Leino *et al.*
Large-scale Assessments in Education (2022) 10:25
https://doi.org/10.1186/s40536-022-00146-4

Large-scale Assessments in Education

RESEARCH Open Access

Associations between teacher quality, instructional quality and student reading outcomes in Nordic PIRLS 2016 data

Kaisa Leino*, Kari Nissinen and Marjo Sirén

*Correspondence:
kaisa.j.leino@jyu.fi

University of Jyväskylä, Finnish Institute for Educational Research, P.O. Box 35, 40014 Jyväskylä, Finland

Abstract

Progress in International Reading Literacy Study (PIRLS) focuses on the reading proficiency of students mostly in the fourth year of schooling. A wide selection of studies has shown that family background and early literacy activities at home have substantial associations with student achievement in reading literacy. However, research focusing on teacher qualities and teaching processes is inadequate. In this study, we focus on associations of teacher quality (formal qualifications and professional identity) and instructional quality (classroom management, cognitive activation and teacher support) with cognitive and affective-motivational student outcomes (variables Reading Achievement, Students Confident in Reading, and Students Like Reading). We analyzed PIRLS 2016 data from four Nordic countries (Denmark, Finland, Norway, and Sweden), consisting altogether of 923 teachers and 17,161 students. Using path analysis, we considered selected background variables from teacher and student questionnaires in relation to the outcomes. Overall, the associations of student outcomes with teacher quality and instruction quality were found to be weak in all the countries, and there was little variation between the countries. The strongest association observed in all countries was the positive relation between Teacher Support Perceived by Students and Students Like Reading. Further, a positive Working Atmosphere in the Classroom tended to promote Reading Achievement and Students Confident in Reading. Teacher's Specialization in reading and the language of the test was positively associated with Teacher's Self-Efficacy in teaching reading, which in turn was related to measures of instructional quality. The implications for practice are discussed.

Keywords: Reading achievement, Reading attitude, Teacher quality, Instructional quality, PIRLS study

Introduction

The International Association for the Evaluation of Educational Achievement (IEA) conducts international comparative studies in education to enhance knowledge about education systems and the achievement of students. Progress in International Reading Literacy Study (PIRLS) has assessed the reading literacy achievement of students in their fourth year of schooling every five years since 2001. In addition to students' reading achievement data, PIRLS also gathers background information about

students' home environment and support for literacy, as well as about instruction in classrooms, using questionnaires directed to parents, teachers and principals. Sweden and Norway have participated since 2001, Denmark since 2006 and Finland since 2011. In this study we use the data collected in 2016, when data were collected from fourth graders in Denmark, Finland and Sweden. In Norway the target population was fifth graders. However, due to differences in the schooling systems, students in all four countries were similar in age (average age 10.7–10.8 years). In 2016, among 50 participating countries, Finnish students were ranked 5th, Norwegian students were 8th, Swedish students were 12th, and Danish students were 18th (Mullis et al., 2017, Student Achievement section).

Educational systems in Nordic countries are seen to be very similar (see, e.g., Volmari, 2019) and are often referred to collectively as the "Nordic Model," although individual differences have existed and are increasing, even to the extent that some have questioned the whole existence of a common Nordic model (Frønes et al., 2020). The common goals and content of teaching reading are similar, yet such important issue as the level of teacher education differs. Reading is strongly embedded in the teaching of mother tongue or the language of the school (e.g., in Finland the subject called "mother tongue and literature" or "Swedish" in Sweden), but the subject has other content areas also. In addition, teaching reading is not the responsibility of just one teacher or limited to one school subject. Reading is an important way to acquire information in other subjects also, and therefore teaching reading is a cross-curricular activity in all the Nordic countries. Different levels of success in international comparative student assessments have yet again raised questions about whether the Nordic model is so uniform after all and what are the factors that influence students' equal opportunity to learn.

Studies have found many factors affecting students' reading skills and attitudes. Typically, the supportiveness of the home environment (Gustafsson et al., 2013; Hemmerechts et al., 2017; Mullis et al., 2017, Home Environment Support section), socioeconomic background (Hemmerechts et al., 2017; Neff, 2015; Sirin, 2005; Støle et al., 2020), and engagement in reading (Ho & Lau, 2018; Wantchekon & Kim, 2019) are found to be important explanators of reading proficiency. Gender issues also arise regularly in literacy studies since gender difference in large-scale studies has for years favored girls (Gustafsson et al., 2013; OECD, 2019b, 141–149). However, Gustafsson et al. (2013) noted a stronger emphasis on literacy activities at home for girls than for boys, which may explain some of the differences.

Evidence concerning the relationship of teacher quality and reading outcomes is inconsistent or even missing in many large-scale studies (see also Nortvedt et al., 2016; Van Staden & Zimmerman, 2017). It has been suggested that not only teaching practices, but also teacher qualifications, the classroom atmosphere, and time spent learning have a link to student outcomes (e.g., Creemers & Kyriakides, 2008). For example, the high educational level of teachers has been noted as one of the key factors in the success of Finnish students in different international assessments (e.g., Crouch, 2015). Research on teacher quality has more often focused on, for example, student outcomes of mathematics (e.g., Blömeke et al., 2016) or science (e.g., Nilsen et al., 2018) than on reading. Seidel and Shavelson (2007) noted in their large meta-analysis that especially domain-specific learning activities and teachers' knowledge of them affected the cognitive aspect

of learning and also have an important role in promoting motivational outcomes, such as interest or self-efficacy.

Teachers' actions and activities in the classroom derive from their formal education as well as from their beliefs and even individual characteristics. Goe (2007, 8–9) has presented a framework where Teacher Quality is a combination of two inputs and classroom activities. The inputs consist of teacher qualifications (e.g., educational level, specialization, and participation in professional development) and characteristics that influence teachers' instruction (e.g., attitudes, self-efficacy). Whereas teacher qualifications form through formal or informal education, characteristics are rather related to personal character traits and the view of oneself as a teacher. In this study, we use the concept of professional identity (Canrinus et al., 2012), which refers to teachers' attitudes and beliefs about themselves as a teacher. Teachers reflect on their own teaching through their perception of skills and job satisfaction by mirroring their teaching to other teachers. It consists of beliefs, self-efficacy, and collaboration skills, which also form the basis for mastery of specific areas (Epstein & Hundert, 2002; Goe, 2007; Kunter et al., 2013; Nilsen et al., 2018).

Classroom practices form teaching quality, which includes, for example, planning, instructional delivery, classroom management and interactions with students (Goe, 2007, 8–9). Pedagogical content knowledge (Shulman, 1986) can be considered a basis for teachers' actions in the classroom. In this study, we use the concept of instructional quality (e.g., Nilsen & Gustafsson, 2016), which refers to a teacher's knowledge and the quality of instructional practices that are put into use in the classroom, such as activating and supporting students and managing the working environment. Klieme et al. (2009), in their model of quality of instruction, divided teachers' practices into (a) cognitive action and deep content, (b) classroom management, clarity and structure, and (c) a supportive climate.

A teacher's instructional quality is of great significance in forming a well-functioning teacher-student relationship. According to Hattie's (2009) meta-analysis, a constructive teacher-student relationship is more important for a student's school success than the student's socioeconomic background. Instructional quality is a key when a teacher aims to engage students in the learning at hand and activate their cognitive processes (Klieme et al., 2009). However, Goe (2007) has emphasized that teachers' effectiveness cannot be measured only with dimensions related to teacher without considering student outcomes. In a study focusing on fourth-graders' science skills in the TIMSS 2015 assessment, Nilsen et al. (2018) found that instructional quality correlated positively with science achievement, but not with a student's intrinsic motivation.

Based on previous studies, we form a theoretical structure with three top concepts—teacher quality, instructional quality, and student outcomes. We employ path analysis in studying the relations of teacher quality and instructional quality to three student outcomes: students' reading achievement in the PIRLS 2016 test, the level of students' confidence in reading and how much students like to read. We use teacher and student data collected in four Nordic countries (Denmark, Finland, Norway and Sweden) in the PIRLS 2016 assessment. Our aim is to examine which characteristics of teacher quality and instructional quality promote students' proficiency in and attitudes towards reading to gain a better understanding of the qualities and skills needed in successful teaching.

We analyze the data sets of each country separately. Our purpose is not to perform an in-depth analysis of between-country differences, but rather obtain an overview of which variables are associated with the reading-related student outcomes in these four countries.

Next, we define the main concepts and examine previous studies. The focus is on those factors that were measured in PIRLS. In that basis, we examine the theoretical background addressing two subdimensions of teacher quality—namely formal qualifications and professional identity—and three subdimensions of instructional quality—namely classroom management, cognitive activation and teacher support (see also Fig. 1). A presentation of our research questions as well as the variables and methods follows.

Theoretical background and previous studies

Teacher quality

Formal qualifications

Formal qualifications refer to education received by the individual, including level of formal education and its content as well as in-service professional development activities. There are studies that focus on different subjects which state that more qualified teachers correlate with—though are not a guarantee for—better student achievements and provide students with more equal opportunities for success regardless of their socioeconomic statuses, race/ethnicity, or other individual backgrounds (e.g., Akiba et al., 2007; Clotfelter et al., 2010; Darling-Hammond, 2000). In these studies, the quality of the teacher is usually measured by teachers' formal education level, such as a degree or certificate.

Teachers in Nordic countries usually have an extensive *formal education*. According to the PIRLS 2016 study (Mullis et al., 2017, Teachers' and Principals' Preparation section, Exhibit 8.1), the majority of the participating teachers in Nordic countries were qualified through a teacher education program at a university or teacher college. According to the PIRLS 2016 data, Finland stands out in level of education: up to 92% of Finnish fourth-grade teachers had a postgraduate university degree, that is, a master's or doctoral degree or equivalent, whereas in Norway the figure was 22%, in Sweden 13% and in Denmark 4%. A bachelor's degree or equivalent was the highest degree for 81% of

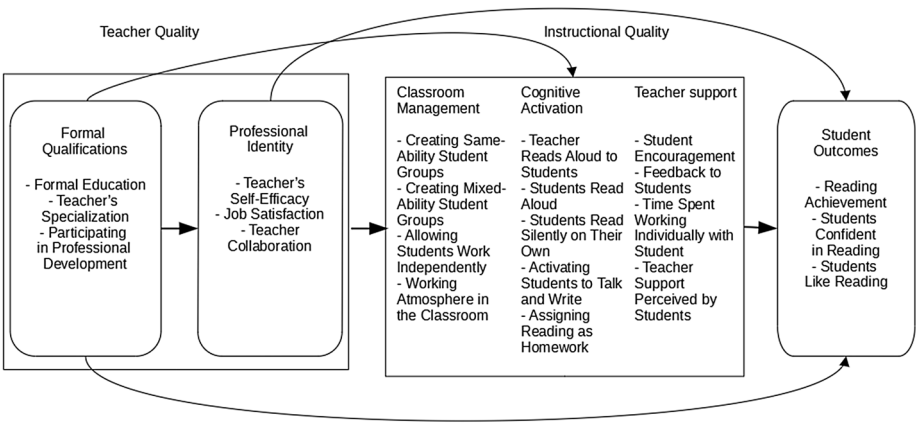

Fig. 1 General structure of path analysis models

teachers in Sweden, for 79% in Denmark, for 73% in Norway, and for 6% in Finland. However, starting from 2017, teacher education in Norway has also been arranged in a five-year master's degree program (Gabrielsen, 2017).

