Studies in Educational Evaluation 81 (2024) 101347

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Teaching quality and student reading outcomes: Evidence from a longitudinal study from grade 5 to 7

Manja Attig [a,f,1], Frances Hoferichter [b,f,1], Isa Steinmann [c,f,1], Rolf Strietholt [d,e,f,*,1]

[a] Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany
[b] University of Greifswald, Ernst-Lohmeyer-Platz 3, 17489 Greifswald, Germany
[c] Oslo Metropolitan University, Pilestredet 42, 0167 Oslo
[d] TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany
[e] International Association for the Evaluation of Educational Achievement (IEA), Überseering 27, 22297 Hamburg, Germany
[f] College for Interdisciplinary Educational Research (CIDER), Rostocker Straße 6, 60323 Frankfurt am Main, Germany

ARTICLE INFO

Keywords:
Reading competence
Attitudes towards reading
Teaching quality
NEPS
Lower secondary school students

ABSTRACT

The literature on the effects of teaching behavior on student competence and motivation has primarily focused on single teaching facets, with limited attention to multiple components simultaneously. To address this gap, this study examined a model of teaching and learning by investigating 10 teaching quality components reported by both students and teachers. The study used a sample of 3067 students in 150 schools in Germany and regressed reading competence and attitudes towards reading in grade 7 on these 10 variables. The study controlled for prior reading competence and attitudes towards reading in grade 5, as well as other student background characteristics, and estimated school-fixed effects. The results did not identify significant associations between the investigated teaching quality components and the reading competence and attitudes towards reading measures when considered simultaneously in one model, after taking into account previous competence and attitudes. The study discusses limitations and implications of these findings.

Hardly any other topic in education research has received as much attention as teaching quality. However, a pivotal issue in the research on teacher quality is its characterization as a rather abstract meta-construct, which includes an array of facets, notably the teachers' deliberate design of learning arrangements, alongside various dimensions of their professional actions. Several studies have examined the effects of different facets and dimensions of teaching quality on student achievement, and the results of these studies have been summarized in multiple reviews and meta-analyses (e.g., Brophy, 1986; Creemers, 1994; Fraser, Walberg, Welch, & Hattie, 1987; Levine & Lezotte, 1990; Scheerens, 2000; Scheerens & Bosker, 1997; Seidel & Shavelson, 2007; Wang, Haertle, & Walberg, 1993). While there are several reviews concerning the effects of various components of teaching quality, the challenge of effectively integrating research findings persists. A fundamental obstacle in this field of research lies in the meaningful integration of individual studies, primarily due to the absence of widely accepted theoretical models and classification systems. Seidel and Shavelson (2007) concluded that the models in use significantly influence the outcomes of meta-analyses on teaching effectiveness.

An inherent challenge when assessing teaching quality lies in the dynamic and multifaceted nature of the instructional process. Isolating individual components of teaching quality for examination is hardly meaningful. For instance, if a teacher provides goal-oriented instruction but manages instructional time inadequately, the effectiveness of their teaching may be compromised. Over the past two decades, scholars have increasingly acknowledged various overlapping components of teaching quality, giving rise to a multidimensional and process-oriented approach to learning (Antoniou & Kyriakides, 2011, 2013;Baumert, Blum, & Neubrand, 2004; Fauth et al., 2019; Graham, White, Cologon, & Pianta, 2020; Klieme & Rakoczy, 2003; Seidel, Rimmele, & Prenzel, 2003; Senden, Nilsen, & Blömeke, 2021). Currently, three popular conceptual models hold influence and are widely employed in research. Bolhuis (2003) presented a process-oriented model of teaching for self-directed learning. This model encompasses cognitive, behavioral, emotional, and

* Corresponding author at: TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany.
E-mail addresses: manja.attig@lifbi.de (M. Attig), frances.hoferichter@uni-greifswald.de (F. Hoferichter), isa.steinmann@oslomet.no (I. Steinmann), rolf.strietholt@iea-hamburg.de (R. Strietholt).

[1] All authors contributed equally to this paper and are listed in alphabetic order.

https://doi.org/10.1016/j.stueduc.2024.101347
Received 20 April 2023; Received in revised form 11 January 2024; Accepted 22 February 2024
Available online 1 March 2024

contextual components, treating learning as a complex, process-oriented social phenomenon. Her model involves teaching components such as goal setting, goal orientation, execution of learning activities, evaluation, and regulation, all underscored by the pivotal role that the social context plays in the learning process. Creemers and Kyriakides (2008; see also Kyriakides et al., 2020) proposed the Dynamic Model of Educational Effectiveness, which aims to consider a broad set of educational outcomes and constructivist learning theories. The dynamic model comprises dimensions like orientation, structuring, modeling, application, questioning, assessment, management of time, and the classroom as a learning environment. Klieme, Pauli, and Reusser (2009) have proposed Three Basic Dimensions (TBD) of teaching quality, namely, cognitive activation, classroom management, and a supportive climate. These dimensions are also denoted as a "deep structure," and their classification scheme is underpinned by insights from international video research.

Each of the three modern models of teaching quality introduced previously has been applied in numerous studies and cited extensively (e.g., Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Kyriakides, Creemers, & Antoniou, 2009; Kyriakides, Christoforou, & Charalambous, 2013; Seidel & Shavelson, 2007). Interestingly, these three models have evolved independently of each other, and in the cited sources, there are hardly any references to one another. The models share several similarities, as they all view teaching quality as a multidimensional meta-construct and they consider learning as a co-constructive process in which students engage in self-directed learning. However, they differ in their emphasis on whether and to what extent teachers establish learning goals, engage students, and conduct ongoing evaluation and adaptation of instruction. This paper specifically adopts and discusses the framework suggested by Bolhuis (2003).

1. Bolhuis' extended model of teaching and learning

The present study is underpinned by the theoretical model proposed by Bolhuis (2003), which has been empirically operationalized in an extensive meta-analysis conducted by Seidel and Shavelson (2007). The authors of this meta-analysis have noted that they have extended Bolhuis's model in various key aspects to enhance its applicability and scope. Specifically, they distinguish between distal and proximal executive processes of teaching and learning. Distal executive processes cover the broader frame for learning and include the teaching quality components 1–7 below. In comparison to these distal executive processes, proximal executive processes (see component 8 below) refer to actual learning activities and are thought to have a larger impact on learning. In the following, we provide a detailed description on the distal (1−7) and proximal (8) teaching components:

1) Domain of learning refers to the knowledge domain (e.g., mathematics, language).
2) Time for learning reflects the duration of teaching.
3) Organization of learning includes classroom management activities such as providing a functional and orderly classroom setting.
4) Social context pertains to teaching activities that focus on the creation of a social learning climate (e.g., student discussions).
5) Goal setting and orientation relate to clarifying goals, providing structured and clear teaching, and activating students' knowledge on the topic (i.e., cognitive activation).
6) Evaluation entails assessing students' learning progress and providing feedback.
7) Regulation and monitoring encompasses supervising students during learning processes, considering students' individual differences (i.e., adaptive teaching), and teaching students to apply self-directed learning and self-monitoring strategies.
8) Execution of learning includes three subcategories.
   a) Initiating social interactions means that the learning situations entail engaging in social interactions.
   b) Basic processing implies that the learning situations help processing basic learning and information.
   c) Domain-specific processing entails that the learning situations also focus on specific learning processes.

By providing this set of teaching quality components, the model takes a holistic view and brings together various research insights on effective learning and teaching. Drawing their conclusions mainly on correlational studies, Seidel and Shavelson (2007) investigated the associations between the teaching quality components and student cognitive and motivational-affective outcomes. Cognitive outcomes referred to student performance measured in standardized tests. Motivational-affective outcomes included attitudes towards school and learning, motivational orientations, and the development of stable interests.

1.1. Empirical findings on cognitive and motivational outcomes

The most crucial empirical evidence for our study is the meta-analysis by Seidel and Shavelson, as it explicitly references the theoretical model developed by Bolhuis. Therefore, we will first summarize their findings. Subsequently, we will consolidate the results of previous meta-analyses and reference recent research.

The aforementioned meta-analysis (Seidel & Shavelson, 2007) found Cohen's *d* effect sizes ranging from.00 to.05 for the correlation between teaching quality components and cognitive student outcomes, and a range from 0.00 to.13 for motivational outcomes. The sole exceptions were domain-specific processing strategies, which demonstrated a more substantial effect size of.22 for cognitive and.21 for motivational outcomes. It seems worth to note that domain-specific processing strategies typically focus on narrowly defined subject content areas and, accordingly, can be generalized only to a limited extent. In contrast, the other generic components of teaching quality are not specific to individual content areas.

