clearly separable, the unique contribution of each variable to predicting the gender–achievement relationship became relatively small. Consequently, no predictor was statistically significant in this model (not shown in Table 1).

Discussion

Summary of Findings

Our study investigated the extent to which change in the relationships between students' gender and their reading achievement and enjoyment of reading during 9th grade is associated with student-perceived quality of teaching. Seven instructional features related to teaching quality—representing the dimensions structure, cognitive activation, and support—were used to predict initial status and change in the gender–achievement and gender–enjoyment relationships.

In line with our expectations, two variables representing the dimension structure (classroom management, pacing) and one variable representing the dimension cognitive activation (teachers' focus on students' language competencies) were related to a less pronounced increase of girls' advantage in reading achievement during 9th grade. Together with the control variables, classroom management, pacing, and focus on language

competencies explained 8.2%, 10.1%, and 11.2%, respectively, of between-classroom variation in the change of the gender–achievement relationship. The resulting differences in the gender–achievement relationship are moderate but not irrelevant. For instance, in classrooms with inadequate pacing (say, 2 *SD* below mean), the predicted change in the gender–achievement relationship was 0.24 *SD* in favor of the girls. In classrooms with adequate pacing (say, 2 *SD* above mean), it was only 0.06 *SD* in favor of the girls.

Importantly, a reduced increase of the gender gap did not come at cost of the girls. Several of the variables above—classroom management, pacing, focus on language competencies, and additionally, clarity of instruction—were related positively to change in the reading achievement level of the entire classroom, that is, all students profited from higher teaching quality.

Contrary to our expectations, the two variables representing the quality dimension support (teacher support, student orientation) were not associated with change of the gender–enjoyment relationship during 9th grade. Instead, a high level of teacher support as well as a strong focus on students' language competencies were related to a smaller gender gap in the enjoyment of reading at the beginning of the school year. Together with the control variables, these predictors explained 1.8% and 3.7%, respectively, of between-classroom variation in the gender–enjoyment relationship at the beginning of 9th grade. For example, in classrooms with a weak focus on language competencies (2 *SD* below mean), the predicted gender gap in the enjoyment of reading was 0.65 *SD* in favor of the girls. In classrooms with a strong focus on language competencies (2 *SD* above mean), it was 0.42 *SD* in favor of the girls. Given that the majority of classrooms had the same teacher for one or more years before 9th grade, and presupposing that these teachers' quality of teaching was rather stable over the secondary years, these associations might represent some accumulated effect of differences in teaching

quality before 9th grade. Of course, this interpretation must be viewed with utmost caution and would have to be corroborated by future studies starting earlier in secondary school.

Implications for Theory and Practice

Our results fit well to previous findings on how the gender gap in reading develops across secondary school. Longitudinal research indicates that gender differences in reading increase during middle school and then decrease again until Grade 12 (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Robinson & Lubienski, 2011). In our sample, there was no statistically significant gender gap in reading achievement at the beginning of 9th grade. At the end of Grade 9, girls reported not only a higher enjoyment of reading but also outperformed boys. In Germany, 9th-graders are around 15 years old. This appears to be an age in which increased socialization pressures lead to an intensification of gender-related expectations (Hill & Lynch, 1983). Thus, an increase in gender-specific subject motivation and achievement may be one sign of this process of “gender intensification”.

However, even if the gender gaps widen during 9th grade, the size of this increase varies between classrooms, and in some of them gender differences even narrow. This shows that, while part of the gender gap may be explained with biological factors, early reading experiences in the family, and societal norms about gender roles, there is some scope for schools to promote gender equity. The associations between several instructional features and the gender–achievement and gender–enjoyment relationships indicate that increasing teaching quality might mitigate to some extent the increasing gender gap in reading commonly observed in secondary school. One might argue that these associations are small. However, students experience high-quality and low-quality instruction across their entire school career. Given that the effects of teaching quality are potentially cumulative, the gender gap in reading might considerably increase or decrease if students are consistently exposed to low or high teaching quality over several school years.

The “boys’ debate” has stressed the importance of tailoring curricula and instructional materials to the interests of boys and providing them with male role models for reading (e.g., Millard, 1997; Newkirk, 2000). However, the research evidence is rather inconclusive. Our study suggests an alternative (or additional) strategy: Improving teaching quality has the advantage of reducing the disadvantage of boys, while at the same time being beneficial for girls. Additionally, there is evidence that it might contribute to attenuating achievement gaps more generally (Geier et al., 2008; Möller et al., 2002; Snow & Lohman, 1984; White & Frederiksen, 1998). Hence, helping teachers to realize a high level of structure, cognitive activation, and support in their lessons appears to be a promising strategy for increasing overall proficiency and equity at the same time.