In the Nordic countries, reading is taught within all subjects throughout primary education. According to the PIRLS 2016 study (Mullis et al., 2017, Teachers' and Principals' Preparation section, Exhibit 8.2), *specialization in reading* included in teachers' formal education varied greatly between Nordic countries: Sweden had the highest number of students whose teacher had specialized in the language of the test (82%) and Denmark outnumbered others in pedagogy of reading (57%) and reading theory (42%). Finland had the least specialization in these three topics. Overall, the PIRLS 2016 results (Mullis et al., 2017, Teachers' and Principals' Preparation section, Exhibit 8.2) showed no relationship between specialization in the language of the test and students' average reading achievement. The same applied for specialization in pedagogy or reading theory.

According to the PIRLS 2016 study (Mullis et al., 2017, Teachers' and Principals' Preparation section, Exhibit 8.4), *participation in professional development* also varied. In Norway 64% of students had teachers who had participated in at least six hours of professional development during the two years preceding the survey. The corresponding figure was 62% in Sweden, 43% in Denmark, and only 17% in Finland.

Blömeke et al. (2016) focused on mathematics achievement, which is not directly comparable to learning to read. They found that the teacher's level of education showed significant positive relations to instructional quality and student achievement in mathematics in several countries, but student achievement was not well-predicted by instructional quality. In addition, participation in professional development was one of the strongest predictors of instructional quality across all 47 countries studied, including in the Nordic countries. Correspondingly, the ISCED level of teacher education was, on average, the strongest predictor of student achievement across all countries.

Professional identity

Professional identity refers to personal characteristics, such as teachers' attitudes and beliefs towards their own skills, workplace and profession, the opportunity to influence student learning, and learning pedagogy in general. In the PIRLS 2016 context, a teacher's professional identity can be studied through self-efficacy, job satisfaction and collaboration.

Teachers' high *self-efficacy*—indicating a high level of confidence in possessing the knowledge and skills needed in successful teaching (e.g., Bong, 2006)—has been shown to have significant and positive relations with instructional quality and student achievement (Nilsen et al., 2018). Furthermore, teachers with high self-efficacy are able to create and promote surroundings that enhance job satisfaction (Caprara et al., 2006).

Evans (1997) defines *job satisfaction* as "a state of mind determined by the extent to which the individual perceives her/his job-related needs to be being met" (p. 328). Furthermore, job satisfaction consists of two main components: job comfort and job fulfilment. The former refers to how satisfactory conditions and circumstances are to an individual, and the latter refers to self-assessment of personal accomplishments within meaningful aspects of the job (Evans, 1997). Banerjee et al. (2017) found that teacher job satisfaction has a modest but significant and direct association with

students' reading achievement (see also Caprara et al., 2006). According to Dicke et al. (2020), different aspects of teacher quality are interdependent, showing both direct and indirect associations with teacher job satisfaction and student outcomes.

Teacher collaboration is often connected to school activities such as professional development within school, but it is also related to job satisfaction and belonging to the work community. According to TALIS results (OECD, 2019c, cp. 4), teachers who took part in the interdependent forms of collaboration reported high job satisfaction and self-efficacy levels and they used cognitive activation practices in teaching more frequently than other teachers did. Despite the unclarity of causality, some studies (e.g., Fuglestad et al., 2017; Goddard & Tschannen-Moran, 2007; Ronfeldt et al., 2015) have found clear associations between higher marks in mathematics and reading and teacher collaboration for improvement in issues such as curriculum, instruction, and professional development. Nilsen et al. (2018), studying the Nordic countries, found an improved level of instructional quality (self-reported) among science teachers who collaborate more often than other teachers do. In addition, their findings included better results in student achievement and greater student motivation to learn science.

Instructional quality

Classroom management

Classroom management refers to organizing the students' work in the classroom and disciplinary practices, such as reducing distractions (Doyle, 1986). Studies have shown that student-centered classrooms, where teachers and students engage in interpersonal interaction, facilitate high student achievement and positive learning environments (Freiberg, 2013; Freiberg et al., 2009). Korpershoek et al. (2016) found that effective classroom management creates a positive *learning atmosphere* and significantly increases students' academic achievement and decreases behavioral problems. Instead, feelings of fear or being bullied at school usually result in lower achievement (e.g., Milam et al., 2010; Ponzo, 2012).

One way of classroom management is organizing and *grouping students* in the class. Social interaction about reading may help students see different views (Almasi & Garas-York, 2009), argument their understanding (Unrau, 1992), and understand the social aspect of reading (Alvermann & Moje, 2013). Working with other students promotes student engagement and is related to better achievement (Hattie, 2009). Grouping can be done randomly or by students' abilities, which is not, however, entirely unambiguous. Lleras and Rangel (2009), among others, found in their study of elementary students that low performing students may experience homogeneous groupings negatively while more skilled students may benefit from working with students of similar level. However, struggling readers benefit from instruction given individually, in pairs, or in very small groups instead of whole class instruction (Elbaum et al., 1999; Hattie, 2009; Vaughn et al., 2003).

In teaching reading, teachers can also decide whether to use teacher-led read-aloud or independent reading in the classroom. *Independent reading* in school reinforces out-of-school reading and with intentional instruction it consolidates and helps to take ownership of reading skills and strategies (ILA, 2018).

Cognitive activation

To enhance learning, teachers use cognitive activation, that is, instructional approaches and learning tasks that aim to help students learn different kinds of strategies so they can analyze, evaluate and create information (e.g., Klieme et al., 2009). In the context of reading skills, cognitive activation includes tasks for *students to read aloud or in silent* and content-related tasks at lessons or at home. Students reading aloud by themselves enhances their memory (e.g., Lafleur & Boucher, 2015). Robinson et al. (2018) found that among students with reading disabilities, oral reading facilitated higher reading comprehension than silent reading did. Reutzel et al. (2008), however, found that scaffolded silent reading (ScSR) improved third-grade students' fluency and comprehension as effectively as guided repeated oral reading (GROR) among students with no reading difficulties.

Sometimes teachers also read to students. *A teacher's reading aloud* is a normal part of elementary school work in many countries and teachers in the primary grades frequently read to their students, but in the upper grades, teachers do not necessarily read aloud to students despite the numerous benefits it has even for older students (Ariail & Albright, 2005; ILA, 2018; Jacobs et al., 2000). Reading aloud to a child has a positive effect on children's language development and vocabulary, especially when involving versatile texts and combined with various activities that support learning, such as naming the objects and things in the book and using the words learned in other situations (e.g., Lane & Wright, 2007; Wasik & Bond, 2001). It is also connected to increased enthusiasm for reading and willingness to read later in life, as well as higher academic achievements (see e.g., Ariail & Albright, 2005; Lerkkanen et al., 2018; Torppa et al., 2022). According to Hurst and Griffity (2015), reading to students models fluent reading and provides opportunities for discussion and hence hearing the text read aloud most benefits the less fluent readers.

Regardless of who has read the text, effective instruction includes tasks that support learning. The main comprehension processes of reading, which are also evaluated in the PIRLS study (Mullis & Martin, 2015), include finding and using information, making inferences, interpreting and integrating ideas and information as well as evaluating the content and textual elements (see also OECD, 2019a, Ch. 2). To reinforce these processes, teachers can create a discussion of what has been read by selecting appropriate questions (Morgan & Meier, 2008). Teachers activate students' learning by *asking them to do tasks* such as thinking about their prior knowledge, connecting information to their own lives, comparing texts with similar content or summarizing the content. These cognitive activations before, during and after reading help students to gain skills and metacognitive strategies to better master the comprehension processes (e.g., Baker & Beall, 2009). This may benefit student outcomes in many ways. Huang and Chen (2018), using PIRLS 2011 data from Hong Kong, showed that the frequency of reading strategy instruction was significantly related to student attitudes toward reading and motivation to read, and student attitudes toward reading were significantly associated with reading achievement. Berge et al. (2017) found that Norwegian teachers guided fourth-grade students to use reading strategies only monthly, which was considered to be too seldom, and, in addition, they guided more often to use less effective metacognitive strategies than highly effective in-depth strategies.

Cognitive activation may extend beyond the classroom. *Giving homework* for students is a common pedagogical practice to reinforce daily learning and foster study skills (Bempechat, 2004). There is some evidence that homework completion relates positively to academic achievement (Cooper et al., 1998). Bempechat (2004), however, emphasizes that the value of homework cannot be measured only with test grades because the type and the aim of the homework varies as well as the quality and amount of support for homework given by parents at home. For example, giving reading homework at school is not limited only to reading lessons and the process of the reading is different for increasing fluency compared to reading to learn. Hattie's (2009) analysis refers to the finding that homework may be more beneficial in older age and for already better students. However, this may correlate to student's socioeconomic background and parental involvement, which may give an advantage to some students in the form of home support, as Bempechat (2004) suggests, adding that older students may also have better learning skills to perform tasks by themselves without a teacher's support.

Teacher support

The third important form of instructional quality is teacher support for students. Hamre, Pianta and their colleagues (Hamre & Pianta, 2007; Hamre et al., 2007) define three different kinds of support in the classroom: emotional, organizational, and instructional. Emotional support influences classroom climate and how the student feels that he or she is being encountered as an individual (see also Federici & Skaalvik, 2014). Emotional support also refers to the degree to which the *teacher encourages* and conveys confidence in students' abilities (Strati et al., 2017). Organizational support, as mentioned earlier, refers to how a teacher is able to maintain peaceful working conditions and make activities progress smoothly. Instructional support includes teachers' efforts to *help students with the task at hand* and to develop their learning, higher-order thinking skills, and working skills. Teacher support has been linked to better learning engagement and outcomes (Curby et al., 2013; Jensen et al., 2019; Klem & Connell, 2004). According to Curby et al. (2013), different kinds of support also correlate: emotional support given in an early stage predicted higher instructional support later on, and vice versa. Relations may not always be as straightforward. A study among Norwegian first-graders (Jensen et al., 2019) showed significant positive relations between teachers' emotional support and students' self-concept, which then mediated the effect on reading achievement. In reading development, teacher support is essential in facilitating students' reading development through strategy instruction and engaging them in independent and collaborative reading activities (Ho & Lau, 2018; Housand & Reis, 2008).

As previously presented, a teacher's effectiveness is a sum of many instructional factors. However, some commonalities can be highlighted. Stronge et al. (2007) found that effective teachers demonstrated a higher degree of fairness toward students and that they understood the need to alter the lesson materials in order to reach different kinds and levels of learners. They noticed that even though effective as well as less effective teachers asked the same number of lower-level questions, the effective teachers asked approximately seven times more higher-level questions (i.e., application, analysis, synthesis, evaluation). In addition, effective teachers had much less

disruptive behavior of students. Finally, Yair (2000) has proposed that effective teachers also motivate students by connecting the task at hand to reality, which emphasizes the relevance of the task.