In a more recent meta-analysis, Kyriakides *et al.* (2013) tested the Dynamic Model he co-developed. This review identified comparatively high effects for all components of teaching quality, approximately.35. Similarly high effects were also found for additional components of teaching quality that were not situated within the Dynamic Model. When examining earlier meta-analyses, it is imperative to underscore that their classification systems do not align with the modern models of learning that have been previously referenced. Notably, the effect sizes reported in these earlier analyses are typically much larger. The two moth comprehensive previous meta-analyses by Fraser et al. (1987) and Scheerens and Bosker (1997) document standardized effect sizes for variables like reinforcement, feedback, cooperative learning, mastery learning, learning time, and adaptive instruction, ranging from.3 to 1.2. Nonetheless, these findings have not been replicated in more recent studies. Applying Scheerens and Bosker's (1997) categorization system to contemporary research, Seidel and Shavelson (2007) found predominantly very small effects. Scheerens, Luyten, Steen, & de Thouars, (2007, p. 200) stated that the effect sizes of teaching quality on student learning "appeared to be disappointingly small."

Recent international comparative studies found inconsistent results. Using international data from the Trends in International Mathematics and Science Study (TIMSS) and a range of instructional quality items from a teacher questionnaire (including clarity of instruction, cognitive activation, supportive climate), Blömeke, Olsen, and Suhl (2016) reported that student mathematics achievement was not well predicted by a composite measure of the teaching quality that combined measures of cognitive activation, clarity of instruction, and a supportive climate in 47 countries. Nortvedt, Gustafsson, and Lehre (2016) used the student questionnaire from TIMSS and PIRLS (Progress in International Reading Literacy Study) to measure instructional quality and found mixed

results. While teaching quality showed a positive correlation with student achievement in some countries, it was negatively correlated in others, and in some, no significant relationship was observed. A limitation of both studies utilizing TIMSS and PIRLS data is the cross-sectional design, and the validity of the measures of teaching quality is limited, as each component of teaching quality was measured with only two to three items (Klieme & Nilsen, 2022).

Using the most recent TIMSS 2019 data from Norway, Senden, Nilsen, and Teig (2023) employed refined measures to assess the teaching quality components of cognitive activation, classroom management, and supportive climate. The interconnectedness of the teaching quality components became evident as the correlations apparent in simple regression models vanished when all three were analyzed together, highlighting the need for a holistic approach in understanding and modeling teaching quality to mitigate potential confounding effects. In their final model, none of to components of teaching quality correlated with math and science achievements in the fifth grade when the socio-economic backgrounds of the classes were controlled for. However, in the ninth grade, significant correlations were found: cognitive activation and classroom management with math achievement, and cognitive activation with science achievement.

The TALIS (Teaching and Learning International Survey) video study employed a longitudinal design. Conducted across nine countries, it revealed that classroom management, socio-emotional support, and the overall quality of instruction had positive correlations with student performance in some countries. However, these effects diminished and were no longer statistically significant when student performance was controlled for longitudinally (Doan, Mihaly, & McCaffrey, 2020; Praetorius, Herbert, Decristan, & Köhler, 2020). Additionally, more favorable effects were seen for student interest and self-efficacy, with modest impacts from classroom management, socio-emotional support, and overall instructional quality observed in approximately one third of the countries studied.

The aforementioned meta-analyses and international studies have uncovered evidence of small effects of teaching quality on performance outcomes and slightly more pronounced effects on motivational outcomes. All in all, however, Scheerens et al. (2007, p. 200) concludes that the effect sizes of teaching quality on student learning "appeared to be disappointingly small." Nevertheless, a growing number of smaller, national studies have been investigating the impacts of teaching quality in recent years. This paper does not aim to conduct a systematic review of this growing body of studies published subsequent to the latest meta-analyses. Nonetheless, recent research appears to corroborate the varied outcomes of earlier studies, indicating effects that range from negligible to modestly positive. For instance, utilizing longitudinal expansions of representative samples from PISA (Programme for International Student Assessment) and TEDS-M (Teacher Education and Development Study in Mathematics) datasets, Baumert et al. (2010) and Blömeke, Jentsch, Ross, Kaiser, and König (2022) studied the effects of teaching quality on growth in math achievement. While Baumert et al. found evidence for moderate effects of classroom management but not for individual leaning support, Blömeke et al. did not find evidence that a general measure of teaching quality – combining measures classroom management, structure, student support, and cognitive activation–predicts growth in a achievement. Likewise, a number of international studies utilizing non-representative samples have reported varied outcomes for both academic performance (e.g., Burgess, Rawal, & Taylor, 2023; Idris, 2023; Praetorius, et al., 2017; Lauermann, & ten Hagen, 2021) and motivational-affective outcomes, including positive school attitudes (Hoferichter, Hirvonen, & Kiuru, 2021), higher school engagement (Martin & Collie, 2018), higher achievement motivation (Hoferichter, Kulakow, & Raufelder, 2020; Scherer & Nilsen, 2016; Wentzel, 2009), and grit, that is, perseverance and interest (Kulakow & Hoferichter, 2021).

1.2. Reading competence and attitudes towards reading

While the previous results referred to domain-general findings, the present study focuses on the reading domain. *Reading competence* is defined as a comprehensive cognitive ability that includes a "competent handling of written texts in different and typical everyday situations" (Gehrer, Zimmermann, Artelt, & Weinert, 2012, p. 2). *Attitudes towards reading* include feelings related to reading that cause students to approach or avoid reading situations (Tunnell, Calder, Justen, & Phaup, 1991). Hence, reading attitude is an affective concept that strongly overlaps with reading motivation (McKenna, Kear, & Ellsworth, 1995). In their meta-analysis, Seidel and Shavelson (2007) emphasized the importance of domain-specific activities; for reading they found an overall effect size of $d = .15$.

Various studies, with very diverse results, have investigated different reading methods and evaluated interventions with the aim of enhancing reading competence and motivational-affective outcomes, as reading is an important asset for societal participation and career development (Muis, Ranellucci, Trevors, & Duffy, 2015; Solheim, Frijters, Lundetrae, & Uppstad, 2018; Spoerer & Schuenemann, 2014; Vansteelandt, Mol, Vanderlinde, Lerkkanen, & Van Keer, 2020). While there are many studies on teacher instruction in relation to reading interventions, there are few studies on other teaching quality components. Hochweber and Vieluf (2016) found that high teaching quality, defined as effective classroom management, adequate pacing, and a strong focus on language competencies, was related to higher reading achievement and reading enjoyment over one school year. High teaching quality also mitigated the commonly observed increase of gender gaps in reading motivation and achievement. In their cross-sectional analysis of the perceived motivational role of German language teachers for girls' and boys' learning strategies, Schweder and Raufelder (2020) found that, among girls, perseverance, elaboration, memorizing, and control strategies were higher if they perceived their language teachers as a motivational source, whereas for boys, only perseverance was related to teachers' motivational role.

1.3. The current study

In sum, Seidel and Shavelson's (2007) meta-analysis took a multidimensional and process-oriented approach to teaching and learning and investigated the effect of teaching quality components on students' cognitive and motivational-affective outcomes. The meta-analysis included mainly correlational and few experimental studies. The results varied depending on the design of the primary studies. The largest effects were observed for proximal teaching components (see components 1–7 above) that were mainly investigated in (quasi-)experimental studies. Furthermore, the conceptual inconsistency of the individual studies, combined with the failure to view teaching and learning as a multidimensional process, made it difficult to interpret the results. Hence, specific effects may have been masked because effect sizes were aggregated over different teaching components and student outcomes. Overall, the meta-analysis (Seidel & Shavelson, 2007) found small effects of teaching quality components on cognitive outcomes and small to moderate effects on motivational-affective outcomes. The highest effect sizes were found for the execution of domain-specific activities.

To our knowledge, longitudinal, domain-specific survey studies that investigated the impact of a comprehensive set of teaching quality components on both cognitive and motivational-affective outcomes have not been conducted so far in the domain of reading. Using a longitudinal design and considering school-fixed effects as well as students' prior competence and background variables, this study therefore promises to move the research field forward. We follow Seidel and Shavelson's (2007) model and, simultaneously and step-wise, investigate the effects of a set of teaching quality components on students' reading competence and attitudes towards reading in German lower secondary schools.

Based on the presented literature, we tested two hypotheses:

1) We expected small positive effects of a set of 10 teaching quality components on reading competence from grades 5 to 7.
2) We expected small to moderate positive effects of a set of 10 teaching quality components on attitudes towards reading from grades 5 to 7.

To consider potential confounding effects, we included a set of control variables in addition to prior reading competence and attitudes towards reading. These included demographic, socio-economic, and school-related variables. Previous studies usually find higher reading competences and more positive attitudes towards reading in girls than in boys (e.g., OECD, 2019). Since students become more reading proficient over the school years and their cognition evolves further, we also included age. Students' competences and attitudes also differ by the educational levels of parents linked with the family's economic situation and immigration background (e.g., Bergen, Zuijen, Bishop, & Jong, 2016; Flisi, Meroni, & Vera-Toscano, 2016). Further, students' grade in the subject German as well as how satisfied they feel at school may correlate with their reading skills and attitudes towards reading (e.g., Marshik, Ashton, & Algina, 2017; Zhang et al., 2020).