Limitations

One substantial limitation of this study derives from our use of student ratings of instruction. Although researchers have pointed out that student ratings may provide valuable information about students’ learning environment (e.g., Aleamoni, 1999; De Jong & Westerhof, 2001; Marsh et al., 2011), there is general agreement that classroom observations or analyses of classroom artifacts (e.g., task analysis) provide more accurate measures of several important aspects of teaching quality. The usefulness of student ratings strongly depends on the construct in question and the specific information the researcher is interested in (Kunter & Baumert, 2006). They typically work best if the instructional feature is relatively easily observable, for instance, in the judgment of teachers’ classroom management. On the other hand, for instructional features such as cognitive activation, obtaining additional measures based on different methodological approaches seems crucial for future research on this topic (Benton & Cashin, 2012; Kunter & Baumert, 2006).

Furthermore, in case of some scales, the number of available items was small. This is due to the limited number of items in the original dataset we used for this study, rather than

resulting from an extensive elimination of items. As is often the case in large-scale assessments, the number of constructs assessed in the questionnaires was high, limiting the number of items available for each construct. For similar reasons, only one instructional feature representing cognitive activation, teachers' focus on language competencies, was included. Future studies might collect additional indicators of cognitive activation. For example, enhanced cognitive activity of students might be expected if the reading materials used in the lessons match both students' prior knowledge and their reading interests.

Regarding the outcome variables, one limitation arises from our use of a short PISA scale to measure enjoyment of reading. Two of the items assess reading enjoyment without specific context ("Because reading is fun, I wouldn't want to give it up."; "When I read, I sometimes get totally absorbed."). One item explicitly refers to reading enjoyment experienced out of school ("I read in my spare time."). Given our research interest was to identify relationships to instructional features, using measures that capture reading enjoyment as experienced during the lessons might have been more appropriate and would have possibly led to results more in line with expectations for the instructional features' effects.

Finally, the generalizability of our findings is limited. First, the sample included only 9th-graders, but gender differences in reading occur already at the start of kindergarten (Robinson & Lubienski, 2011). Thus, more comprehensive information on the role of schools for aggravating or alleviating gender differences in reading could be gained by following children throughout longer periods of their school career. Second, cultural differences might affect the results. Our study examined a large and representative sample, but it included German students only. It remains unclear how far the results can be generalized to other educational systems.

References

- Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*, 153-166.
- Alloway, N., Freebody, P., Gilbert, P., & Muspratt, S. (2002). *Boys, literacy and schooling: Expanding the repertoires of practice*. Canberra, Australia: Commonwealth Department of Education, Science & Training. Retrieved from <http://eprints.jcu.edu.au/637/1/BoysLiteracySchooling.pdf>
- Asparouhov, T., & Muthén, B. O. (2010). *Multiple imputation with Mplus*. Retrieved from <http://www.statmodel.com/download/Imputations7.pdf>
- Bank, B. J., Biddle, B. J., & Good, T. L. (1980). Sex roles, classroom instruction, and reading achievement. *Journal of Educational Psychology, 72*, 119-132.
- 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*, 133-180.
- Benton, S. L., & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature* (IDEA Paper No. 50). Manhattan, KS: IDEA Center.
- Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*, 87-100.
- Brown, A. L. (1994). The advancement of learning. *Educational Researcher, 23*(8), 4-12.
- Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology, 34*, 319-337.
- Burbules, N. C., & Linn, M. C. (1988). Response to contradiction: Scientific reasoning during adolescence. *Journal of Educational Psychology, 80*, 67-75.

- Cadima, J., Leal, T., & Burchinal, M. (2010). The quality of teacher–student interactions: Associations with first graders’ academic and behavioral outcomes. *Journal of School Psychology, 48*, 457-482.
- Chambers, S. K., & Andre, T. (1997). Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current. *Journal of Research in Science Teaching, 34*, 107-123.
- Coles, M., & Hall, C. (2002). Gendered readings: Learning from children’s reading choices. *Journal of Research in Reading, 25*, 96-108.
- Cresswell, J., Rowe, K., & Withers, G. (2002). *Boys in school and society*. Camberwell, Australia: Australian Council for Educational Research. Retrieved from http://research.acer.edu.au/cgi/viewcontent.cgi?article=1000&context=boys_edu
- Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., ... Barbarin, O. (2009a). The relations of observed pre-k classroom quality profiles to children’s achievement and social competence. *Early Education & Development, 20*, 346-372.
- Curby, T. W., Rimm-Kaufman, S. E., & Ponitz, C. C. (2009b). Teacher–child interactions and children’s achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology, 101*, 912-925.
- Dar, Y., & Resh, N. (1997). Separating and mixing students for learning: Concepts and research. In R. Ben-Ari & Y. Rich (Eds.), *Enhancing education in heterogeneous schools* (pp. 191-213). Ramat-Gan, Israel: Bar-Ilan University Press.
- De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*, 51-85.
- Deci, E. L., Ryan, R. M., & Williams, G. C. (1996). Need satisfaction and the self-regulation of learning. *Learning and Individual Differences, 8*, 165-183.