Research questions

The aim of our study was to examine the relations of teachers' education, attitudes and classroom activities with students' reading-related outcomes in four Nordic countries, using the variables available in PIRLS 2016 data. Our research questions (RQ) derived from findings reported in recent literature on associations between student outcomes and characteristics of teachers and instruction, described above. Accordingly, the variables to be considered were selected from the available data on the basis of findings reported in literature. The interrelations between teacher quality and instructional quality, and their associations with student outcomes, were analyzed through path models. Even though we analyzed each country separately, we did not include the possible between-country differences in the research questions, as our central target was to investigate variables that may affect student outcomes in the context of Nordic countries.

The general structure of the model, which was the starting point of our analysis, is illustrated in Fig. 1. According to the suppositions of the path model, the associations are directed in the sense that there are dependent variables whose variance is explained by independent variables. From this viewpoint we can also call associations 'effects', although the cross-sectional nature of data does not allow strictly causal inferences. In the model, we hypothesize causality between the subdimensions of teacher quality, that is, we assume that formal qualifications can affect professional identity. Regarding instructional quality, we do not hypothesize such causality. Therefore, the subdimensions of teacher quality and instructional quality are treated differently in the model.

We hypothesize that teachers' formal qualifications are associated with their professional identity, which in turn associates with instructional quality, and this finally associates with student outcomes (Fig. 1). But, in addition, we allow for the possibility that formal qualifications affect directly both instructional quality and student outcomes, and professional identity affects directly student outcomes. Student outcomes may thus depend on teacher characteristics both directly and indirectly. However, within this conceptual upper level path structure, our approach is exploratory at the level of variables. This means that we do not explicitly hypothesize any associations between the observed variables which measure teacher quality and instructional quality or are student outcomes. Instead, we let the empirical data suggest through significance tests which associations are relevant and which are not.

The research questions are as follows:

RQ 1. How are teachers' formal qualifications associated with professional identity? Are formal qualifications directly associated with instructional quality, and reading-related outcomes of ten-year-old students?

RQ 2. How is professional identity associated with instructional quality? Is instructional quality directly associated with reading-related outcomes of fourth (fifth) graders?

RQ 3. How is instructional quality associated with reading-related outcomes of fourth (fifth) graders?

Methods

Data

We employed the PIRLS 2016 assessment data from four Nordic countries: Denmark, Finland, Norway, and Sweden. In each country, the PIRLS data are collected with a stratified cluster sampling design, where one to four classrooms are drawn from each sampled school, and principally every student in the classroom is tested. The hierarchical nature of the data was taken into account in all statistical analyses. Background data concerning students and their families as well as teachers were collected with student and teacher questionnaires.

We merged the national PIRLS 2016 teacher questionnaire data set with the respective student-level data sets. In total, the Danish data consisted of 186 teachers and 3,508 students, while the Finnish data consisted of 295 teachers and 4,896 students, the Norwegian data consisted of 215 teachers and 4,232 students, and the Swedish data consisted of 227 teachers and 4,525 students. In the data sets there was a perfect one-to-one match of teachers and classrooms so that each student had precisely one teacher in the data. In the Danish data, on average there were 3,508/186 = 18.9 students per teacher. In Finland, the respective ratio was 16.6, in Norway 19.7, and in Sweden 19.9 students per teacher.

The students were fourth-grade students in all countries except for in Norway, where they were in fifth grade. The average age of the students was 10.7 years in Sweden and 10.8 years in all the other Nordic countries (Mullis et al., 2017, About PIRLS 2016 section). Next, we describe the different variables used in the study.

Student outcome variables

Student outcomes included three different variables: Reading Achievement, Students Confident in Reading and Students Like Reading. *Reading Achievement* in PIRLS consists of two major purposes of reading (Mullis & Martin, 2015): reading for literary experience, and to acquire and use information. In both purposes the processes of comprehension are the following: (a) focus on and retrieve explicitly stated information, (b) make straightforward inferences, (c) interpret and integrate ideas and information, and (d) evaluate and critique content and textual elements. In measuring reading achievement in this study, we utilized five plausible values (variables ASRREA01–ASRREA05) available in the PIRLS student data. Plausible values are estimates of the latent proficiency of a student, based on the student's success in the PIRLS reading literacy test and conditioned on background information (Foy & Yin, 2017). The estimated test reliability of the PIRLS 2016 reading assessment was 0.88 for Denmark, 0.88 for Finland, 0.87 for Norway, and 0.88 for Sweden, with the international median reliability being 0.89 (Martin et al., 2017, Exhibit 10.7, p. 10.15).

To examine whether students are confident in reading and whether students like reading, we employed two scale variables (ASBGSCR = Students Confident in Reading scale, and ASBGSLR = Students Like Reading scale) derived from student questionnaire statements by the PIRLS International Study Center, using IRT methodology. The *Students Confident in Reading* scale was based on six statements (Martin et al., 2017, Appendix 14A, p. 14.83), and its reliability was 0.83 in Denmark, 0.80 in Finland, 0.82 in Norway, and 0.82 in Sweden. The *Students Like Reading* scale was based on ten statements

(Martin et al., 2017, Appendix 14A, p. 14.91), and its reliability was 0.85 in Denmark, 0.89 in Finland, 0.87 in Norway, and 0.88 in Sweden.

Controlling variables

Earlier studies have shown that students' reading achievement tends to be correlated with socioeconomic status, gender, and whether the test is taken in the student's home language (e.g., Mullis et al., 2017; OECD, 2019b). It is thus possible that such background variables may intervene in the relation of teacher and instruction to the student outcomes: outcomes may be better if the student has favorable background characteristics, independent of the quality of teacher and instruction. We wanted to reduce the risk of distorted analysis by controlling for three student questionnaire variables typically associated with reading achievement: Male Gender of student (derived from variable ITSEX in the student data set), Home Language, meaning the frequency the students speak the language of test at home (variable ASBG03 in the student data set, scale inverted), and Number of Books at Home (variable ASBG04 in the student data set) which is typically included when measuring socioeconomic status of the family. We did not use data collected from parents due to the large amount of missing data in some countries.

Variables measuring teacher quality

We originally considered over 20 background variables or indices obtained from the PIRLS teacher questionnaire. The variables representing teacher quality were grouped into two subdimensions, namely, teacher's formal qualifications and professional identity.

The variables measuring *formal qualifications* were the following: highest level of completed Formal Education (variable ATBG04 in the teacher data set, measured on the ISCED scale), a sum index measuring Teacher's Specialization in reading pedagogy and the language of the test as a part of formal education, constructed from six items (variables ATBG05BA, ATBG05BB, ATBG05BC, ATBG05BE, ATBG05BF, ATBG05BI) in the teacher questionnaire, and Participating in Professional Development that related to teaching reading in the past 2 years (variable ATBG06 in the teacher data set).

Professional identity was represented by three variables: Teacher Collaboration, Job Satisfaction, and Teacher's Self-Efficacy, which were factor scores constructed from statements in the teacher questionnaire. Teacher Collaboration was formed from five statements (variables ATBG09A–ATBG09E), dealing with several types of interactions with other teachers. The reliability of this score was 0.74 for Denmark, 0.78 for Finland, 0.76 for Norway, and 0.80 for Sweden. Similarly, Job Satisfaction was formed from five statements (variables ATBG10A–ATBG10E). The reliability of this score was 0.91 for Denmark, 0.93 for Finland, 0.91 for Norway, and 0.89 for Sweden. Teacher's Self-Efficacy was formed from six statements in the teacher questionnaire. These statements were a national option in the Nordic countries (i.e., they were not used in all PIRLS countries, and their variable names also vary between the national data sets). They concerned how confident teachers felt in performing various tasks related to teaching reading to their classes. The reliability of this measure was 0.83 for Denmark, 0.86 for Finland, 0.77 for Norway, and 0.82 for Sweden.

Variables measuring instructional quality

The instructional quality was considered through three subdimensions: classroom management, cognitive activation, and teacher support. Variables obtained from the background questionnaires were grouped accordingly. *Classroom management* was examined with three statements in the teacher questionnaire and one index variable extracted from statements in the student questionnaire. The teacher questionnaire statements concerned creating same-ability student groups when teaching reading (variable ATBR08B), creating mixed-ability student groups when teaching reading (variable ATBR08C), and allowing students to work independently (variable ATBR08E). From the student questionnaire we employed six statements in creating an index of Working Atmosphere in the Classroom, perceived by students. These statements also were a national option in the Nordic countries, and they concerned, for instance, the frequency of noise and disorder in the classroom. Again, the index was a factor score. The reliability of this measure was 0.85 for Denmark, 0.82 for Finland, 0.84 for Norway, and 0.84 for Sweden.

To evaluate the *cognitive activation* of students, we considered several items in the teacher questionnaire. First, we looked at a set of items dealing with reading in the classroom: how often the teacher reads aloud to students (variable ATBR10A), how often the teacher asks students to read aloud (variable ATBR10B), and how often the teacher asks students to read silently on their own (variable ATBR10C). Then, we considered an index of Activating Students to talk or write about what they had read. This was a factor score formed from three statements (variables ATBR13A–ATBR13C). The reliability of this index was 0.66 for Denmark, 0.69 for Finland, 0.82 for Norway, and 0.83 for Sweden. The appearance of rather low reliabilities (Denmark, Finland) is not that unexpected, given the small number of items in the index. Finally, we used the question asking how often the teacher assigns reading as homework (variable ATBR17).

Teacher support was examined through five variables. From the teacher questionnaire, we first created an index, which we call Student Encouragement. This was again a factor score extracted from five questionnaire items, mainly dealing with encouraging students in various activities related to reading (variables ATBR11C–ATBR11G). The reliability of this index was 0.85 for Denmark, 0.84 for Finland, 0.85 for Norway, and 0.80 for Sweden. Next, we employed two questions concerning the frequency of feedback to students (variables ATBR11I and ATBR19A), and time spent working individually with a student who struggles with reading (variable ATBR21C). Finally, we employed an index of Teacher Support Perceived by Students. This was a factor score derived from four statements in the student questionnaire (variables ASBR01F–ASBR01I), concerning encouragement and advice received from the teacher. The reliability of this index was 0.78 for Denmark, 0.79 for Finland, 0.74 for Norway, and 0.77 for Sweden.

Analysis

We examined the relations of teacher quality (in terms of formal qualifications and professional identity), instructional quality (in terms of classroom management, cognitive activation and teacher support), and student outcomes (Reading Achievement, Students Confident in Reading and Students Like Reading), illustrated in Fig. 1 above, by path analysis. We performed the path analyses under the framework of structural equation modelling (e.g., Bowen & Guo, 2012; Maruyama, 1998), and each country was

considered separately. As the latent factor structures were not of our interest, we did not consider any measurement models in our analysis in order to keep the models reasonably simple. All variables in the analysis were therefore regarded as manifest variables. We first employed exploratory factor analyses to create scales and index variables to be used in the path models and estimate their reliabilities. It is worth mentioning that the PIRLS data already contain well validated (manifest) sum indices created by the consortium, such as the Students Confident in Reading Scale and the Students Like Reading Scale. Regarding the aim of our study, re-evaluating the validity of these indices by introducing latent structures in our models would not bring added value.