2. Materials and Methods

2.1. Data

This study used data from a starting cohort of fifth grade students of the National Educational Panel Study in Germany (NEPS; Blossfeld, Roßbach, & von Maurice, 2011; Blossfeld & Roßbach, 2019), who were followed from the school year 2010/2011 until grade 7 in 2012/2013.[2] We investigated students from the main regular school survey who remained at the same schools between grades 5 and 7 to study teaching effects on student outcomes longitudinally.[3] We therefore excluded students who changed schools between grades 5 and 7 or repeated a grade. Our sample covered 15 out of 16 federal states because in all Berlin schools that participated in NEPS, the transition between primary and secondary school was after the sixth grade. In the other states, the transition between primary and secondary school occurs after the fourth grade, so that students are typically at the same school from the fifth to the seventh grade (and beyond). We also excluded students with missing data on the attended German class. Students who attended German classes for which no teacher information on the teaching quality was available at all in grades 5, 6, and 7 were also excluded. The effective sample amounted to $N = 3067$ students nested in 150 schools. The students were on average almost 11 years old in grade 5, about half of them were girls, and two thirds did not have a migration background (see Table 1). The schools included 25 lower-track schools (*Hauptschule*), 31 middle-track schools (*Realschule*), 66 higher-track schools (*Gymnasium*), and 28 comprehensive schools (*Gesamtschule* or *Schule mit mehreren Bildungsgängen*).

To understand our analyses, it is important to stress that students typically do not remain in the same class from grade 5 to 7 in Germany. Some schools establish new classes after grade 6 and individual students can switch to parallel classes. We observed that the students in our sample belonged to 250 classes in grade 5, 271 classes in grade 6, and 330 classes in grade 7. Only 820 students attended the same German class in all three grades, 1566 attended two different German classes, and 681 changed German class every year. Furthermore, even if the students remained in the same class, the teachers often changed between the grades. Within grades, the students of a class had the same German teacher, though. Therefore, we created a dataset that entailed the information from the individually attended German classes in grades 5, 6, and 7 per student. The student- and teacher-reported teaching quality measures were therefore only identical for students of the same school, who attended the same German classes in grades 5, 6, and 7.

Table 1
Descriptive Statistics of Variables on Student Level.

Variable	Source	% missing	*M* (*SD*) or % in categories
Reading competence			
Reading competence, grade 5[1]	student test	5%	0.13 (1.44)
Reading competence, grade 7[1]	student test	6%	0.81 (1.52)
Attitudes towards reading			
Attitudes towards reading, grade 5[2]	student quest.	11–12%[3]	3.04 (0.78)
Attitudes towards reading, grade 7[2]	student quest.	8–9%[3]	2.73 (0.83)
Control variables			
Gender[4]	sample info.	0%	51% female
Age in years[5]	sample info.	5%	10.92 (0.62)
Grade in subject German[5,6]	student quest.	12%	4.72 (1.06)
Satisfaction with school[5,7]	student quest.	8%	7.89 (2.59)
Number of books at home[5,8]	student quest.	11%	4.02 (1.54)
Parental education level[5,9]	parent interview	30%	6.04 (3.10)
Language of origin[5,10]	parent interview	29%	89% German
Migration background[5,11]	student quest.	6%	68% no migration background

Note. Reported are unstandardized descriptive statistics. [1]WLE score with higher values indicating better reading competence. [2]From 1 =*completely disagree* to 4 =*completely agree*. [3]Range reflects item level missingness. [4]Dummy 0 =*male* and 1 =*female*. [5]Measured in grade 5. [6]From 1 =*failing* to 6 =*very good*. [7]From 0 =*completely dissatisfied* to 10 =*completely satisfied*. [8]From 1 =*none or only very few (0 to 10 books)* to 6 =*enough to fill a shelf unit (more than 500 books)*. [9]From 0 =*no certificate* to 10 =*award of a doctorate, habilitation*. [10]Dummy on language child learned in family in first three years 0 =*not (only) German* and 1 =*German*. [11]Dummy 0 =*child, parent(s), and/or grandparent(s) born abroad* and 1 =*child, parents, and grandparents born in Germany*.

2.2. Instruments

2.2.1. Reading competence outcome

In NEPS, reading competence was assessed by two different standardized paper-pencil tests in grades 5 and 7 following the same framework established by Weinert et al. (2011) and Gehrer et al. (2012). The reading competence test aimed to evaluate students' reading skills across various text types and cognitive requirements. The theoretical framework included five distinct text types: information texts, commenting or argumenting texts, literary texts, instruction texts, and advertising texts. Each text type presented specific characteristics and challenges that required students to apply different reading strategies. Within each text type, three cognitive requirements were assessed: finding information in the text, drawing text-related conclusions, and reflecting and assessing the content. These cognitive requirements were designed to measure students' abilities to comprehend, analyze, and critically evaluate texts. The test consisted of a total of 29 items in grade 5 and 42 items in grade 7. The response formats utilized in the test included simple multiple-choice (MC) items, complex multiple-choice

[2] Starting Cohort Grade 5, doi:10.5157/NEPS:SC3:8.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). Since 2014, NEPS has been carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

[3] Newly immigrated students who do not master German well enough to take the test, students at special schools, and students from later additional samples are not included in the regular sample.

(CMC) items, and matching (MA) items. MC items presented four response options, with one correct solution and three distractors. CMC items involved multiple subtasks, each with two response options. MA items required test takers to match responses to a given set of statements, often used for assigning headings to paragraphs in a text.

The age-specific versions of the test were linked between grades 5 and 7 to measure changes in competence on the same measurement scale (Fischer, Rohm, Gnambs, & Carstensen, 2016). The weighted maximum likelihood estimate (WLE) scores had high reliabilities of.767 in grade 5 and.791 in grade 7 (Krannich et al., 2017; Pohl, Haberkorn, Hardt, & Wiegand, 2012). As Table 1 shows, reading competence increased on average by a half standard deviation over the two years (0.13 in grade 5 to 0.81 in grade 7). In grade 7, the intra-class correlation coefficient (ICC) of reading competence between schools was.236.

2.2.2. Attitudes towards reading

The students' attitudes towards reading were assessed in student questionnaires in grades 5 and 7. The identical scales were introduced with the question "what do you think about reading?" followed by six items like "I find reading interesting," "if I had enough time, I would read even more," or "I like to read about new things". The response scale ranged from 1 =*completely disagree* to 4 =*completely agree* (LIfBi, 2016, 2018). Cronbach's α for this scale was.830 in grade 5 and.871 in grade 7, indicating a good internal consistency of the scale. For the analyses, we calculated a mean scale from the six items. Consistent with previous research indicating that attitudes towards reading decline during the transition from one school level to the next (e.g., Nootens et al., 2019), Table 1 shows that the mean self-reported attitudes towards reading slightly decreased between grades 5 and 7. In grade 7, the ICC of attitudes towards reading between schools was.114.

2.2.3. Teaching quality

NEPS surveys contain 10 scales to capture different components of teaching quality in German language classes (LIfBi, 2016, 2018). According to the developers of the NEPS survey instruments (Frahm et al., 2011; Fabian et al., 2019), the 10 teaching quality scales in NEPS are grounded in the theoretical model of teaching and learning developed by Bolhuis (2003), as conceptualized by Seidel and Shavelson (2007). Table 2 shows that each components of teaching quality was measures with three to eleven items. Half of the scales were administered to students, while the other half were given to teachers, thereby leveraging the specific validity of both perspectives on teaching quality. We view teaching quality as an inherent attribute of teachers, and, as a result, we aggregate the feedback from all students who rate the same teacher.

The same instruments were administered in grades 5, 6, and 7. To summarize the information on each teaching quality component across grades, we first computed the scale mean (across items) for each grade and then calculated the means across the three grades. Table 2 depicts the descriptive statistics for each scale. The intraclass correlations (ICCs) for all scales are greater than.5, indicating substantial variation in teaching quality across schools. Appendix A displays all teaching quality items, along with the proportion of missing values and the internal consistency of each scale (Cronbach's alpha) by grade. The 10 scales exhibit reliability ranging from alpha = .5 to.8.

The correlations between the teaching quality components indicate that both student and teacher ratings were generally more highly correlated within the rating group than between the groups (see Table 3). Beyond that, some teaching quality components were more highly correlated than others.

2.2.4. Control Variables

To account for confounding effects, next to prior reading competence and attitudes towards reading in grade 5, demographic (gender, age, language of origin, migration background), socio-economic (books at home, parental education), and school-related (grade in subject German, school satisfaction) variables were considered in the analyses. Table 1 provides operationalization details and descriptive statistics for the control variables.

2.3. Imputation of missing values

The NEPS dataset contained missing data on different levels (see Table 1 and Appendix A). Since we could not assume that this missingness occurred completely at random, we imputed missing data using predictive mean matching. To optimally predict missing values on both the student and German class level, we performed two parallel imputations for these levels. Both imputation models included the student and German class sample described in the data section. Afterwards, we combined the imputed student-level datasets with the respective imputed teacher information from the German classes attended in grades 5, 6, and 7. We imputed each missing value five times, replicated all analyses, and combined the estimates using Rubin's (1987) rules.

Specifically, we imputed the missing data on the teacher ratings of teaching quality in a model that incorporated a number of further imputation covariates from all three waves.[4] In order to impute the student-level teaching quality items, reading competence, attitudes towards reading, and student control variables, we set up a similar imputation model in which we also included a number of variables from different waves.[5] To replicate the data structure, we included the sampling weights as an imputation predictor and accounted for the nesting in schools by including school identification dummies in both types of imputation models. We employed the R-package mice for imputing missing data (van Buuren & Groothuis-Oudshoorn, 2011).