- 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.
- DESI-Konsortium (Ed.). (2008). Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie [Learning and instruction of German and English. Results of the DESI study]. Weinheim, Germany: Beltz.
- Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98*, 382-393.
- Eccles, J. S., Jacobs, J. E., & Harold, R. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, *46*, 183-201.
- Eccles, J. S., Wigfield, A., Harold, R., & Blumenfeld, P. B. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*, 830-847.
- Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*, 103-112.
- Fan, W. (2011). Social influences, school motivation and gender differences: An application of the expectancy-value theory. *Educational Psychology*, *31*, 157-175.
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*, 148-162.
- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, *21*, 1-56.

- Gates, A. I. (1961). Sex differences in reading ability. *Elementary School Journal*, 61, 431-434. Retrieved from <http://www.jstor.org/stable/999823>
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45, 922-939.
- Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence*, 13, 21-43.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206-213.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. 3, pp. 403-422). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72, 625-638.
- Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn & A. Petersen (Eds.), *Girls at puberty: Biological and psychosocial perspectives* (pp. 201-228). New York, NY: Plenum.
- Hochweber, J., Steinert, B., & Klieme, E. (2012). Lehrerkooperation, Unterrichtsqualität und Lernergebnisse im Fach Englisch [The impact of teacher cooperation and instructional quality on learning in English as a foreign language]. *Unterrichtswissenschaft*, 40, 351-370.

Howes, C., Burchinal, M., Pianta, R. C., Bryant, D., Early, D., Clifford, R., & Barbarin, O.

(2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly, 23*, 27-50.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

International Labour Office (1990). *ISCO-88: International Standard Classification of Occupations*. Geneva, Switzerland: ILO.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509-527.

Johnson, W., McGue, M., & Iacono, W. G. (2005). Disruptive behavior and school grades: Genetic and environmental relations in 11-year-olds. *Journal of Educational Psychology, 97*, 391-405.

Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts "Pythagoras" [Quality dimensions and effectiveness of mathematics instruction. Theoretical background and selected findings of the Pythagoras project]. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule* (pp. 127-146). Münster, Germany: Waxmann.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137-160). Münster, Germany: Waxmann.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: Aufgabenkultur und Unterrichtsgestaltung [Mathematics instruction at lower secondary level. Task culture and instructional design]. In Bundesministerium für

- Bildung und Forschung (BMBF) (Ed.), *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente* (pp. 43-57). Berlin, Germany: BMBF.
- Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*, 231-251.
- 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*, 805-820.
- LePore, P. C., & Warren, J. R. (1997). A comparison of single-sex and coeducational catholic secondary schooling: Evidence from the National Educational Longitudinal Study of 1988. *American Educational Research Journal, 34*, 485-511.
- Lietz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation, 32*, 317-344.
- Lipowsky, F., Rakoczy, K., Pauli, C., 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.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom and school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*, 120-131.
- Marinak, B., & Gambrell, L. (2010). Reading motivation: Exploring the elementary gender gap. *Literacy Research and Instruction, 49*, 129-141.
- Marsh, H. W., Ginns, P., Morin, A. J. S., Nagengast, B., & Martin, A. J. (2011). Use of student ratings to benchmark universities: Multilevel modeling of responses to the

- Australian Course Experience Questionnaire (CEQ). *Journal of Educational Psychology, 103*, 733-748.
- Marsh, H. W., Martin, A. J., & Cheng, J. H. S. (2008). A multilevel perspective on gender in classroom motivation and climate: Potential benefits of male teachers for boys? *Journal of Educational Psychology, 100*, 78-95.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., ... Early, D. M. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732-749.
- Masters, G. N., & Forster, M. (1997). *Mapping literacy achievement. Results of the 1996 National School Literacy Survey*. Canberra, Australia: Department of Employment, Education, Training and Youth Affairs.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14-19.
- McGeown, S., Goodwin, H., Henderson, N., & Wright, P. (2012). Gender differences in reading motivation: Does sex or gender identity provide a better account? *Journal of Research in Reading, 35*, 328-336.
- McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: Results of a U.S. survey. *Reading Research Quarterly, 47*, 283-306.
- McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30*, 934-956.
- Millard, M. (1997). Differently literate: Gender identity and the construction of the developing reader. *Gender and Education, 9*, 31-48.

- Miller, J. W., & McKenna, M. C. (2016). *World literacy: How countries rank and why it matters*. New York, NY: Routledge.
- Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung [Fostering the science understanding of primary school students by structuring the learning environment]. *Zeitschrift für Pädagogik, Beiheft 45*, 176-191.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- National Institute of Child Health and Human Development Early Child Care Research Network (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *Elementary School Journal*, 102, 367-387.
- Newkirk, T. (2000). Misreading masculinity: Speculations on the great gender gap in writing. *Language Arts*, 77, 294-300.
- Organisation for Economic Cooperation and Development (2009). *PISA 2006 technical report*. Paris, France: OECD.
- Organisation for Economic Cooperation and Development (2014). *PISA 2012 results: What students know and can do – student performance in mathematics, reading and science* (Volume I, revised edition, February 2014). Paris, France: OECD.
- Pauley, F. R. (1951). Sex differences and legal school entrance age. *The Journal of Educational Research*, 45, 1-9. Retrieved from <http://www.jstor.org/stable/27529330>