We were only interested in correlational relations of the chosen variables (in terms of regression coefficients); we did not aim to examine any mean structures or estimate population variances. Therefore, we used the correlation matrix of the variables as the input data, meaning that we treated all variables as (nationally) standardized. In calculating the correlations, listwise deletion of missing data was adopted.

The structural equation models, including path models, are usually applied for confirmatory purposes. However, the number of variables representing teacher quality and instructional quality, which are potentially relevant in explaining the variance of student outcomes or other variables, was large, and, additionally, the variables tend to be correlated with each other. For this reason, we first conducted preliminary path analyses in an exploratory sense to sieve out variables which did not show any explanatory power and/or suffered from collinearity problems. These analyses were carried out separately for each country, and all the variables were included in these analyses together. We carefully examined the results of the preliminary analyses and consequently reduced the number of variables to be used in the eventual path modelling. Variables, which did not show statistically significant relations with other variables in any of the four countries or showed excessive overlap with other variables in the models, were dropped from the final modelling phase. We considered a relation statistically non-significant when its p value was larger than 0.10. The preliminary analysis resulted in a remarkably reduced number of variables for final analysis. The remaining variables are presented at the beginning of the Results section.

We included the controlling variables (Male Gender, Home Language, and Number of Books at Home) in all models, regardless of their statistical significance.

We performed all analyses at the student level (i.e., the unit of analysis was student), and applied student weights scaled to sum up to national sample sizes in all analyses (such weights are called house weights in the PIRLS context). The clustering of students was taken into account by introducing classroom as the clustering variable, to correct for otherwise underestimated standard errors.

As Reading Achievement was operationalized through five plausible values (Martin et al., 2017, pp. 12.14–12.15), we used them in the analyses. That is, we performed all analyses five times, only varying the plausible value included in the model, and then merged the five obtained results to produce single estimates using the multiple imputation methodology. This is the recommended approach for the analysis of large assessment data with plausible values as it adequately handles the uncertainty related to estimating latent proficiency (Khorramdel et al., 2020; Rutkowski et al., 2010; von Davier et al., 2009; Wu, 2005).

The path analyses were performed with the Mplus 7 software (Muthén & Muthén, 2015). The "complex option" for analyzing clustered survey data was used, and the chosen estimation method was MLR, which is maximum likelihood robust to non-normality and non-independent observations. The standard errors were computed using a sandwich estimator, supported by Mplus, which is an alternative to the replicate weights (jackknife) approach for complex sampling designs. The analysis of plausible values and the respective multiple imputations were performed by the imputation facility of Mplus. The model fit was assessed with the usual goodness-of-fit criteria (standardized root mean square residual SRMR, root mean square error of approximation RMSEA, comparative fit index CFI, Tucker-Lewis index TLI) of structural equation models (see, e.g., Bentler, 1995; Steiger & Lind, 1980). Exploratory factor analyses to create factor score variables and estimate their reliabilities were performed by the Factor procedure of SAS® software, using iterative principal axis factoring.

Results

Results of preliminary analysis

As mentioned above, several variables were dropped from the final analysis, based on their statistically non-significant (i.e., $p > 0.10$) performance in the preliminary analyses, or excessive collinearity with other variables. In the following, we describe the main findings of the preliminary analyses. It is informative to recognize not only variables with explanatory power, but also variables, which are not associated with other variables of interest. We do not present the details of preliminary results, because the size of the results tables would become very large.

Considering teacher quality, two variables measuring *formal qualifications* remained for the final modelling phase, namely Teacher's Specialization in reading pedagogy and language of the test and Participating in Professional Development. The highest level of completed Formal Education did not have any significant associations in the models. Of the variables measuring *professional identity,* Teacher Collaboration and Job Satisfaction were dropped; thus Teacher's Self-Efficacy was the only variable employed in final models.

Considering instructional quality, the subdimensions of classroom management and teacher support only had statistical significance. However, not all variables were relevant in them either. The only remaining variable, which represented c*lassroom management,* was Working Atmosphere in the Classroom (perceived by students), meaning that grouping students according to their ability or allowing them to work independently had no role in the paths explaining variance in student outcomes. Similarly, the only remaining variable representing *teacher support* was Teacher Support Perceived by Students. Students' view regarding teacher support thus seems most significant while teachers' responses about encouraging students, helping them individually, or frequency of feedback were not associated with outcomes. The five variables measuring c*ognitive activation* (e.g., students reading aloud, reading as homework, and Activating Students) were all dropped. These findings were uniform in all four countries.

Based on the findings of preliminary analysis, we decided to start the final modelling phase with the path model illustrated in Fig. 2. Teacher's Specialization and Participating in Professional Development explain Teacher's Self-Efficacy, which then explains

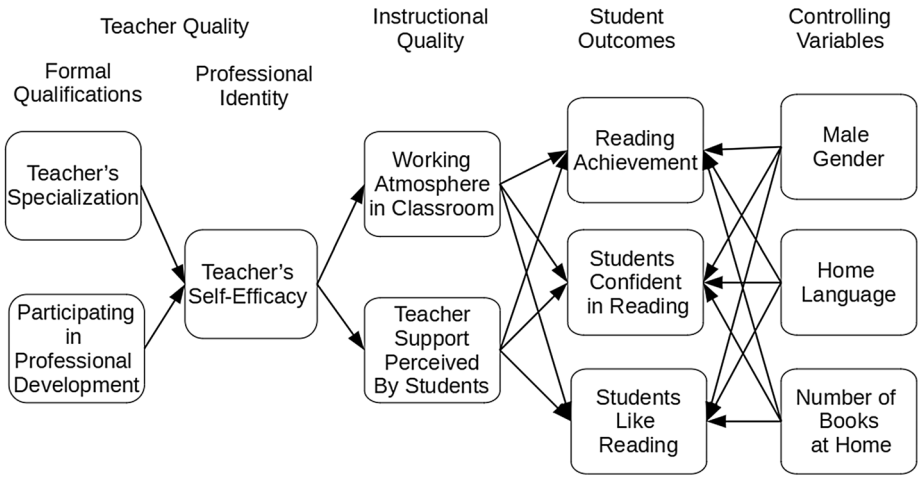

Fig. 2 Variables and paths in the path analysis model

Working Atmosphere in the Classroom and Teacher Support Perceived by Students, which eventually explain student outcomes (Reading Achievement, Students Confident in Reading and Students Like Reading). On the other hand, we let the controlling variables (Male Gender, Home Language, Number of Books at Home) have direct effects on the outcomes. Both the controlling variables and the variables of teacher quality and instructional quality were entered into the model simultaneously. The explanatory variables and error terms were allowed to be correlated when needed. During the modelling process some paths postulated in Fig. 2 appeared non-significant in some countries. We did not keep them in the final models to be reported, because we wanted to respect the principle of model parsimony (e.g., Bentler & Mooijaart, 1989), while seeking as good-fitting a model as possible for each country. In addition, Denmark appeared different from the other countries in that we had to add two extra paths (from Teacher's Specialization and Participating in Professional Development to Teacher Support Perceived by Students) to model in Fig. 2, to obtain a model of sufficient fit. Consequently, the final models of the countries were not completely similar. The country-specific results are presented in more detail in what follows. In all the figures significance levels are indicated as follows: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; +$p < 0.10$. The correlation matrices of the variables used in the final analysis are shown in appendices A–D (Additional file 1).

Main results by country

Denmark

We present the estimated path model for Denmark in Fig. 3. For clarity, we have omitted the part of the controlling variables from the figure. We present the numerical results for controlling variables in Table 1 instead. Figure 3 shows the standardized path coefficient estimates and their standard errors, as well as the non-zero error correlations among error terms. The model fit was good: SRMR = 0.03, RMSEA = 0.01, CFI = 0.98, TLI = 0.97. The number of students used in the estimation was $n = 2896$, and number of classrooms (teachers) was 176. It is worth noting that in the following models the standard errors of the relations between the teacher quality variables (Teacher's Specialization, Professional Development and Self-Efficacy) are larger than

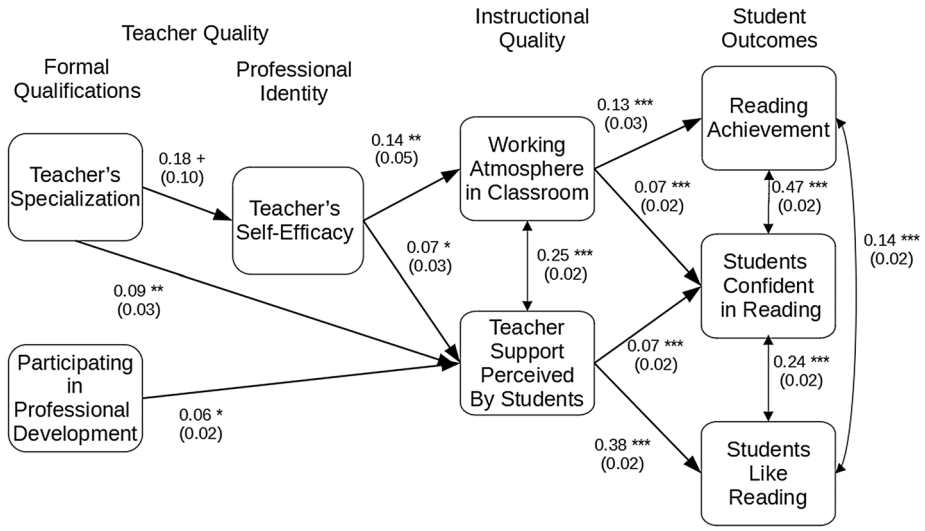

Fig. 3 Path analysis model for Denmark

Table 1 Standardized regression coefficients of controlling variables in the path model for Denmark

Controlling variable	Coefficient (std. error)		
	Reading achievement	Students confident in reading	Students like reading
Male gender	− 0.08 (0.02)***	0.01 (0.02)^{ns}	− 0.16 (0.02)***
Home language	0.09 (0.02)***	0.00 (0.02)^{ns}	− 0.09 (0.02)***
Number of books at home	0.30 (0.02)***	0.19 (0.02)***	0.21 (0.02)***

***p < 0.001; **p < 0.01; *p < 0.05; $^{+}$p < 0.10; ^{ns}p ≥ 0.10

the other standard errors, because these variables are measured at teacher level, while the others are measured at student level. In Fig. 3 this is seen in the standard error of the relation between Teacher's Specialization and Self-Efficacy.