2.4. Analytical strategy

We used nonexperimental data to identify the effects of teaching quality on student competence. The consequential empirical issues are best illustrated by considering student competence as a function of student, teacher, and school features: First, student variables and teaching quality may be confounded due to selection mechanisms. In the tracked German school system, teaching quality may be distributed unequally across different school types (e.g., lower-track vs. higher-track schools). Prior research has shown that competence and social background are important predictors of the allocation of students to school types (e.g., Maaz, Trautwein, Lüdtke, & Baumert, 2008). This makes a confounding of student variables and teaching features likely. Second, different components of teaching quality are correlated with each other. For example, teachers who clearly articulate performance goals evaluate learning progress more frequently, making it difficult to disentangle the effects. Third, teaching quality could be correlated with general school quality (e.g., leadership, infrastructure).

Our approach for identifying the impact of teaching quality is best illustrated by a linear equation with school-fixed effects:

$$Y_{ijk7} = \alpha B_{ijk5} + \beta T_{jk} + \mu_k + \epsilon_{ijk} \quad (1)$$

In this model, Y is the reading test score of student i in class j and school k in grade 7. The outcome was regressed on a rich set of student variables, B, which included prior competence, attitude, and background measures, all of which were observed in grade 5. The main explanatory variables were a vector of 10 teaching quality scales T.

[4] These included background characteristics of the German teachers (year of birth, gender, migration background, general attitudes about education and schooling) and the aggregated student ratings on teaching quality.

[5] We added additional competence test scores (mathematics, orthography, perceptual speed, declarative metacognition, information and communication technology literacy, scientific literacy, reading speed), a self-assessment in reading, repetitions of the student controls in later waves, and additional background information (mathematics school grades, homework help from parents and tutors).

Table 2
Descriptive Statistics of Teaching Quality Scales.

Short name	Teaching quality component (Seidel & Shavelson, 2007)	Source	# of items	Example item	*M* (*SD*)	ICC
[01] Organization[1]	Organization of learning	student	4	I think my German teacher has the class under control.	3.41 (0.35)	.574
[02] Autonomy[1]	Social context–perceived teacher autonomy support	student	3	My German teacher encourages me to ask questions.	3.38 (0.33)	.570
[03] Interaction[1]	Social context–promoting interaction	student	3	My German teacher encourages us to exchange ideas with each other in class.	3.37 (0.34)	.593
[04] Goal setting[1]	Goal setting and orientation	student	5	I think my German teacher expects me to try my very best.	3.39 (0.28)	.657
[05] Orientation[1]	Goal setting and orientation	student	3	My German teacher sums up the most important things at the end of the lesson.	3.14 (0.35)	.619
[06] Activation[2]	Execution of learning–basic processing	teacher	11	The students are requested by me to relate to the questions and comments of their classmates.	3.76 (0.42)	.558
[07] Challenging assignments[1]	Execution of learning–basic processing	teacher	4	I give them assignments that require explanations and in depth comments rather than simple solutions.	3.47 (0.53)	.556
[08] Student discussion[3]	Execution of learning–social interaction	teacher	4	Discussion rounds.	4.11 (0.70)	.570
[09] Evaluation[3]	Evaluation	teacher	9	Oral testing of students.	3.32 (0.45)	.587
[10] Adaptivity[1]	Regulation and monitoring	teacher	6	I demand considerably less from students who are less capable.	3.12 (0.40)	.545

Note. Reported are unstandardized descriptive statistics. Teacher questionnaire variables calculated as means across scale items (see Appendix A) and measurements in grades 5–7. Student questionnaire variables calculated as means across students, scale items (see Appendix A), and measurements in grades 5–7. [1]From 1 =*does not apply at all* to 5 =*applies completely*. [2]From 1 =*very rarely* to 5 =*very often*. [3]From 1 =*never* to 6 =*(almost) every lesson*.

Table 3
Correlations between the Teaching Quality Components.

	[01]	[02]	[03]	[04]	[05]	[06]	[07]	[08]	[09]	[10]
[01] Organization	1									
[02] Autonomy	.685 *	1								
[03] Interaction	.551 *	.780 *	1							
[04] Goal setting	.434 *	.263 *	.166 *	1						
[05] Orientation	.530 *	.554 *	.458 *	.594 *	1					
[06] Activation	.134 *	.099 *	.047 *	-.100 *	-.012	1				
[07] Challenging assignm.	.023	-.047 *	-.098 *	-.085 *	-.056 *	.544 *	1			
[08] Student discussion	.153 *	.192 *	.297 *	.085 *	.092 *	.404 *	.234 *	1		
[09] Evaluation	.060 *	.017	-.069 *	.185 *	.126 *	.207 *	.215 *	.301 *	1	
[10] Adaptivity	-.019	-.031	-.026	.222 *	.116 *	.082 *	.141 *	.281 *	.474 *	1

Note. Reported are bivariate correlations between scale scores. * $p < .050$.

Given that various teaching quality components were confounded, we included them in a joint model where we estimated the effects of all components simultaneously and estimated the impact of each teaching component while controlling for the others.

The most significant remaining issues were unobserved school factors that are correlated with teaching quality. It is difficult to distinguish between teacher and school effects if the data available per school stems from only one teacher. However, NEPS sampled students from different German classes within a school and most students transited through different classes from grade 5 to 7. This design allowed us to add school-fixed effects, μ, to account for bias from any unobserved confounding variables on school (or higher) level. By implication, the estimation of β is based upon within-school variation in teaching quality.[6]

To generalize our findings to the underlying student population, we used NEPS sampling weights that accounted for different sampling probabilities and initial and wave-specific nonresponse (*w_t_cal*; Steinhauer & Zinn, 2016). All analyses were replicated using attitudes towards reading as the outcome variable. We employed the R-package lfe for the estimation of the the fixed effects models (Gaure, 2013).

[6] Note that the model is not a multilevel model because school-level variation is absorbed by school fixed effects (see Allisson, 2009).

3. Results

To provide a point of comparison with our main analysis below, we first estimated a simple cross-sectional model in which we regressed standardized reading competence scores in grade 7 on the 10 standardized measures of teaching quality without controlling for other covariates and without school-fixed effects. Table 4 shows mixed findings for the cross-sectional associations between students' reading competence and teaching variables (column 1). The standardized regression coefficients for organization and activation were small but statistically significant; the coefficients for autonomy, interaction, challenging assignments, and student discussion were not significantly different from zero. The parameters for goal setting and orientation, evaluation, and adaptivity were significant and negative. It is of important to acknowledge that these associations derive from cross-sectional data, without the incorporation of controls for potential confounding variables. Therefore, we refrain from interpreting the direction and magnitude of any of these associations. The strength of the associations, however, decreased considerably after controlling for prior competence and other key student covariates (column 2). This observed change in the parameters indicates that selection mechanisms biased the estimation of teaching quality effects. We notice that performance in grade 5 serves as the most significant indicator of performance in grade 7. It is s important to note that the impact of social background variables is minimal due to our control for prior achievement. In our main model,

Table 4
Results of the Regression of Reading Competence and Attitudes towards Reading on Teaching Quality and Student Controls in School-Fixed Effect Models.

	Reading competence grade 7			Attitudes towards reading grade 7		
	β (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)
	[01]	[02]	[03]	[04]	[05]	[06]
Teaching quality						
Organization	0.12 *	0.06 *	0.01	0.08 *	0.04	0.01
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
Autonomy	0.08	0.03	0.00	0.09 *	0.02	0.03
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.06)
Interaction	0.07	0.05	0.09	0.09	0.06	0.07
	(0.04)	(0.03)	(0.05)	(0.04)	(0.03)	(0.05)
Goal setting	-0.22 *	-0.11 *	-0.05	-0.13 *	-0.06 *	-0.02
	(0.03)	(0.03)	(0.06)	(0.02)	(0.02)	(0.04)
Orientation	-0.21 *	-0.10 *	-0.03	-0.14 *	-0.04	-0.03
	(0.03)	(0.03)	(0.05)	(0.03)	(0.02)	(0.04)
Activation	0.06 *	0.02	-0.05	0.04	0.02	0.04
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
Challenging assignments	0.00	-0.02	-0.01	0.07 *	0.04	-0.04
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
Student discussion	-0.04	-0.03	0.01	-0.03	-0.03	-0.05
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
Evaluation	-0.05 *	-0.04	0.02	-0.05 *	-0.05 *	0.10 *
	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
Adaptivity	-0.05 *	-0.03	-0.04	0.00	0.02	0.00
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)
Student controls						
Reading competence grade 5	-	0.31 *	0.30 *	-	0.11 *	0.08 *
		(0.02)	(0.02)		(0.02)	(0.02)
Attitudes towards reading grade 5	-	0.04 *	0.04 *	-	0.31 *	0.32 *
		(0.02)	(0.02)		(0.02)	(0.02)
Gender	-	0.04 *	0.04 *	-	0.10 *	0.10 *
		(0.02)	(0.02)		(0.02)	(0.02)
Age in years	-	-0.02	-0.01	-	-0.02	0.00
		(0.02)	(0.02)		(0.02)	(0.02)
Grade in subject German	-	0.10 *	0.09 *	-	0.05 *	0.04 *
		(0.02)	(0.02)		(0.02)	(0.02)
Satisfaction with school	-	0.00	-0.01	-	-0.04 *	-0.03
		(0.02)	(0.02)		(0.02)	(0.02)
Number of books at home	-	0.06 *	0.04	-	0.06 *	0.04 *
		(0.02)	(0.02)		(0.02)	(0.02)
Parental education level	-	0.03	0.01	-	0.05 *	0.03
		(0.02)	(0.02)		(0.02)	(0.02)
Language of origin	-	0.02	0.02	-	0.01	0.02
		(0.02)	(0.02)		(0.02)	(0.02)
Migration background	-	0.00	0.01	-	0.03	0.03
		(0.02)	(0.02)		(0.02)	(0.02)
School-fixed effects	no	no	yes	no	no	yes
R^2	0.15	0.27	0.31	0.07	0.23	0.26

Note. Reported are standardized regression coefficients. * $p < .050$.

we added school-fixed effects to effectively control for any confounding variable at school or higher levels (column 3). In these models, the regression coefficients of the teaching quality variables were not significant.