- Pauli, C., Drollinger-Vetter, B., Hugener, I., & Lipowsky F. (2008). Kognitive Aktivierung im Mathematikunterricht [Cognitive activation and mathematics instruction]. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology*, 22, 127-133.
- Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, 45, 365-397.
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38, 109-119.
- Pianta, R. C., La Paro, K., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore: Paul H. Brookes.
- Rakoczy, K. (2006). Motivationsunterstützung im Mathematikunterricht [Motivational support in mathematics instruction]. *Zeitschrift für Pädagogik*, 52, 822-843.
- Räty, H., Kasanen, K., & Kärkkäinen, R. (2006). School subjects as social categorizations. *Social Psychology of Education*, 9, 5-25.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Reeve, J. (2002). Self-determination theory applied to educational settings. In E. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 183-203). Rochester, NY: The University of Rochester Press.
- 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.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009).

The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*, 958-972.

Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E., & Stanat P. (2014).

Socioeconomic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction, 32*, 63-72.

Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., & Stanat, P. (2015). Classroom

composition and language minority students' motivation in language lessons. *Journal of Educational Psychology, 107*, 1171-1185.

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in

mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal, 48*, 268-302.

Rosenshine, B., & Furst, N. (1971). Research on teacher performance criteria. In B. O. Smith

(Ed.), *Research in teacher education*. Englewood Cliffs, NJ: Prentice-Hall.

Sainsbury, S., & Schagen, I. (2004). Attitudes to reading at ages nine and eleven. *Journal of*

Research in Reading, 27, 373-386.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art.

Psychological Methods, 7, 147-177.

Silinskas, G., Parrila, R., Lerkkanen, M.-K., Poikkeus, A.-M., Niemi, P., & Nurmi, J.-E.

(2010). Mothers' reading-related activities at home and learning to read during kindergarten. *European Journal of Psychology of Education, 25*, 243-264.

- Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76*, 347-376.
- Stanat, P., & Kunter, M. (2002). Geschlechterspezifische Leistungsunterschiede bei Fünfzehnjährigen im internationalen Vergleich [An international comparison of gender-specific differences in achievement of 15-year-olds]. *Zeitschrift für Erziehungswissenschaft, 5*, 28-48.
- Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist, 39*, 97-110.
- Steinert, B., Hartig, J., & Klieme, E. (2008). Institutionelle Bedingungen der Sprachkompetenzen [Institutional prerequisites of language competencies]. In DESI-Konsortium (Ed.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (pp. 411-450). Weinheim, Germany: Beltz.
- Steyer, R., Partchev, I., & Shanahan, M. J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little, J. Baumert, & K.-U. Schnabel (Eds.), *Modeling longitudinal and multilevel data* (pp. 109-126). Mahwah, NJ: Erlbaum.
- Swalander, L., & Taube, K. (2007). Influences of family-based prerequisites, reading attitude, and self-regulation on reading ability. *Contemporary Educational Psychology, 32*, 206-230.
- Thorndike, R. L., & Hagen, E. (1993). *The Cognitive Abilities Test: Form 5*. Itasca, IL: Riverside.
- Van Buuren, S. (2010). Multiple imputation of multilevel data. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 173-196). New York, NY: Routledge.

- Van de gaer, E., Pustjens, H., Van Damme, J., & De Munter, A. (2006). The gender gap in language achievement: The role of school-related attitudes of class groups. *Sex Roles, 55*, 397-408.
- Veenman, M. V. J., & Elshout, J. J. (1995). Differential effects of instructional support on learning in simulation environments. *Instructional Science, 22*, 363-383.
- Von Secker, C. E., & Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching, 36*, 1110-1126.
- Weinert, F. E., & Helmke, A. (1995). Learning from wise Mother Nature or Big Brother Instructor: The wrong choice as seen from an educational perspective. *Educational Psychologist, 30*, 135-142.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*, 3-118.
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*, 51-59.

Table 1

Predicting Initial Status (T1) and Change (T2–T1) of the Gender Relationship With Reading Achievement and Enjoyment of Reading (Classroom-Level Results)