First, we look at the results regarding the controlling variables (Table 1). In Denmark, Number of Books at Home clearly had the strongest associations with outcomes. Male Gender had a small negative association with Reading Achievement and Students Like Reading. The students with the same Home Language as the test language achieved slightly better in reading than the others did, but they also liked reading less. These three controlling variables together explained 11.9% ($R^2 = 0.119$) of variance in Reading Achievement, 3.8% ($R^2 = 0.038$) of variance in Students Confident in Reading, and 9.1% ($R^2 = 0.091$) of variance in Students Like Reading. There was a small but significant correlation (0.09) between Number of Books at Home and Home Language, meaning that the Number of Books at Home tends to be higher in homes where the language spoken is the same as the language used in PIRLS test. Male Gender was not associated with Home Language or Number of Books at Home.

When we look at the associations between variables measuring teacher quality and instructional quality, and student outcomes (Fig. 3), we note that the estimated path coefficients are generally very small even if they are statistically significant at the

adopted 10% level. Thus, the teacher quality and instructional quality does not seem to transfer strongly to positive student outcomes. In the Danish data both Teacher's Specialization and Participating in Professional Development had a significant direct effect on Teacher Support Perceived by Students. These effects were not included in the starting model (and they did not appear in any other country), but they were very small anyway. By far the strongest association was found between Teacher Support Perceived by Students and Students Like Reading, indicating that students who have a positive experience of the teacher's support in the classroom tend to like reading more. Teacher's Specialization was associated with Teacher's Self-Efficacy, but this association was not strong.

The weak associations are mirrored in the low values of R-squared. Overall, the model explains 3.3% ($R^2 = 0.033$) of variance in Teacher's Self-Efficacy, 1.9% ($R^2 = 0.019$) of variance in Working Atmosphere in the Classroom, and 1.8% of variance in Teacher Support Perceived by Students. The R-squared of the overall model is 12.6% ($R^2 = 0.126$) for Reading Achievement, 4.7% ($R^2 = 0.047$) for Students Confident in Reading, and 22.2% ($R^2 = 0.222$) for Students Like Reading. But when we subtract the variance explained by the controlling variables from these R-squared values, we find that teacher quality and instructional quality explain 13.1% of variance in Students Like Reading, but only 0.7% of variance in Reading Achievement, and 0.9% of variance in Students Confident in Reading. The amount of explained variance in Students Like Reading is conclusively due to Teacher Support Perceived by Students.

The errors of Reading Achievement and Students Confident in Reading were quite strongly correlated (0.47). Their correlations with the error of Students Like Reading were weaker. In addition, the errors of Working Atmosphere in the Classroom and Teacher Support Perceived by Students were positively correlated (0.25). As for Formal Qualifications, correlation of Teacher's Specialization with Participating in Professional Development was zero in the Danish data.

Finland

Next, we turn to the path model estimated for Finland (Fig. 4; Table 2). The values of the goodness-of-fit criteria again showed good fit: SRMR = 0.03, RMSEA = 0.02, CFI = 0.96, TLI = 0.95. The number of students used in the estimation was $n = 4387$, and number of classrooms (teachers) was 288.

The Finnish results regarding the controlling variables (Table 2) were largely similar to those observed in Denmark. Again, the Number of Books at Home was positively associated with student outcomes, while Male Gender was negatively associated with them. The students with the same Home Language as the test language achieved slightly better scores in reading than the others did. In Finland, the controlling variables explained 13.5% ($R^2 = 0.135$) of variance in Reading Achievement, 3.8% ($R^2 = 0.038$) of variance in Students Confident in Reading, and 9.1% ($R^2 = 0.091$) of variance in Students Like Reading (the last two R-squared values were exactly the same as in Denmark). The controlling variables were not correlated with each other in Finland.

In Finland both Teacher's Specialization and Participating in Professional Development were statistically significantly associated with Teacher's Self-Efficacy. Actually, among the four Nordic countries the relation between Participating in Professional

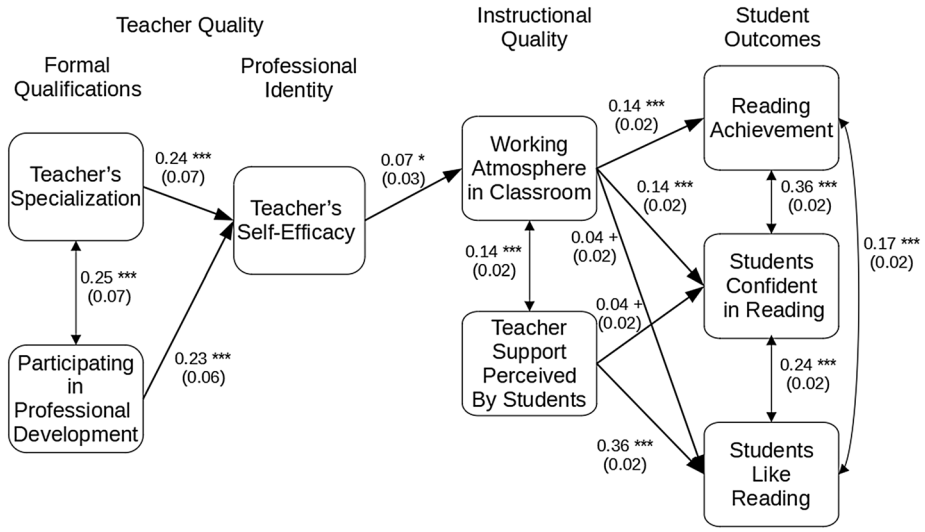

Fig. 4 Path analysis model for Finland

Table 2 Standardized regression coefficients of controlling variables in the path model for Finland

Controlling variable	Coefficient (std. error)		
	Reading achievement	Students confident in reading	Students like reading
Male gender	− 0.16 (0.02)***	− 0.08 (0.02)***	− 0.18 (0.02)***
Home language	0.11 (0.02)***	0.05 (0.02)**	− 0.05 (0.02)*
Number of books at home	0.31 (0.02)***	0.16 (0.02)***	0.20 (0.02)***

***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.10; ^{ns}p ≥ 0.10

Development and Teacher's Self-Efficacy was found to be significant in Finland only, while the relation between Teacher's Specialization and Teacher's Self-Efficacy appeared to be significant in all countries. The link from Teacher's Self-Efficacy to instructional quality was negligible, although there was a small statistically significant path coefficient to Working Atmosphere in the Classroom (Fig. 4). It seems again that teacher quality (when measured with the chosen variables) does not necessarily result in high instructional quality. Furthermore, the relations between measures of instructional quality and student outcomes are small except the one between Teacher Support Perceived by Students and Students Like Reading (standardized path coefficient 0.36). The same was already observed in the Danish data.

The Finnish model explained 13.8% ($R^2 = 0.138$) of variance in Teacher's Self-Efficacy and only 0.5% ($R^2 = 0.005$) of variance in Working Atmosphere in the Classroom. There were no significant explanatory variables for Teacher Support Perceived by Students. The R-squared of the overall model was 15.4% ($R^2 = 0.154$) for Reading Achievement, 5.9% ($R^2 = 0.059$) for Students Confident in Reading, and 20.7% ($R^2 = 0.207$) for Students Like Reading, but when we remove the variance explained by the controlling variables, we find that the contribution of teacher quality and

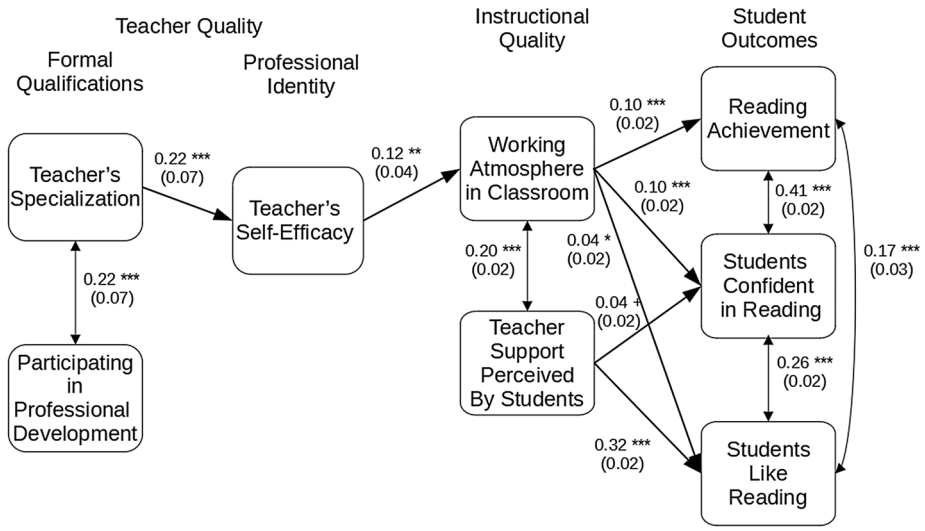

Fig. 5 Path analysis model for Norway

Table 3 Standardized regression coefficients of controlling variables in the path model for Norway

Controlling variable	Coefficient (std. error)		
	Reading achievement	Students confident in reading	Students like reading
Male gender	− 0.12 (0.02)***	− 0.04 (0.02)$^{+}$	− 0.12 (0.02)***
Home language	0.06 (0.02)*	0.02 (0.02)^{ns}	− 0.09 (0.02)***
Number of books at home	0.26 (0.02)***	0.20 (0.02)***	0.18 (0.02)***

***p < 0.001; **p < 0.01; *p < 0.05; $^{+}$p < 0.10; ^{ns}p ≥ 0.10

instructional quality to the R-squared was 1.9% for Reading Achievement, 2.1% for Students Confident in Reading, and 11.6% for Students Like Reading.

The correlation of errors of Reading Achievement and Students Confident in Reading was 0.36 in Finland, and the correlation of errors of Working Atmosphere in the Classroom and Teacher Support Perceived by Students was 0.14. Unlike in Denmark, Teacher's Specialization and Participating in Professional Development had a positive correlation (0.25) in Finland, so that in Finland teachers with specialization in reading pedagogy and language of the test as a part of their formal degree had participated somewhat more often in formal professional development.

Norway

The path model results for Norway are presented in Fig. 5 and Table 3. The model fit was again good: SRMR = 0.02, RMSEA = 0.01, CFI = 0.99, TLI = 0.99. The number of students used in the estimation was $n = 3676$, and the number of classrooms (teachers) was 207.

The results for Norway were close to those obtained for Finland. The main difference between the Norwegian and Finnish models is that there is no significant direct effect of Participating in Professional Development in Norway.

The controlling variables explained 10.0% ($R^2 = 0.100$) of variance in Reading Achievement in Norway, while the respective R-squared was 4.4% ($R^2 = 0.044$) for Students Confident in Reading, and 5.9% ($R^2 = 0.059$) for Students Like Reading in Norway. Number of Books at Home was again the variable with the strongest contribution. Like in Denmark, there was a small but significant correlation (0.10) between Number of Books at Home and Home Language. Male Gender also had a small negative association with Number of Books at Home (correlation − 0.09).