The analyses using attitudes towards reading as the dependent variable showed similar patterns as the previously reported findings for reading competence. Cross-sectional associations (column 4) mostly vanished after controlling for prior attitudes towards reading (column 5). In the main model with school-fixed effects, no teaching variable—except for evaluation—had significant effects on student attitudes towards reading (column 6). However, the standardized regression parameter for evaluation was very small. A central element of our analysis strategy is to analyze different dimensions of teaching quality in parallel in order to control for potential confounding between these dimensions. A drawback of this approach, however, is that the Type I error rate increases as multiple comparisons are being made. Because of this, we applied a Bonferroni correction, and as a result, the parameter for evaluation is also no longer significant.

We conducted a set of robustness checks. Conventional two-level analyses with controls (see Appendix B) using the R-package lme3 (Bates, Mächler, Bolker, & Walker, 2015) as well as models in which the teaching quality components were introduced separately (see Appendix C) resulted in qualitatively similar findings.

4. Discussion

In the realm of educational research, teaching quality emerges as a predominant focus, yet the field is marked by a persistent ambiguity in its conceptual framework. This lack of consensus is further compounded by the methodological limitations that characterize much of the existing literature, including issues such as non-representative sampling, reliance on cross-sectional study designs, and a narrow concentration on isolated facets of teaching quality. Addressing these pervasive gaps, our study embarks on a comprehensive examination underpinned by a robust theoretical model. We have utalized representative large-scale longitudinal data from Germany to investigate the impact of ten distinct components of teaching quality on both student achievement and motivation.

In a landscape of research seeking to quantify the influence of educational practices on student outcomes, our findings make a pivotal contribution. Contrary to the anticipated outcomes, our analysis reveals that none of the assessed components of teaching quality exert a measurable effect on student achievement. This conclusion aligns with the observations made in the seminal meta-analysis by Seidel and Shavelson (2007), reinforcing the notion that the effectiveness of teaching practices, as traditionally defined and measured, may not translate into quantifiable academic benefits for students. Our research not only echoes these earlier insights but also extends them by offering empirical evidence from a contemporary, large-scale dataset, thereby providing a nuanced perspective on the intricate dynamics of educational efficacy. We will elaborate on these issues below.

4.1. Teaching as a Multidimensional Process

Teaching is a multidimensional endeavor, which, when analyzed in empirical studies, requires researchers to consider different aspects in parallel. Studies that examine individual teaching components in isolation may be biased if the component under investigation is correlated with other unobserved teaching aspects. Although confounding is a well-established concept, it has been largely ignored in a large number of studies on teaching. Hence, applying a multidimensional approach to learning provides a basis to capture the complex situation of students in classes and schools. In line with this idea, the model of teaching and learning components (Bolhuis, 2003; Seidel & Shavelson, 2007) has identified a set of teaching quality components that are considered to be important for student's cognitive and motivational-affective outcomes. It was assumed that teachers play an important role in shaping students' learning environment, which in turn influences cognitive (e.g., reading competence) and motivational-affective (e.g., attitudes towards reading) outcomes (Hattie, 2009; Muijs et al., 2014). But as the idea of teaching operates on a meta-level and research on teaching has used different and overlapping terms to describe teaching quality components, it is difficult to compare and integrate results from different studies.

Using data from the NEPS study enabled us to account for the effects of a wider set of teaching quality components on student outcomes in parallel and to control for prior achievement and attitudes, a large set of further controls, and to estimate school fixed effects. Aiming at high transparency, we presented results including bivariate associations between teaching components, models that considered students' prior achievement, attitudes, background information, and models that eliminated unobserved variance between schools.

Our results indicated that, based on bivariate correlations, teaching components are associated. That is, several components (such as

organization, autonomy, and interaction) do not occur independently of each other. Although these components are theoretically distinctive, they are correlated in the NEPS data (see Table 3). This underscores the significance of accounting for these confounding by modeling them simultaneously. At the same time, it is noteworthy that the correlations do not reach levels indicative of multicollinearity concern, as all correlations remain below 0.8, as suggested by Berry and Feldmann (1985). Further, we observed some cross-sectional associations between the teaching components and reading competence respectively attitudes towards reading. However, when prior achievement, attitudes, and background variables were taken into account, the significant effects on students' outcomes vanished. Although there were a few significant coefficients, the effect sizes were unanimously small. When we additionally accounted for school-fixed effects, we found only one significant, very small effect of evaluation on students' attitudes towards reading. Even if previous research has found a (negative) effect of evaluation on mathematics achievement (Meijnen, Lagerweij, & de Jong, 2003), this finding should be interpreted with caution as the effect was very small and the robustness checks did not confirm the finding. All other variables were not significantly related to the two outcomes in grade 7. Further, after a Bonferroni correction, even the parameter for evaluation was no longer significantly different from zero. Overall, the robustness checks confirmed the pattern of these findings, which indicated only some small significant effects.

Hence, we could not confirm either of our hypotheses. We did not find the assumed small to moderate effect sizes of the 10 proximal and distal teaching quality components for reading competence or attitudes to reading. Some previous studies found a significant impact of some teaching quality on students' cognitive and motivational-affective outcomes while others did not (Scheerens & Bosker, 1997; Scheerens et al., 2007; Seidel & Shavelson, 2007). We would also like to emphasize that our study results are in line with previous German studies. Previous large-scale longitudinal studies with random samples have not found consistent effects of individual components of teaching quality on achievement development (e.g., Baumert et al., 2010; Helmke et al., 2008; Klieme et al., 2008; Praetorius et al., 2020). A consistent finding is, however, that even the few significant effects were small in size, especially if student characteristics, prior competence, and factors on the class and school level were considered (Klieme et al., 2009). One possible explanation is that there is little variation in learning gains that could be explained by instructional quality in the first place (Klieme et al., 2009; Seidel & Shavelson, 2007). This is one explanation why longitudinal multi-level survey studies have found few significant effects of teaching quality, as in our study. In a similar vein, recent internationally comparative studies have found inconsistent results (Blömeke et al., 2016; Nortvedt et al., 2016). For instance, a study by Nortvedt et al. (2016) found significant positive correlations between instructional quality and reading and mathematical competence in fewer than half of the participating educational systems. Germany belonged to the countries with insignificant associations. Similarly, Blömeke et al., (2016) found insignificant associations in most countries. In Germany, they found no significant effect of instructional quality on student achievement, but the teachers' education degree and major focus of teacher education had an effect (Blömeke et al., 2016). Furthermore, studies using observational measures (e.g. TALIS video-study) showed no significant effects on the achievement and interest of students, once students' previous interest and performance were taken into account (Doan et al., 2020; Praetorius et al., 2020). Therefore, the observation-based and questionnaire-based studies showed similar results and emphasize the importance of longitudinal studies in the investigation of teaching quality effects. Future research could use multi-modal and longitudinal analyses to further investigate robust teaching quality effects.

Finally, it is worth noting that our analyses covered only two years of learning and teachers alone are not the only relevant factor when considering the development of students' cognitive and motivational-affective outcomes, especially in reading. Based on the conceptual framework of determinants of student outcomes (Creemers & Kyriakides, 2018) the students' background, the school level (e.g. school climate), and the national level (e.g. educational policy) could account for additional variance. Hence, the present study included several student controls. As our results indicate, and in line with Nilsen, Gustafsson, and Blömeke (2016), both the inclusion of student controls and the school-fixed effects increased the amount of explained variation in the model. This reveals the complexity of the educational process.