Variables	M0 (B)		M1 (B)		M2 (B)		M3 (B)		M4 (B)		M5 (B)		M6 (B)		M7 (B)	
	T1	T2–T1	T1	T2–T1	T1	T2–T1	T1	T2–T1	T1	T2–T1	T1	T2–T1	T1	T2–T1	T1	T2–T1
Reading achievement level																
Intercept	0.131***	-0.040*	0.160***	-0.098**	0.158***	-0.095**	0.160***	-0.095**	0.161***	-0.095**	0.160***	-0.096**	0.160***	-0.097**	0.160***	-0.096**
CM			0.005	0.045**												
Structuredness					-0.012	0.018										
Clarity							0.020	0.052**								
Pacing									0.063**	0.093***						
Focus on LC											0.044*	0.077**				
Teacher support													0.012	0.033		
Student orientation															0.022	0.013
Proportion of girls			0.022	0.007	0.021	0.009	0.021	0.004	0.013	-0.005	0.014	-0.008	0.022	0.006	0.020	0.007
Teacher gender			-0.044	0.095*	-0.042	0.090*	-0.045	0.091*	-0.046	0.092*	-0.046	0.091*	-0.044	0.093*	-0.045	0.092*
Intermediate ST			-0.077**	-0.008	-0.080**	-0.011	-0.074**	-0.003	-0.071**	-0.005	-0.074**	-0.006	-0.076**	-0.009	-0.072**	-0.010
Highest ST			0.438***	0.016	0.435***	0.021	0.435***	0.010	0.403***	-0.035	0.421***	-0.012	0.439***	0.020	0.435***	0.015
R ²	–	–	75.9	7.3	75.9	3.3	76.0	8.2	77.3	14.9	76.6	13.1	75.9	5.4	76.0	3.5
Reading enjoyment level																
Intercept	-0.214***	-0.162***	-0.190***	-0.173***	-0.186***	-0.173***	-0.188***	-0.173***	-0.188***	-0.173***	-0.189***	-0.172***	-0.189***	-0.173***	-0.189***	-0.173***
CM			0.033	0.003												
Structuredness					0.037*	-0.001										
Clarity							0.044*	-0.012								
Pacing									0.045*	-0.017						
Focus on LC											0.070**	-0.012				
Teacher support													0.041*	-0.010		
Student orientation															0.043*	-0.007
Proportion of girls			0.006	<0.001	0.009	<0.001	0.004	0.001	0.001	0.002	-0.008	0.002	0.004	0.001	0.002	0.001
Teacher gender			-0.035	0.020	-0.042	0.020	-0.038	0.020	-0.037	0.020	-0.038	0.020	-0.036	0.020	-0.037	0.020
Intermediate ST			-0.094***	-0.030	-0.091***	-0.031	-0.089***	-0.033	-0.094***	-0.032	-0.091***	-0.032	-0.092***	-0.032	-0.087***	-0.033
Highest ST			0.200***	0.009	0.210***	0.009	0.195***	0.011	0.176***	0.019	0.174***	0.014	0.205***	0.008	0.195***	0.010
R ²	–	–	41.0	2.7	41.1	3.0	42.0	3.9	41.4	4.0	45.6	4.1	41.6	3.5	41.6	3.2

Table 1 (continued)

Variables	M0 (B)		M1 (B)		M2 (B)		M3 (B)		M4 (B)		M5 (B)		M6 (B)		M7 (B)	
	T1	T2-T1	T1	T2-T1	T1	T2-T1	T1	T2-T1	T1	T2-T1	T1	T2-T1	T1	T2-T1	T1	T2-T1
Gender-achievement relationship																
Intercept	0.024	0.097***	0.022	0.149***	0.022	0.146***	0.022	0.148***	0.023	0.148***	0.022	0.149***	0.022	0.149***	0.022	0.148***
CM			0.008	-0.034 [†]												
Structuredness					-0.005	-0.017										
Clarity							-0.002	-0.026								
Pacing									0.018	-0.044 [†]						
Focus on LC											0.004	-0.046 [†]				
Teacher support													-0.008	-0.016		
Student orientation															0.004	-0.003
Proportion of girls			0.014	-0.006	0.014	-0.008	0.014	-0.005	0.010	-0.002	0.014	0.004	0.015	-0.005	0.014	-0.007
Teacher gender			-0.009	-0.090	-0.009	-0.086	-0.009	-0.088	-0.011	-0.090	-0.010	-0.088	-0.009	-0.089	-0.010	-0.089
Intermediate ST			0.002	0.026	<0.001	0.028	<0.001	0.025	0.004	0.028	0.001	0.026	<0.001	0.028	0.002	0.030
Highest ST			0.039	-0.023	0.038	-0.028	0.039	-0.020	0.029	0.002	0.037	-0.005	0.038	-0.025	0.038	-0.023
R ²	–	–	32.3	8.2	32.7	3.7	32.3	6.2	33.1	10.1	31.3	11.2	32.5	4.6	32.5	3.2
Gender-enjoyment relationship																
Intercept	0.560***	0.046*	0.532***	0.046	0.528***	0.046	0.531***	0.047	0.535***	0.048	0.532***	0.047	0.532***	0.046	0.532***	0.047
CM			-0.033	0.006												
Structuredness					-0.042	<0.001										
Clarity							-0.040	0.018								
Pacing									0.053	0.035						
Focus on LC											-0.059*	0.038				
Teacher support													-0.043*	0.012		
Student orientation															-0.027	0.012
Proportion of girls			-0.017	0.001	-0.020	0.001	-0.015	<0.001	-0.028	-0.005	-0.005	-0.006	-0.015	<0.001	-0.014	-0.001
Teacher gender			<0.001	<0.001	0.008	<0.001	0.003	-0.001	-0.003	-0.001	0.003	-0.002	0.002	<0.001	0.001	-0.001
Intermediate ST			0.012	-0.002	0.009	-0.003	0.008	0.001	0.025	0.001	0.010	0.001	0.010	-0.001	0.009	0.001
Highest ST			-0.011	0.029	-0.022	0.029	-0.006	0.026	-0.039	0.009	0.012	0.015	-0.016	0.030	-0.008	0.027
R ²	–	–	0.4	4.5	1.3	4.4	1.6	5.9	3.1	8.6	3.7	8.9	1.8	5.1	– ^a	5.0