In the Norwegian data, the model explained 5.0% ($R^2 = 0.005$) of variance in Teacher's Self-Efficacy and 1.3% ($R^2 = 0.013$) of variance in Working Atmosphere in the Classroom. Like in Finland, there were no significant explanatory variables for Teacher Support Perceived by Students. The R-squared of the overall model was 10.7% ($R^2 = 0.107$) for Reading Achievement, 5.5% ($R^2 = 0.055$) for Students Confident in Reading, and 16.1% ($R^2 = 0.161$) for Students Like Reading. By subtracting the contribution of the controlling variables, we found that the variables representing teacher quality and instructional quality only explained 0.7% of the variance in Reading Achievement, and 1.1% of variance in Students Confident in Reading in Norway. The explained variance for Students Like Reading was again the largest being 10.2%. Again, this resulted mainly from the positive contribution of Teacher Support Perceived by Students.

The errors of Reading Achievement, Students Confident in Reading and Students Like Reading were all correlated in Norway also. The highest correlation (0.41) was again met between Reading Achievement and Students Confident in Reading. The correlation of errors of Working Atmosphere in the Classroom and Teacher Support Perceived by Students was 0.20. Teacher's Specialization and Participating in Professional Development had a positive correlation (0.22) in Norway. This is again close to the value observed in the Finnish data.

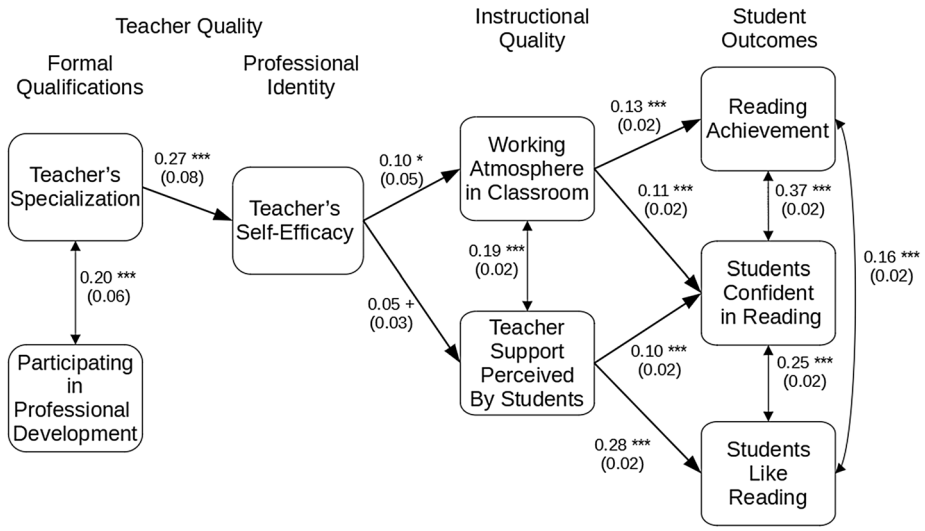

Fig. 6 Path analysis model for Sweden

Table 4 Standardized regression coefficients of controlling variables in the path model for Sweden

Controlling variable	Coefficient (std. error)		
	Reading achievement	Students confident in reading	Students like reading
Male gender	− 0.08 (0.02)***	–0.02 (0.02)^{ns}	–0.12 (0.02)***
Home language	0.09 (0.03)***	0.01 (0.02)^{ns}	–0.11 (0.02)***
Number of books at home	0.34 (0.02)***	0.17 (0.02)***	0.23 (0.02)***

***p < 0.001; **p < 0.01; *p < 0.05; $^{+}$ p < 0.10; ^{ns}p ≥ 0.10

Sweden

Finally, we present the results for Sweden in Fig. 6 and Table 4. Again, the goodness-of-fit criteria indicated good model fit: SRMR = 0.02, RMSEA = 0.01, GFI = 0.99, CFI = 0.98. The number of students used in the estimation was $n = 3757$, and number of classrooms (teachers) was 217.

The key findings for Sweden did not differ from those already obtained in other Nordic countries. There was a positive association between Teacher's Specialization and Teacher's Self-Efficacy, but the further connections to instructional quality were weak. Like in the other countries, the strongest association between instructional quality and student outcomes was the one between Teacher Support Perceived by Students and Students Like Reading.

The controlling variables explained 14.5% ($R^2 = 0.145$) of variance in Reading Achievement in Sweden, and the respective R-squared was 2.9% ($R^2 = 0.029$) for Students Confident in Reading, and 7.7% ($R^2 = 0.077$) for Students Like Reading. The correlation between Number of Books at Home and Home Language (being the same as the language of the test) was larger (0.19) in Sweden than in the other countries. Male Gender had a small negative association with Number of Books at Home (correlation − 0.09) in Sweden also.

The model for Sweden explained 7.4% of ($R^2 = 0.074$) variance in Teacher's Self-Efficacy (resulting from the effect of Teacher's Specialization in reading pedagogy and language of the test), but only 0.9% ($R^2 = 0.009$) of variance in Working Atmosphere in the Classroom and 0.3% ($R^2 = 0.003$) of variance in Teacher Support Perceived by Students. This again gives rise to the conclusion that formal qualifications and professional identity of teacher are not associated with instructional quality in the Nordic countries (as measured with the PIRLS questionnaires).

For students' Reading Achievement, the R-squared of the overall model was 16.2% ($R^2 = 0.162$) in Sweden. The respective values were 5.5% ($R^2 = 0.055$) for Students Confident in Reading, and 15.0% ($R^2 = 0.150$) for Students Like Reading. If we remove the effects of the controlling variables, the contribution of teacher quality and instructional quality becomes 1.7% for Reading Achievement, 2.6% for Students Confident in Reading, and 7.3% for Students Like Reading.

The correlations of errors of Reading Achievement, Students Confident in Reading and Students Like Reading were almost the same in Sweden as in the other countries. The highest correlation (0.37) was again between Reading Achievement and Students Confident in Reading. The correlation of errors of Working Atmosphere in the Classroom

and Teacher Support Perceived by Students was 0.19 in Sweden. In addition, Teacher's Specialization and Participating in Professional Development had a correlation of 0.20. These two variables were thus almost similarly correlated in all Nordic countries except in Denmark.

Discussion

Discussion of findings

In this study, we used the PIRLS 2016 data in conducting path analyses that explored the relations of two subdimensions of teacher quality (formal qualification and professional identity), along with three subdimensions of instructional quality (classroom management, cognitive activation and teacher support) and student outcomes in four Nordic countries. First, we were interested in the relations of teachers' formal qualifications and professional identity as well as direct associations of formal qualifications with instructional quality and student outcomes.

Of the original three variables measuring formal qualifications, Teacher's Specialization had a significant positive association with Teacher's Self-Efficacy, which, in turn, was the only remaining measure of professional identity in the final models. This association was weak, but it was observed in all of the countries we considered. In addition, Participating in Professional Development was associated with Teacher's Self-Efficacy in Finland. Teacher's Specialization and Participating in Professional Development were intercorrelated in all the other countries except for in Denmark, suggesting that those who had specialized in reading issues in their formal education tend to participate more in reading-related professional development activities. Formal qualifications had no associations with student outcomes in any of the countries.

In our analysis, teachers' Formal Education had no statistically significant associations with student outcomes or teacher-related variables. It was not therefore retained in the final path model for any country. This finding contradicts some earlier studies (e.g., Blömeke et al., 2016; Nilsen et al., 2018). One possible explanation is that the educational level of teachers in Nordic countries is uniform and relatively high. As noted in the Introduction, most of the participating teachers had at least a bachelor's degree. From the viewpoint of statistical modelling, a variable with little variation appears insignificant even though the issue itself may be important.

Another possible explanation is that the original teacher training may not contain enough material specific to teaching reading, regardless of the formal degree. Due to the wide range of content a primary school teacher has to study, there are differences in how much teaching reading is included in the compulsory studies of teacher training, even within a single country. According to our analysis, Teacher's Specialization in teaching reading plays a more important role than the level of Formal Education does in regards to a teacher's professional identity and, further, to instructional quality and student outcomes.

As noted, the frequency of participating in reading-related professional development was lowest in Finland and less than 50% in Denmark. This means there is considerable heterogeneity for the variable among teachers in Denmark and Finland, which can be the reason why it emerges in models of these countries. This suggests that reading-related professional development, just like specialization in reading in

formal education, is important for teachers to teach reading efficiently, and teachers should be encouraged to participate in such training. Overall, the issue of professional development is important for teacher quality in Nordic countries. According to Taajamo (2016), in-service training in the Nordic countries seems fragmented because there are many different organizations offering it. His study, based on TALIS 2013 data, showed that, in the Nordic countries, teachers spent minimal time on professional development activities even though they find the activities beneficial. Fuglestad et al. (2017), however, have also observed increased participation in professional development among Norwegian teachers.

The second research question focused on professional identity. The only variable, which measured professional identity and remained in the final model, was Teacher's Self-Efficacy, and it had a positive effect on Working Atmosphere in the Classroom in all the countries and also on Teacher Support Perceived by Students in Denmark and Sweden. Nilsen et al., (2018) found an association between self-efficacy regarding pedagogical content knowledge and instructional quality in science teaching. They also found associations with science-related student outcomes, but in our study direct associations with student outcomes were not evident. Job Satisfaction showed no significant associations, which differs from the findings of Banerjee et al. (2017), who found modest but positive relations regarding job satisfaction and reading achievement. Moreover, Teacher Collaboration showed no significant relations in this study, which raises the question of differences in collaboration on teaching reading and science (cf. Nilsen et al., 2018). According to Nilsen et al. (2018), effective teacher collaboration should include professional development embedded within school and classroom practices, clearly defined learning goals, and structures and processes that support teaching innovations.

Finally, the third research question focused on instructional quality and its relations. The analysis started with several variables related to instructional quality, but only a couple of them showed associations which could be considered significant. Teacher Support Perceived by Students was the only relevant variable measuring teacher support. It had a rather strong positive relation with Students Like Reading and a weak but positive relation with Students Confident in Reading in all four Nordic countries. Similarly, Working Atmosphere in the Classroom was the only relevant variable measuring classroom management. It had a positive association with Reading Achievement and Students Confident in Reading in all four countries and, additionally, a positive relation to Students Like Reading in Finland and Norway. So even though variables of instructional quality had few connections to Reading Achievement, they seem to be important for affective-motivational student outcomes, which then correlate to achievement. However, cognitive activation, such as a teacher reading aloud to students, had no significant associations with student outcomes in our study.

In general, the associations of student outcomes with teacher quality and instructional quality, as measured above, did not appear particularly strong. This suggests that, at least in the Nordic countries, the variation in students' reading results cannot be straightforwardly reduced to differences in teaching. Teachers and instruction of high quality do not necessarily manifest in high student achievements. The explained variance was the largest for Students Like Reading, especially through Teacher Support Perceived by Students, and the smallest for Reading Achievement.