4.2. Strengths and Limitations

The current study has several strengths. First, we simultaneously considered 10 different components of the Seidel and Shavelson (2007) teaching and learning components model, which meant the analysis could account for their confounding. In meta-analyses (e.g., Scheerens & Bosker, 1997; Seidel & Shavelson, 2007), such effects might have been underestimated (Muijs et al., 2014). Second, we investigated whether teaching quality matters for both students' reading competence and attitudes, which is important since the effects might vary between outcomes (Seidel & Shavelson, 2007). Third, the longitudinal design allowed to both improve the reliability of the teaching quality scales and to control for students' prior competence and attitudes. Fourth, applying school-fixed effects in the models controlled for possible unobserved confounders (omitted variable bias) on the school and higher levels. Fifth, as the NEPS data contains missing values like most longitudinal studies, we circumvented potential generalizability and validity issues by applying a sophisticated two-level data imputation strategy. Sixth, the study took on board the advice to use a mix of data sources (Seidel & Shavelson, 2007; Raudenbush, 2005) by combining both student- and teacher-reported teaching quality measures.

This also points to one of the study's limitations. Although we used a mix of data sources, we were not able to use classroom observations or video data. Such data would add another valuable perspective (Seidel & Shavelson, 2007). Further, for economic reasons, NEPS surveyed each teaching component by teachers *or* students, but not both. For this reason, it is not possible to compare student and teacher ratings of the same teaching component. Methodological limitations also included the reliability of some of the components. For example, the Cronbach's alpha for evaluation of learning success, student discussion, adaptivity, and goal setting were rather low (see Appendix A). However, we do not consider reliability a major issue because we used large-scale data from three survey waves. Although the analyses are based on longitudinal data, both students and teachers changed learning groups over the waves. We circumvented this natural limitation by analyzing the means across the three waves. Most students were thus exposed to a mixture of different instructional quality levels. Another natural limitation concerns the fact that reading is taught as a part of the subject German language, not as a separate subject. This implies that the associations between teaching and reading outcomes might be closer in education systems in which teaching to read can be investigated, separately.

Lastly, we would also like to discuss the generalizability of our results. Although NEPS uses random sampling to generalize the results to Germany, there are a few issues that need to be taken into account. First, no data were available for Berlin, as the Berlin school system differs from that of other German states. Second, NEPS developed separate instruments for special education schools and as the relevant variables were not administered in the special education school, we could not use their data. Third, children who changed schools, repeated a grade, or skipped a grade were not tested and surveyed in seventh grade. These are important aspects that should be taken into account when interpreting the results of our study.

5. Conclusion

To sum up, the present study took the multifaceted nature and goals

of schooling into account (Reynolds et al., 2014) when disentangling the effects of a set of teaching quality components on reading competence and attitudes towards reading. Using longitudinal data in a regression with school-fixed effects and student controls, the study found mostly insignificant effects of teaching quality. Hence, the present study complements the state of literature with a study that investigates a comprehensive set of student- and teacher-reported teaching qualities and students' cognitive and motivational-affective outcomes in the domain of reading with a longitudinal approach. Taking previous competence and attitudes into account, our study could not identify significant associations between the investigated teaching quality components and the reading competence and attitudes towards reading measures, when considered simultaneously in one model.

CRediT authorship contribution statement

Rolf Strietholt: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Isa Steinmann:** Conceptualization, Formal analysis, Writing – original draft. **Manja Attig:** Conceptualization, Formal analysis, Writing – original draft. **Frances Hoferichter:** Conceptualization, Formal analysis, Writing – original draft.

Acknowledgements

We thank Tobias Linberg for his support in preparing the manuscript, Roisin Cronin for copy editing, and Hannah Heister for her assistance with the analyses.

Statements and declarations

The authors declare that they have no financial or non-financial interests to disclose that are directly or indirectly related to the work submitted for publication. This work was supported by the College for Interdisciplinary Educational Research (CIDER), a joint initiative of the BMBF, the Jacobs Foundation and the Leibniz Association.

Appendix

Appendix Table A
Properties of the Teaching Quality Scales.

			Grade 5		Grade 6		Grade 7		Item details	
Scales	Source	# of items	% missing	α	% missing	α	% missing	α	Response scale	Scale and item wording
Organization	SQ	4	8-10%	.70	8-9%	.72	9%	.75	From 1 =*does not apply at all* to 5 =*applies completely*	*I think my German teacher* *… is aware of everything that happens in class.* *… manages to quickly involve me again, if I don't pay attention for a moment.* *… instantly notices when I don't pay attention.* *… has the class under control.*
Autonomy	SQ	3	12-13%	.66	10-11%	.71	11-12%	.69	From 1 =*does not apply at all* to 5 =*applies completely*	*My German teacher* *… first tries to understand my point of view, and then tells me what he/she would do.* *… listens to my suggestions and takes them seriously.* *… encourages me to ask questions.*
Interaction	SQ	3	11-12%	.69	10-11%	.64	10-11%	.67	From 1 =*does not apply at all* to 5 =*applies completely*	*My German teacher* *… allows us to discuss our assignments with each other.* *… encourages us to help each other in class.* *… encourages us to exchange ideas with each other in class.*
Goal setting	SQ	5	9-12%	.55	9-10%	.54	10-11%	.57	From 1 =*does not apply at all* to 5 =*applies completely*	*I think my German teacher* *… expects me to try my very best.* *… tells me that she/he thinks that I can do better than I have done so far.* *… finds it very important that we do our work very thoroughly.* *… uses students that achieve good grades as an example for us all.* *… tells us where we stand compared to our classmates.*
Orientation	SQ	3	12%	.67	11%	.64	10-11%	.67	From 1 =*does not apply at all* to 5 =*applies completely*	*My German teacher …* *… sums up the most important things at the end of the lesson.* *… gives us information as to what is especially important in the lesson.* *… explains to us how old and new topics relate to each other.*
Activation	TQ	11	31-32%	.77	27-28%	.79	43-44%	.78	From 1 =*very rarely* to 5 =*very often*	*How often do the following statements apply to German lessons in your class? The students …* *… have the freedom to develop their own understanding during reading and writing.* *… consciously and purposefully discuss things with me and their classmates.* *… are requested to comment orally, express their own views or personal impressions.* *… are asked questions that show if they have understood the subject matter in depth.* *… are asked questions that show if they are able to critically assess and analyze the subject matter.* *… may steer discussions in new directions.*

(*continued on next page*)

Appendix Table A (*continued*)

Scales	Source	# of items	Grade 5		Grade 6		Grade 7		Item details	
			% missing	α	% missing	α	% missing	α	Response scale	Scale and item wording
										… are requested by me to relate to the questions and comments of their classmates. *… actually relate to the questions and comments of their classmates.* *… provide counterarguments, comments or opinions to their classmates or to my own statements.* *… question the interpretation of texts* (e.g. *by showing alternative perspectives).* *… are asked questions during which the subject matter has to be critically reviewed.*
Challenging assignments	TQ	4	31-33%	.71	27-28%	.74	42-43%	.73	From 1 =*does not apply at all* to 5 =*applies completely*	*To what extent do the following statements apply to the assignments you give your students during German lessons?* *… I give them assignments that do not only involve the identification of standard solutions but also the selection of the right approach.* *… I give them assignments in which the students need time to think in order to find solutions.* *… I give them assignments in which the students have to show different approaches.* *… I give them assignments that require explanations and in depth comments rather than simple solutions.*
Student discussion	TQ	4	31-33%	.51	27-30%	.49	41-43%	.51	From 1 =*never* to 6 = *(almost) every lesson*)	*How often do you use the following social methods of learning in this German class?* *… Work with small student groups* *… Partner work* *… Discussion rounds* *… Students acting as tutors (“Learning by Teaching“, peer tutoring)*
Evaluation	TQ	9	31-34%	.46	27-31%	.48	43-45%	.47	From 1 =*never* to 6 = *(almost) every lesson*)	*How often do you use the following types of learning success control methods in your German lesson?* *… Tests developed by me* *… Parallel/comparative tests,* i.e. *the same tests for all classes and learning groups* *… Observation of oral participation* *… Oral testing of students* *… Diagnostic tests or externally developed standardized tests* *… Project-based work* *… Homework, home assignments* *… Student folders* *… Portfolio of the learning process*
Adaptivity	TQ	6	31-32%	.59	27-28%	.58	42-44%	.56	From 1 =*does not apply at all* to 5 =*applies completely*	*To what extent do the following statements apply to your German lessons in this class?* *… I demand considerably less from students who are less capable.* *… I form groups of students with similar capabilities.* *… I give students homework ranging in complexity based on their capability.* *… I allow students who work faster to move on to the next assignment while I am still practicing or reviewing things with the ones that work slower.* *… If students have difficulties in understanding, I give them additional assignments.* *… I give more capable students extra assignments that are really challenging for them.*

Note. TQ is short for teacher questionnaire and SQ for student questionnaire.

Appendix B

Robustness Checks: Conventional Two-Level Analyses with Controls.

The main analyses with school-fixed effects controlled for unobserved confounding variables at school and higher levels. A drawback of this approach is that school-fixed effects reduce the variance in the explanatory variables that is used in the estimation of teaching effects by the amount of variation that exists between schools. This loss of variation leads to a decrease in the precision of our estimates. Our approach to explore consequences for the present study was to estimate a conventional two-level linear regression model without school-fixed effects. The multilevel model did not control for school-level confounding factors. To reduce such bias, we controlled for two observed key covariates on school level, the school type (i.e., upper school track, lower school track, etc.) and the location of the school (i.e., 15 federal states).