Note. CM = classroom management; LC = language competencies; ST = school type. ^a Negative R² obtained (uninterpretable).

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed). [†] $p < .05$, ^{††} $p < .01$, ^{†††} $p < .001$ (one-tailed).

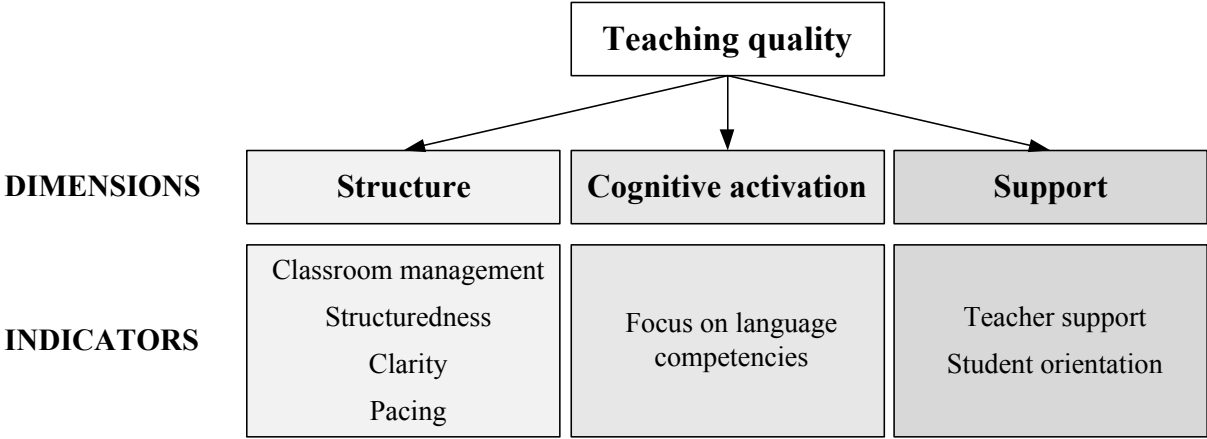


Figure 1. Basic dimensions of teaching quality and indicators (instructional features) used in this study.