Limitations and future research

Due to the cross-sectional nature of data, true causal effects cannot be detected. In addition, despite the large number of considered variables, there may always exist variables which are not measured and which may intervene in the observed relations. In this data, the number of teachers was small even though the school samples were nationally representative. The coefficients of determination (R-squared) of the outcomes were generally small. One reason for this is that in several considered variables there was little national variation between teachers. On the other hand, if there was remarkable between-teacher variation, it did not necessarily coincide with the variation in student outcomes. Among the variables we were originally interested in, there were many which lacked significant associations with the outcomes, or which gave inconsistent results due to intercorrelations with other explanatory variables. This led to the dropping of several variables from the final analyses. One of those was the collaboration of the teachers, which showed no significance here. Ronfeldt et al. (2015) have noted that teacher collaboration can vary a lot within schools and this within-school variation is larger than the variation between schools.

These analyses were done country by country. The aim was not to compare the statistical differences between countries, but rather to learn which kind of associations appear in the four Nordic countries. This clarifies the focus of this article but also leaves some questions. For example, the level of education did not show any significance in the models. Teacher's educational level, however, is clearly highest in Finland while the level of reading-related professional development is the lowest. It would be interesting to complement these results with a study of the content and quality of formal teacher education and reading specialization studies in Nordic countries.

In the large-scale assessments, the questionnaires are usually targeted to cover many areas, and consequently there is no room to focus very deeply on the phenomenon in question (see also Nilsen et al., 2016). An example of this kind of shortcoming is measuring self-efficacy, which was based on six items only and did not include items on efficacy in disciplinary matters or any variables related to school operations and work organization (cf. Bandura, 1997; Friedman & Kass, 2002). Our hypothesis is that these issues might also correlate to job satisfaction and classroom management because the atmosphere at the school usually transfers to the classroom. This should be considered in future research.

The data in this study were largely based on teachers' self-reported responses to questions in the PIRLS teacher questionnaire. Similarly to Van Staden et al. (2019) as well as Shiel and Eivers (2009), we found that the PIRLS teacher questionnaire data show little that is new about the instructional quality of the teachers. As they have suspected, one reason may be that teachers reported the frequency of their activities too positively or too much in line with social expectations (e.g., what is expected in the curriculum) rather than reporting the reality of what they do. In addition, some of the teachers' actions, such as support given to students, may vary individually, but the questionnaire captures only the teacher's average estimation for the whole class. Even though student-specific questionnaires are laborious for teachers, they would give more precise information about the adaptation and differences of teaching for each student in the class. In

such a study, a student's participation in special education should also be included in the analysis.

Finally, we note that the focus of this study was only on reading and reading-related outcomes. As Emler et al. (2019) point out, teachers work with a wide range of outcomes, such as creativity, problem-solving, organization of knowledge, self-monitoring skills, and entrepreneurship. Unlike in China (see Zhao, 2014), for example, in the Nordic countries the curriculum and study content are not guided by participation in large-scale assessments. Therefore, teachers' work includes a lot that has not been measured in PIRLS or other large-scale assessments.

Conclusion

In this study, the four Nordic countries appear, despite small differences, very similar when looking at the relationship between teacher quality, professional identity, instructional quality and student outcomes. Our findings emphasize three factors that should be addressed in teacher training and among in-service teachers. First, Teacher's Self-Efficacy had positive associations with certain variables measuring instructional quality. Every teacher should have the right to participate in professional development in order to gain additional expertise in content about which they feel uncertain. A student's reading skills are the basis for all learning and therefore this issue is especially important.

Second, the most significant variables of instructional quality were Work Atmosphere in the Class and Teacher Support Perceived by Students measured by the students' view. Students have the right to a safe, peaceful, and encouraging learning atmosphere. The emergence of issues related to classroom management implies that this right is not exercised for all students. For one reason or another, teachers are unable to organize class activities so that there is a peaceful working atmosphere for everyone.

Third, teachers' work is not limited only to test scores and achievements, meaning it is important to measure a range of outcomes. Even though students' skills at some levels may be deficient, the teacher can have a significant influence on students' attitudes and interest, which feed students' activities and can later be reflected in improved grades. This is especially important because students' interest in reading has declined in many countries, and according to the PIRLS 2016 survey (Mullis et al., 2017, Student Engagement and Attitudes section), the Nordic countries were at the very bottom when looking at the percentage of students who like reading. A supportive environment provided by the teacher can be crucial in increasing students' reading motivation and enjoyment.

This study emphasizes that teacher training and professional development activities should, on one hand, focus on subject-targeted pedagogical issues, such as teaching reading but, on the other hand, also on general pedagogy, such as classroom management and supporting students' learning through, for example, feedback. This study suggests that teachers need more knowledge and tools for these areas to achieve more

equal education for students. The research shows that teaching and especially teachers' actions are important. More attention should be paid to the areas identified here than is presently the case, because they can support both skills and attitudes towards reading.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40536-022-00146-4.

Additional file 1. Correlation matrices of final analysis for all four countries.

Acknowledgements
Not applicable.

Author contributions
KL was a major contributor in writing the manuscript, especially theoretical background and discussion. She has also been National research coordinator when collecting the data. KN performed methodological analysis and contributed describing those. MS contributed to theoretical background and editing the manuscript. All authors contributed to and approved the final manuscript.

Authors' information
KL, PhD., senior researcher at University of Jyväskylä, the Finnish Institute for Educational Research. KN, PhD., senior researcher at University of Jyväskylä, the Finnish Institute for Educational Research. MS, MA., project researcher at University of Jyväskylä, the Finnish Institute for Educational Research..

Funding
This article was funded by University of Jyväskylä, the Finnish Institute for Educational Research which acted as national research center in PIRLS 2016.

Availability of data and materials
The dataset supporting the conclusions of this article is available in the PIRLS repository https://timssandpirls.bc.edu/pirls2016/international-database/index.html.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 17 February 2022 Accepted: 5 December 2022
Published online: 19 December 2022

References

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher, 36*(7), 369–387. https://doi.org/10.3102/0013189X07308739

Almasi, J., & Garas-York, K. (2009). Comprehension and discussion of text. In S. Israel, & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 470–493). Routledge. https://www.academia.edu/37378134/Handbook_of_Research_on_Reading_Comprehension_Duffy_and_Israel_Taylor_Francis_2014_pdf. Accessed 12 June 2022.

Alvermann, D., & Moje, E. (2013). Adolescent literacy instruction and the discourse of "every teacher a teacher of reading." In D. Alvermann, N. Unrau, & R. Ruddell (Eds.), *Theoretical models and processes of reading* (pp. 1072–1103). International Reading Association. https://doi.org/10.1598/0710.39

Ariail, M., & Albright, L. K. (2005). A survey of teachers' read-aloud practices in middle schools. *Literacy Research and Instruction, 45*(2), 69–89. https://doi.org/10.1080/19388070609558443

Baker, L., & Beall, L. (2009). Metacognitive processes and reading comprehension. In S. Israel, & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 373–388). Routledge. https://www.academia.edu/37378134/Handbook_of_Research_on_Reading_Comprehension_Duffy_and_Israel_Taylor_Francis_2014_pdf. Accessed 12 June 2022.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Banerjee, N., Stearns, E., Moller, S., & Mickelson, R. A. (2017). Teacher job satisfaction and student achievement: The roles of teacher professional community and teacher collaboration in schools. *American Journal of Education, 123*(2), 203–241. https://doi.org/10.1086/689932

Bempechat, J. (2004). The motivational benefits of homework: A social cognitive perspective. *Theory into Practice, 43*(3), 189–196. https://doi.org/10.1207/s15430421tip4303_4

Bentler, P. M. (1995). *EQS: Structural equations program manual, Program version 5.0*. Multivariate Software.

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106*(2), 315–317. https://doi.org/10.1037/0033-2909.106.2.315

Berge I. M., Helgevold L., & Schulz-Heidorf K. (2017). Hva lærerne rapporterer om egen undervisning i strategier [How teachers teach reading strategies]. In E. Gabrielsen (Ed.), *Klar framgang! Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv*. Oslo, Universitetsforlaget. https://doi.org/10.18261/9788215030258-2017-10

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 Open. https://doi.org/10.1007/978-3-319-41252-8

Bong, M. (2006). Asking the right question: How confident are you that you could successfully perform these tasks? In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 287–305). Information Age.

Bowen, N. K., & Guo, S. (2012). Structural equation modelling. *Oxford Scholarship Online*. https://doi.org/10.1093/acprof:oso/9780195367621.001.0001

Canrinus, E. T., Helms-Lorenz, M., Beijaard, D., Buitink, J., & Hofman, A. (2012). Self-efficacy, job satisfaction, motivation and commitment: Exploring the relationships between indicators of teachers' professional identity. *European Journal of Psychology of Education, 27*, 115–132. https://doi.org/10.1007/s10212-011-0069-2

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44*, 473–490. https://doi.org/10.1016/j.jsp.2006.09.001

Clotfelter, C., Ladd, H., & Vigdor, J. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*(3), 655–681.

Cooper, H., Lindsay, J. J., Nye, B., & Greathouse, S. (1998). Relationships among attitudes about homework, amount of homework assigned and completed, and student achievement. *Journal of Educational Psychology, 90*(1), 70–83. https://doi.org/10.1037/0022-0663.90.1.70

Creemers, B. P. M., & Kyriakides, L. (2008). The dynamics of educational effectiveness: A contribution to policy, practice, and theory in con-temporary schools. *Routledge*. https://doi.org/10.4324/9780203939185

Crouch, D. (2015). Highly trained, respected and free: why Finland's teachers are different. *The Guardian*. https://www.theguardian.com/education/2015/jun/17/highly-trained-respected-and-free-why-finlands-teachers-are-different. Accessed 15 May 2022.

Curby, T. W., Rimm-Kaufman, S. E., & Abry, T. (2013). Do emotional support and classroom organization earlier in the year set the stage for higher quality instruction? *Journal of School Psychology, 51*, 557–569. https://doi.org/10.1016/j.jsp.2013.06.001

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis, 8*(1). https://epaa.asu.edu/ojs/index.php/epaa/article/view/392/515&gt%3b. Accessed 12 June 2022.

Dicke, T., Marsh, H. W., Parker, P. D., Guo, J., Riley, P., & Waldeyer, J. (2020). Job satisfaction of teachers and their principals in relation to climate and student achievement. *Journal of Educational Psychology, 112*(5), 1061–1073. https://doi.org/10.1037/edu0000409

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching. A project of the American Educational Research Association* (pp. 392–431). Macmillan.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children, 65*(3), 399–415. https://doi.org/10.1177/001440299906500309

Emler, T. E., Zhao, Y., Deng, J., Yin, D., & Wang, Y. (2019). Side effects of large-scale assessments in education. *ECNU Review of Education, 2*(3), 279–296. https://doi.org/10.1177/2096531119878964

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA Journal of the American Medical Association, 287*, 226–235. https://doi.org/10.1001/jama.287.2.226

Evans, L. (1997). Addressing problems of conceptualization and construct validity in researching teachers' job satisfaction. *Educational Research, 39*(3), 319–331. https://doi.org/10.1080/0013188970390307

Federici, R. A., & Skaalvik, E. M. (2014). Students' perceptions of emotional and instrumental teacher support: Relations with motivational and emotional responses. *International Education Studies, 7*(1), 21–36. https://doi.org/10.5539/ies.v7n1p21

Foy, P., & Yin. L. (2017). Scaling the PIRLS 2016 achievement data. In M. O. Martin, M.O., I. V. S. Mullis, I.V.S., & M. Hooper, M. (Eds.), *Methods and procedures in PIRLS 2016* (pp. 12.1–12.38). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-12.html. Accessed 12 June 2022.