The results of the two-level analysis with the outcome student competence are summarized in Table Appendix B (column 1). The estimates for the different components of teaching quality were similar to the estimates of the main analyses with school-fixed effects. Most estimates were not significant. In contrast to the main analyses, however, three regression coefficients were significant. Goal setting and orientation showed significant small negative and interaction significant very small positive effects on reading competence. Altogether the effects are qualitatively the same.

The two-level analysis on attitudes towards reading was largely consistent with our main results from the fixed effects model with two exceptions (see column 2). The effect of evaluation was not significant as in the main analyses and goal setting showed a very small negative effect.

Appendix Table B
Results of the Two-Level Regression Models of Reading Competence and Attitudes towards Reading on Teaching Quality and Student and School Controls.

	Reading competence grade 7	Attitudes towards reading grade 7
	β (*SE*) [01]	*β* (*SE*) [02]
Teaching quality		
Organization	0.02	0.02
	(0.03)	(0.03)
Autonomy	0.05	0.02
	(0.04)	(0.04)
Interaction	0.07 *	0.06
	(0.04)	(0.04)
Goal setting	-0.11 *	-0.06 *
	(0.03)	(0.03)
Orientation	-0.09 *	-0.05
	(0.03)	(0.03)
Activation	-0.01	0.04
	(0.03)	(0.03)
Challenging assignments	0.00	0.02
	(0.03)	(0.02)
Student discussion	-0.01	-0.03
	(0.02)	(0.02)
Evaluation	-0.01	-0.01
	(0.03)	(0.02)
Adaptivity	-0.04	0.01
	(0.02)	(0.02)
Student controls	yes	yes
School type	yes	yes
Federal state	yes	yes
R^2	0.31	0.25

Note. Reported are standardized regression coefficients. * $p < .050$.

Appendix C

Robustness Checks: Separate Models for Different Components of Teaching Quality.

A tradeoff that arises when modelling the effects of various teaching variables in a joint model—as applied in the main analyses—is the increasing risk of an overadjustment bias. This can lead to an underestimation of the effects (see Castellano, Rabe-Hesketh, & Skrondal, 2014). To test the sensitivity of our analyses, we estimated separate models with student covariates and school-fixed effects where we entered only one teaching component at a time.

The findings of analyses with the outcome reading competence mainly confirm the results of the main analyses (see Table Appendix C.1). The effects of most teaching components were not significant. Interaction and adaptivity showed very small positive and negative effects on reading competence. With respect to attitudes towards reading, no effect was significant when entering each teaching quality component stepwise (see Table Appendix C.2).

Appendix Table C.1
Results of the Separate Regression Models for Regressing Reading Competence on Components of Teaching Quality.

	Reading competence grade 7									
	β (*SE*) [01]	*β* (*SE*) [02]	*β* (*SE*) [03]	*β* (*SE*) [04]	*β* (*SE*) [05]	*β* (*SE*) [06]	*β* (*SE*) [07]	*β* (*SE*) [08]	*β* (*SE*) [09]	*β* (*SE*) [10]
Teaching quality										
Organization	0.02 (0.03)									
Autonomy		0.03 (0.03)								
Interaction			0.07 * (0.03)							
Goal setting				-0.03 (0.03)						
Orientation					-0.01 (0.03)					
Activation						-0.03 (0.03)				
Challenging assignments							-0.04 (0.03)			

(*continued on next page*)

Appendix Table C.1 (*continued*)

	Reading competence grade 7									
	β (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)
	[01]	[02]	[03]	[04]	[05]	[06]	[07]	[08]	[09]	[10]
Student discussion								-0.02 (0.03)		
Evaluation									-0.04 (0.03)	
Adaptivity										-0.06 * (0.03)
Student controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
School-fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31

Note. Reported are standardized regression coefficients. * $p < .050$.

Appendix Table C.2
Results of the Separate Regression Models for Regressing Attitudes towards Reading on Components of Teaching Quality.

	Attitudes towards reading grade 7									
	β (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)	*β* (*SE*)
	[01]	[02]	[03]	[04]	[05]	[06]	[07]	[08]	[09]	[10]
Teaching quality										
Organization	0.04 (0.03)									
Autonomy		0.04 (0.03)								
Interaction			0.05 (0.03)							
Goal setting				0.02 (0.04)						
Orientation					0.02 (0.03)					
Activation						0.02 (0.03)				
Challenging assignments							0.00 (0.03)			
Student discussion								-0.02 (0.03)		
Evaluation									0.04 (0.03)	
Adaptivity										-0.01 (0.03)
Student controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
School-fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
R^2	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26

Note. Reported are standardized regression coefficients. * $p < .050$.

References

Allison, P.D. (2009). Fixed Effects Regression Models. Thousand Oaks: Sage.

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: results from an experimental study. *School Effectiveness and School Improvement, 22*(3), 291–311.

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behavior and student outcomes. *Teaching and Teacher Education, 29*(1), 1–12.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumert, J., Blum, W., & Neubrand, M. (2004). Drawing the lessons from PISA 2000: Long term research implications: Gaining a better understanding of the relationship between system inputs and learning outcomes by assessing instructional and learning processes as mediating factors. In D. Lenzen, J. Baumert, R. Watermann, & U. Trautwein (Eds.), *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (pp. 143–157). VS Verlag für Sozialwissenschaften.

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.

Bergen, E., Zuijen, T., Bishop, D., & Jong, P. F. (2016). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly, 52*, 147–160.

Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen, & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 21–50). Springer International Publishing.

Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction, 79*, Article 101600.

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS). Edition ZfE. (1st ed.). Springer VS.

Blossfeld, H.-P., & Roßbach, H.-G. (Eds.). (2019). Education as a lifelong process: The German National Educational Panel Study (NEPS). Edition ZfE. (2nd ed.). Springer VS.

Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: a multidimensional perspective. *Learning and Instruction, 13*, 327–347.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist, 41*(10), 1069–1077. https://doi.org/10.1037/0003-066X.41.10.1069

Burgess, S., Rawal, S., & Taylor, E. S. (2023). Teachers' use of class time and student achievement. *Economics of Education Review, 94*, Article 102405.

Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Composition, context, and endogeneity in school and teacher comparisons. *Journal of Educational and Behavioral Statistics, 39*(5), 333–367.

Creemers, B.P.M. (1994). The effective classroom. Cassell.

Doan, S., Mihaly, K., & McCaffrey, D. F. (2020). Relationships between teaching practices and student outcomes. In OECD (Ed.), *Global Teaching InSights: A Video Study of Teaching* (pp. 269–283). OECD Publishing.

Fabian, P., Goy, M., Jarsinki, S., Naujokat, K., Porsch, A., Strietholt, R., Blatt, I., & Bos, W. (2019). Transition and Development from lower secondary to upper secondary school. In H.-P. Blossfeld, & H.-G. Roßbach (Eds.), *Education as a Lifelong Process. The*