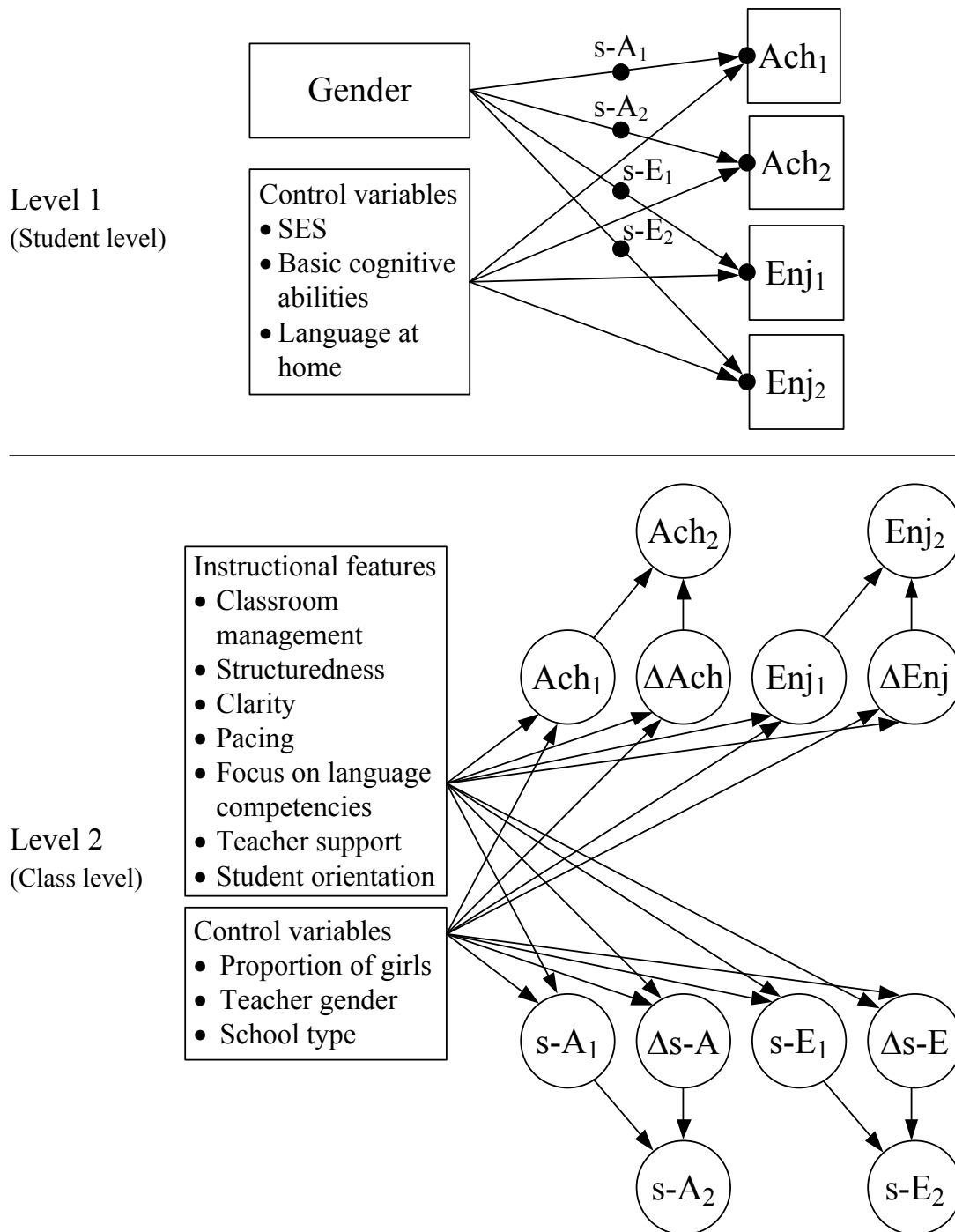


Figure 2. Illustration of the model for predicting change in the classroom-specific gender gaps in reading achievement and enjoyment of reading (Ach₁, Ach₂: achievement at T1 and T2; Enj₁, Enj₂: enjoyment at T1 and T2; s-A₁, s-A₂, s-E₁, s-E₂: random slopes for gender effects on achievement at T1, achievement at T2, enjoyment at T1, enjoyment at T2. Any classroom level change model was specified as: $Y[T2] = Y[T1] + \Delta Y$, and the residual variance of $Y[T2]$ was restricted to zero for identification purposes. At both levels, the relationships between all predictor variables and the relationships between the residuals of all dependent variables were estimated freely.)

Appendix

Table A1

Constructs and Items

Enjoyment of reading

1. Because reading is fun, I wouldn't want to give it up.
2. I read in my spare time.
3. When I read, I sometimes get totally absorbed.

Classroom management

1. My German teacher ensures the students pay attention throughout the entire lesson.
2. My German teacher always knows exactly what is happening in the classroom.

Structuredness

1. At the beginning of the lesson, my German teacher gives an overview of the topic.
2. At the end of the lesson, my German teacher gives a summary of the most important aspects.
3. My German teacher points out the relevant aspects of a lesson.

Clarity

1. The assignments in the German language classes are clear and comprehensible to me.
2. When explaining something, my German teacher uses vivid examples.
3. My German teacher expresses herself/himself clearly and comprehensibly.

Pacing

1. My German teacher proceeds too fast.
2. In German language lessons I often do not get my work done.
3. During group work in German language lessons, I often do not get my assignments done.
4. Group work in German language lessons is frequently interrupted by explanations of our teacher.

Focus on language competencies

How important is it for your German teacher that ...

1. ... you can formulate questions on a text?
2. ... you can listen to a discussion partner and respond to her/him?
3. ... you can speak grammatically correctly?
4. ... you can write grammatically correctly?
5. ... you can argue convincingly?
6. ... you master spelling and punctuation rules?
7. ... you can give a brief talk based on notes?
8. ... you know important grammatical concepts (e.g., present tense, subjunctive, adverb)?
9. ... you can summarize the most important contents of a text?
10. ... you can write texts?

Table A1 (continued)

Teacher support

1. If I need help, I get it from my German teacher.
2. My German teacher gives me advice on how to improve.
3. My German teacher makes sure that I keep up in the lessons.
4. My German teacher treats me fairly.