German National Educational Panel Study (NEPS). Wiesbaden: Springer VS [Edition ZfE].
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.
Fauth, B., Decristan, J., Decker, A.-T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education, 86*(2019), Article 102882.
Fischer, L., Rohm, T., Gnambs, T., & Carstensen, C.H. (2016). Linking the data of the competence tests (NEPS survey paper no. 1). LIfBi, NEPS.
Flisi, S., Meroni, E.C., & Vera-Toscano, E. (2016). Educational outcomes and immigrant background. Publications Office of the European Union.
Frahm, S., Goy, M., Kowalski, K., Sixt, M., Strietholt, R., Blatt, I., Kanders, M., & Bos, W. (2011). Transition and development from Lower Secondary to Upper Secondary School. In: Zeitschrift für Erziehungswissenschaft [Special Issue], 14, 217–232.
Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*, 145–252.
Gaure S. (2013). lfe: Linear group fixed effects. The R Journal, 5(2), 104–117. User documentation of the 'lfe' package, ⟨https://journal.r-project.org/archive/2013/RJ-2013–031/RJ-2013–031.pdf⟩.
Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2012). The assessment of reading competence (including sample items for grade 5 and 9). Scientific use file 2012, Version 1.0.0. University of Bamberg, NEPS.
Graham, L. J., White, S. L. J., Cologon, K., & Pianta, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education, 96*, Article 103190.
Hattie, J.A.C. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge.
Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Klieme, E., Nold, G., & Schröder, K. (2008). Wirksamkeit des Englischunterrichts. In DESI-Konsortium (Ed.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (pp. 382–397). Weinheim: Beltz.
Hochweber, J., & Vieluf, S. (2016). Gender differences in reading achievement and enjoyment of reading: The role of perceived teaching quality. *The Journal of Educational Research, 111*(3), 268–283.
Hoferichter, F., Hirvonen, R., & Kiuru, N. (2021). The development of school well-being in secondary school: High academic buoyancy and supportive class- and school climate as buffers. *Learning and Instruction, 71*(2021), Article 101377. https://doi.org/10.1016/j.learninstruc.2020.101377.
Hoferichter, F., Kulakow, S., & Raufelder, D. (2020). Die Bedeutung von sozio-emotionaler Unterstützung und akademischer Lernunterstützung für epistemische Neugier [The meaning of socio-emotional and academic support from teachers for epistemic curiosity]. In C. Rubach, & R. Lazarides (Hrsg.), *Emotionen in Schule und Unterricht* (S. 155-174). Opladen: Budrich.
Idris, N. (2023). The impact of using Geometers' Sketchpad on Malaysian students' achievement and van Hiele geometric thinking. *Journal of Mathematics Education*, 94–107.
Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive. In J. Baumert, et al. (Eds.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 333–359). Leske + Budrich.
Klieme, E., & Nilsen, T. (2022). Teaching quality and student outcomes in TIMSS and PISA. In T. Nilsen, A. Stancel-Piątak, & J. E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education* (pp. 1089–1134). Springer. ISSN 978-3-030-38298-8. s.
Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study. Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In J. Tomáš, & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.
Klieme, E., Jude, N., Rauch, D., Ehlers, H., Helmke, A., Eichler, W., & Willenberg, H. (2008). Alltagspraxis, Qualität und Wirksamkeit des Deutschunterrichts. In DESI-Konsortium (Ed.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (pp. 382–397). Weinheim: Beltz.
Krannich, M., Jost, O., Rohm, T., Koller, I., Pohl, S., Haberkorn, K., Carstensen, C.H., Fischer, L., & Gnambs, T. (2017). NEPS technical report for reading – scaling results of starting cohort 3 for grade 7 (update NEPS survey paper no. 14). LIfBi, NEPS.
Kulakow, S. & Hoferichter, F. (2021). Der Zusammenhang von Selbstkonzept und Grit: Grit fördern durch Lernunterstützung und Kompetenzerleben [The relationship between self-concept and grit: Supporting grit through teachers' academic and competence support]. In G. Hagenauer & D. Raufelder, Soziale Eingebundenheit: *Sozialbeziehungen im Fokus von Schule und Lehrer*innenbildung* (S. 299-314). Münster: Waxmann.
Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12–23.
Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. In *Teaching and Teacher Education, 36* pp. 143–152).
Kyriakides, L., Creemers, B.P., Panayiotou, A., & Charalambous, E. (2020). Quality and equity in education: Revisiting theory and research on educational effectiveness and improvement. Routledge.
Lauermann, F., & ten Hagen, I. (2021). Do teachers' perceived teaching competence and self-efficacy affect students' academic outcomes? A closer look at student-reported classroom processes and outcomes. *Educational Psychologist, 56*(4), 265–282.
Levine, D.K. & Lezotte, L.W. (1990). Unusually effective schools: A Review and analysis of research and practice. National Center of Effective Schools Research and Development.
LIfBi (2016). Starting cohort 3: grade 5 (SC3) wave 3: questionnaires (SUF version 5.0.0). LIfBi, NEPS.
LIfBi (2018). NEPS starting cohort 3 – grade 5: paths through lower secondary school – educational pathways of students in grade 5 and higher. LIfBi, NEPS.
Maaz, K., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Educational transitions and differential learning environments: How explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives, 2*(2), 99–106.
Marshik, T., Ashton, P. T., & Algina, J. (2017). Teachers' and students' needs for autonomy, competence, and relatedness as predictors of students' achievement. *Social Psychology of Education, 20*, 39–67.
Martin, A. J., & Collie, R. J. (2018). Teacher-student relationships and students' engagement in high school: Does the number of negative and positive relationships with teachers matter? *Journal of Educational Psychology, 111*(5), 861–876.
McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30*(4), 934–956.
Meijnen, G., Lagerweij, N. W., & de Jong, P. F. (2003). Instruction characteristics and cognitive achievement of young children in elemantary schools. *School Effectiveness and School Improvement, 14*(2), 159–187.
Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256.
Muis, K. R., Ranellucci, J., Trevors, G., & Duffy, M. C. (2015). The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction, 28*, 1–13.
Nilsen, T. S., Gustafsson, J.-E., & Blömeke, S. (2016). Conceptual framework and methodology of this report. In T. Nilsen, & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 97–113). Springer International Publishing.
Nootens, P., Morin, M.-F., Alamargot, D., Goncalves, C., Venet, M., & Labrecque, A.-M. (2019). Differences in Attitudes Toward Reading: A Survey of Pupils in Grades 5 to 8. *Frontiers in Psychology, 9*(2773), 9–13.
Nortvedt, G. A., Gustafsson, J.-E., & Lehre, A.-C. W. G. (2016). The importance of instructional quality for the relation between achievement in reading and mathematics. In T. Nilsen, & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 97–113). Springer International Publishing.
OECD (2019). PISA 2018 results volume II. Where all students can succeed. OECD Publishing.
Pohl, S., Haberkorn, K., Hardt, K., & Wiegand, E. (2012). NEPS technical report for reading – scaling results of starting cohort 3 in fifth grade (NEPS working paper no. 15). Otto-Friedrich-Universität, Nationales Bildungspanel.
Praetorius, A. K., Lauermann, F., Klassen, R. M., Dickhäuser, O., Janke, S., & Dresel, M. (2017). Longitudinal relations between teaching-related motivations and student-reported teaching quality. *Teaching and Teacher Education, 65*, 241–254.
Praetorius, A.-K., Herbert, B., Decristan, J., & Köhler, C. (2020). Lernergebnisse und unterrichtliche Wirkungen [Learning outcomes and instructional impacts]. In J. Grünkorn, K. Eckhard, A.-K. Praetorius, & P. Schreyer (Eds.), *Mathematikunterricht im internationalen Vergleich. Ergebnisse aus der TALIS-Videostudie Deutschland*. DIPF.
Raudenbush, S. W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher, 34*(5), 25–31.
Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197–230.
Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. Wiley.
Scheerens, J. (2000). Improving school effectiveness. UNESCO.
Scheerens, J., & Bosker, R.J. (1997). The foundations of educational effectiveness. Pergamon.
Scheerens, J., Luyten, J.W., Steen, R., & de Thouars, Y.C.H. (2007). Review and meta-analyses of school and teaching effectiveness. Universiteit Twente, Afdeling Onderwijsorganisatie en -management.
Scherer, R., & Nilsen, T. (2016). The relationships among school climate, instructional quality, and achievement motivation in mathematics. In T. Nilsen, & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 51–80). Springer International Publishing.
Schweder, S., & Raufelder, D. (2020). Motivationale Beziehungen, fachspezifisches Interesse und Lernstrategien bei Mädchen und Jungen im Deutschunterricht [Motivational relationships, domain-specific interest and learning strategies among girls and boys in German lesson]. Zeitschrift für Erziehungswissenschaft, Special Issue "Motivation in unterrichtlichen Lehr-Lernkontexten" Advance online publication.
Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499.
Seidel, T., Rimmele, R., & Prenzel, M. (2003). Opportunities for learning motivation in classroom discourse—Combination of video analysis and student questionnaires. *Unterrichtswissenschaft, 31*(2), 142–165.
Senden, B., Nilsen, T., & Blömeke, S. (2021). Instructional Quality: A Review of Conceptualizations, Measurement Approaches, and Research Findings. I Klette, Kirsti; Blikstad-Balas, Marte & Tengberg, Michael (Red.), Ways of Analyzing Teaching Quality.Potentials and Pitfalls. Universitetsforlaget. ISSN 978–82-15–04506-1. s. 140–172. doi: doi.org/10.18261/9788215045054–2021-05.

Senden, B., Nilsen, T., & Teig, N. (2023). The validity of student ratings of teaching quality: Factorial structure, comparability, and the relation to achievement. *Studies in Educational Evaluation, 78*, Article 101274. https://doi.org/10.1016/j.stueduc.2023.101274

Solheim, O. J., Frijters, J. C., Lundetrae, K., & Uppstad, P. H. (2018). Effectiveness of an early reading intervention in a semi-transparent orthography: A group randomised controlled trial. *Learning and Instruction, 58*, 65–79.

Spoerer, N., & Schuenemann, N. (2014). Improvements of self-regulation procedures for fifth graders' reading competence: Analyzing effects on reading comprehension, reading strategy performance, and motivation for reading. *Learning & Instruction, 33*, 147–157.

Steinhauer, H.W., & Zinn, S. (2016). NEPS technical report for weighting: Weighting the sample of starting cohort 3 of the National Educational Panel Study (Waves 1 to 5). LIfBi, NEPS.

Tunnell, M. O., Calder, J. E., Justen, J. E., & Phaup, E. S. (1991). Attitudes of young readers. *Reading Improvement, 28*(4), 237–243.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Vansteelandt, I., Mol, S. E., Vanderlinde, R., Lerkkanen, M.-K., & Van Keer, H. (2020). In pursuit of beginning teachers' competence in promoting reading motivation: A mixed-methods study into the impact of a continuing professional development program. *Teaching and Teacher Education, 96*, Article 103154.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249–294.

Wentzel, K. R. (2009). Students' relationships with teachers as motivational contexts. In K. R. Wenzel, & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (pp. 301–322). Taylor & Francis Group.

Zhang, J., Li, M., Yee, D., Park, B.J., Bohrnstedt, G., & Broer, M. (2020). Reading motivation, reading achievement, and reading achievement gaps: Evidence from the NAEP 2015 Reading Assessment. American Institutes for Research.