Student orientation

1. My German teacher allows us to choose topics or tasks during group work.
 2. My German teacher is responsive to our suggestions.
 3. My German teacher encourages us to express our opinion.
 4. If somebody has a good idea, my German teacher acknowledges it.
 5. My German teacher gives me opportunity to express my opinion.
 6. My German teacher is interested in what I have to say.
-

Table A2

Descriptive Statistics and Correlations (Lower Left Triangle), Variances (Diagonal), and Covariances (Upper Right Triangle) of Study Variables

Variable	<i>M</i>	<i>SD</i>	% Missing	1	2	3	4	5	6	7	8	9	10	11
Student level (<i>n</i> = 10543)														
1. Reading test performance T1 (WLE)	-.20	.93	5.22	.86	.49	.26	.26	.03	4.02	4.28	-.07	.00		
2. Reading test performance T2 (WLE)	-.19	.99	7.36	.53	.97	.29	.30	.05	4.59	4.80	-.07	-.01		
3. Enjoyment of reading T1	2.68	.92	11.96	.30	.32	.84	.61	.13	2.58	1.83	-.02	.00		
4. Enjoyment of reading T2	2.57	.97	14.91	.29	.32	.69	.94	.14	2.55	1.99	-.03	.00		
5. Student gender	.51	.50	0.00	.06	.09	.29	.29	.25	-.31	.18	.00	.00		
6. Socio-economic status (HISEI)	49.52	15.81	40.75	.27	.29	.18	.17	-.04	249.93	36.65	-.95	-.09		
7. Basic cognitive abilities	50.00	10.00	9.09	.46	.49	.20	.21	.04	.23	100.01	-.54	-.08		
8. Language use: Other than German	.13	.34	8.93	-.21	-.22	-.07	-.08	-.02	-.18	-.16	.11	-.01		
9. Language use: German and another	.06	.24	8.93	-.01	-.04	.00	.02	.02	-.02	-.03	-.10	.06		
Classroom level (<i>n</i> = 427)														
1. Classroom management	2.63	.44	0.00	.20	.09	.11	.01	.07	.11	.09	.00	-.01	-.01	.00
2. Structuredness	2.32	.34	0.00	.60	.11	.07	-.01	.02	.08	.07	-.01	.01	-.02	-.04
3. Clarity	2.82	.34	0.00	.72	.64	.12	.02	.05	.10	.09	.00	.00	-.02	.02
4. Pacing	2.97	.25	0.00	.06	-.10	.28	.06	.02	.01	.02	.01	.00	-.01	.05
5. Focus on language competencies	3.16	.28	0.00	.54	.25	.56	.33	.08	.05	.05	.01	.00	-.01	.04
6. Teacher support	2.79	.35	0.00	.68	.70	.84	.13	.47	.12	.10	.00	.00	-.01	-.02
7. Student orientation	2.73	.34	0.00	.63	.63	.82	.21	.49	.87	.11	.00	.00	-.03	.02
8. Proportion of girls	.52	.17	0.00	-.01	-.17	.05	.26	.22	-.02	.08	.03	.00	.00	.02
9. Teacher gender	.62	.49	15.46	-.02	.09	-.01	-.02	-.01	.01	-.01	-.02	.24	.02	-.03
10. Intermediate school type (Realschule)	.32	.47	0.00	-.06	-.11	-.13	-.06	-.05	-.09	-.17	.02	.10	.22	-.12
11. Highest school type (Gymnasium)	.38	.49	0.00	.00	-.22	.11	.44	.29	-.09	.11	.22	-.11	-.53	.24

Note. All statistics reported are based on the original variables, before applying multiple imputation. Classroom level variables no. 1 to 7 are based on student ratings of instruction. Percent missing at the student level: Classroom management 14.2%, structuredness 13.2%, clarity 13.5%, pacing 15.0%, focus on language competencies 55.4%, teacher support 13.7%, student orientation 13.6%.

Table A3

Descriptive Statistics for the School Type Categories (Hauptschule, Realschule, Gymnasium)

Variable	Hauptschule		Realschule		Gymnasium	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Student level						
Reading test performance T1 (WLE)	-.72	.83	-.19	.79	.48	.76
Reading test performance T2 (WLE)	-.76	.91	-.18	.82	.54	.78
Enjoyment of reading T1	2.47	.90	2.62	.90	3.01	.86
Enjoyment of reading T2	2.35	.94	2.49	.96	2.93	.91
Classroom level						
Classroom management	2.68	.33	2.59	.46	2.63	.50
Structuredness	2.49 ^{a,b}	.28	2.26 ^a	.35	2.22 ^b	.32
Clarity	2.83	.29	2.76 ^a	.36	2.87 ^a	.35
Pacing	2.82 ^{a,b}	.28	2.95 ^{a,c}	.19	3.11 ^{b,c}	.19
Focus on language competencies	3.05 ^{a,b}	.32	3.13 ^{a,c}	.26	3.26 ^{b,c}	.22
Teacher support	2.89 ^{a,b}	.30	2.74 ^a	.38	2.75 ^b	.35
Student orientation	2.76 ^a	.27	2.65 ^{a,b}	.37	2.78 ^b	.35

Note. At the student level, all school type differences are statistically significant ($p < 0.001$, Bonferroni correction for multiple testing).

At the classroom level, school type differences with equal indices are statistically significant ($p < 0.05$, Bonferroni correction for multiple testing).

Sample size at the student level: $n = 3904$ (Hauptschule), $n = 3756$ (Realschule), $n = 2884$ (Gymnasium).

Sample size at the classroom level: $n = 129$ (Hauptschule), $n = 135$ (Realschule), $n = 163$ (Gymnasium).