Freiberg, J. H. (2013). Classroom management and student achievement. In J. Hattie, & E. M. Anderman (Eds.), *International Guide to Student Achievement* (pp. 228–231). Routledge. https://doi.org/10.4324/9780203850398

Freiberg, J. H., Huzinec, C. A., & Templeton, S. M. (2009). Classroom management—A pathway to student achievement: A study of fourteen inner-city elementary schools. *The Elementary School Journal, 110*(1), 63–80. https://doi.org/10.1086/598843

Friedman, I. A., & Kass, E. (2002). Teacher self-efficacy: A classroom-organization conceptualization. *Teaching and Teacher Education, 18*, 675–686. https://doi.org/10.1016/S0742-051X(02)00027-6

Frønes, T. S., Pettersen, A., Radišić, J., & Buchholtz, N. (2020). Equity, equality, and diversity in the Nordic model of education: Contributions from large-scale studies. In T. S. Frønes, A. Pettersen, J. Radišić, & N. Buchholtz, (Eds.), Equity, Equality and Diversity in the Nordic Model of Education (pp. 1–10). Springer. https://doi.org/10.1007/978-3-030-61648-9

Fuglestad, U., Hoem, T. F., & Schulz-Heidorf, K. (2017). Lærerens betydning for norske elevers leseresultater—hva forteller PIRLS 2016? [The teacher's significance for Norwegian students' reading results - what does PIRLS 2016 say?] In E. Gabrielsen (Ed.), *Klar framgang! Leseferdighet på 4. og 5. trinn i et femtenårsperspektiv*. Oslo, Universitetsforlaget. https://doi.org/10.18261/9788215030258-2017-7

Gabrielsen, E. (2017). Norway. In I. V. S. Mullis, M. O.I. V. S., Martin, M. O., S. Goh, S., & C. Prendergast, C. (Eds.), *PIRLS 2016 encyclopedia: Education policy and curriculum in reading*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/pirls2016/encyclopedia/. Accessed 12 June 2022.

Goddard, Y. L., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877–896.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis.* National comprehensive center for teacher quality. Education Commission of the States, ETS, Learning Point Associates, and Vanderbilt University. https://gtlcenter.org/sites/default/files/docs/LinkBetweenTQandStudentOutcomes.pdf. Accessed 12 June 2022.

Gustafsson, J-E., Hansen, K. Y., & Rosén, M. (2013). Effects of home background on student achievement in reading, mathematics, and science at the fourth grade. In M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade—Implications for early learning* (pp. 181–287). Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timsspirls2011/downloads/TP11_Chapter_4.pdf. Accessed 12 June 2022.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007b). *Building a science of classroom application of the CLASS framework in over 4000 U.S. early childhood and elementary classrooms.* Foundation for Child Development. http://fcdus.org/sites/default/files/BuildingAScienceOfClassroomsPiantaHamre.pdf. Accessed 12 June 2022.

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 49–83). Paul H Brookes Publishing.

Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. *Routledge*. https://doi.org/10.4324/9780203887332

Hemmerechts, K., Agirdag, O., & Kavadias, D. (2017). The relationship between parental literacy involvement, socio-economic status and reading literacy. *Educational Review, 69*(1), 85–101. https://doi.org/10.1080/00131911.2016.1164667

Ho, E. S. C., & Lau, K. L. (2018). Reading engagement and reading literacy performance: Effective policy and practices at home and in school. *Journal of Research in Reading, 41*(4), 657–679. https://doi.org/10.1111/1467-9817.12246

Housand, A., & Reis, S. M. (2008). Self-regulated learning in reading: Gifted pedagogy and instructional settings. *Journal of Advanced Academics, 20*(1), 108–136. https://doi.org/10.4219/jaa-2008-865

Huang, J., & Chen, G. (2018). From reading strategy instruction to student reading achievement. *Psychology in Schools, 56*(5), 724–740. https://doi.org/10.1002/pits.22217

Hurst, S., & Griffity, P. (2015). Examining the effect of teacher read-aloud on adolescent attitudes and learning. *Middle Grades Research Journal 10*(1), 31–47. https://www.thefreelibrary.com/Examining+the+effect+of+teacher+read-aloud+on+adolescent+attitudes...-a0437133007. Accessed 5 Sep 2022.

ILA (2018). *Literacy leadership brief: The power and promise of read-alouds and independent reading*. International Literacy Association. https://www.literacyworldwide.org/docs/default-source/where-we-stand/ila-power-promise-read-alouds-independent-reading.pdf. Accessed 5 Sep 2022.

Jacobs, J. S., Morrison, T. G., & Swinyard, W. R. (2000). Reading aloud to students: A national probability study of classroom reading practices of elementary school teachers. Reading Psychology, 21(3), 171–193. https://doi.org/10.1080/02702710050144331

Jensen, M. T., Solheim, O. J., & Cosmovici Idsøe, E. M. (2019). Do you read me? Associations between perceived teacher emotional support, reader self-concept, and reading achievement. *Social Psychology of Education, 22*(2), 247–266. https://doi.org/10.1007/s11218-018-9475-5

Khorramdel, L., von Davier, M., Gonzalez, E., & Yamamoto, K. (2020). Plausible values: principles of item response theory and multiple imputations. In D.B. Maehler, & B. Rammstedt (Eds.), *Large-scale cognitive assessment. Analyzing PIAAC data* (pp. 27–47). Springer Cham. https://doi.org/10.1007/978-3-030-47515-4

Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262–273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x

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

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research, 86*, 643–680. https://doi.org/10.3102/0034654315626799

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

Lafleur, A., & Boucher, V. J. (2015). The ecology of self-monitoring effects on memory of verbal productions: Does speaking to someone make a difference? *Consciousness and Cognition, 36*, 139–146. https://doi.org/10.1016/j.concog.2015.06.015

Lane, H. B., & Wright, T. L. (2007). Maximizing the effectiveness of reading aloud. *Reading Teacher, 60*(7), 668–675. https://doi.org/10.1598/RT.60.7.7

Lerkkanen, M.-K., Salminen, J., & Pakarinen, E. (2018). Varhaislapsuuden lukuhetket tukevat lukutaitoa [Shared reading in early childhood supports reading literacy]. *Onnimanni, 2018*(1–2), 20–26.

Lleras, C., & Rangel, C. (2009). Ability grouping practices in elementary school and African–American/Hispanic achievement. *American Journal of Education, 115*(2), 279–304. https://doi.org/10.1086/595667

Maruyama, G. (1998). *Basics of structural equation modelling*. Sage.

Martin, M. O, Mullis, I. V. S., & Hooper, M. (Eds.). (2017). *Methods and procedures in PIRLS 2016*. Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/publications/pirls/2016-methods.html. Accessed 2 Feb 2022.

Milam, A. J., Furr-Holden, C. D. M., & Leaf, P. J. (2010). Perceived school and neighborhood safety, neighborhood violence and academic achievement in urban school children. *The Urban Review, 42*(5), 458–467. https://doi.org/10.1007/s11256-010-0165-7

Morgan, P. L., & Meier, C. R. (2008). Dialogic reading's potential to improve children's emergent literacy skills and behavior. *Preventing School Failure, 52*(4), 11–16. https://doi.org/10.3200/PSFL.52.4.11-16

Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 assessment framework* (2nd edn.). Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/pirls2016/framework.html. Accessed 2 Feb 2022.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/pirls2016/international-results/. Accessed 2 Feb 2022.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus User's Guide* (7th ed.). Muthén & Muthén.

Neff, L. (2015). *The relationship between reading enjoyment, gender, socioeconomic status and reading outcomes in PISA 2009*. Doctoral dissertation, Georg Fox University. Paper 54. http://digitalcommons.georgefox.edu/edd/54. Accessed 5 Mar 2022.

Nilsen, T., 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*. IEA Research for Education, vol 2. Springer. https://doi.org/10.1007/978-3-319-41252-8_1

Nilsen, T., & Gustafsson, J.-E. (Eds.) (2016). *Teacher quality, instructional quality and student outcomes*. IEA Research for Education, vol 2. Springer. https://doi.org/10.1007/978-3-319-41252-8

Nilsen, T., Scherer, R., & Blömeke, S. (2018). The relation of science-teachers' quality and instruction to student motivation and achievement in the 4th and 8th grade: A Nordic perspective. In T. Nilsen, R. Scherer, & S. Blömeke (Eds.), *Northern lights on TIMSS and PISA 2018* (pp. 61–94). The Nordic Council of Ministers. TemaNord 2018:524. https://www.udir.no/contentassets/24c699db4e104200ad60541bf54846c6/northern_lights_on_timss_and_pisa_2018.pdf. Accessed 10 Jan 2022.

Nortvedt, G. A., Gustafsson, J.-E., & Lehre, A.-C. W. (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* (pp. 97–113). IEA Research for Education, vol 2. Springer. https://doi.org/10.1007/978-3-319-41252-8_5

OECD (2019c). *TALIS 2018 results (volume II): Teachers and school leaders as valued professionals*. OECD Publishing. https://doi.org/10.1787/19cf08df-en

OECD (2019b). *PISA 2018 results (volume II): Where all students can succeed*. OECD Publishing. https://doi.org/10.1787/b5fd1b8f-en

OECD (2019a). *PISA 2018 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/b25efab8-en

Ponzo, M. (2012). *Does bullying reduce educational achievement? An evaluation using matching estimators*. MPRA Paper No. 36064. https://mpra.ub.uni-muenchen.de/36064/. Accessed 8 Feb 2022.

Reutzel, D. R., Fawson, P. C., & Smith, J. A. (2008). Reconsidering silent sustained reading: An exploratory study of scaffolded silent reading. *The Journal of Educational Research, 102*(1), 37–50. https://doi.org/10.3200/JOER.102.1.37-50

Robinson, M. F., Meisinger, E. B., & Joyner, R. E. (2018). The influence of oral versus silent reading on reading comprehension in students with reading disabilities. *Learning Disability Quarterly, 42*(2), 105–116. https://doi.org/10.1177/0731948718806665

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal, 52*(3), 475–514. https://doi.org/10.3102/0002831215585562

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the last decade: Role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*, 454–499. https://doi.org/10.3102/0034654307310317

Shiel, G., & Eivers, E. (2009). International comparisons of reading literacy: What can they tell us? *Cambridge Journal of Education, 39*(3), 345–360. https://doi.org/10.1080/03057640903103736

Shulman, L. 1986. Those who understand. Knowledge growth in teaching. Educational Researcher 15 (2), 4–14.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytical review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Steiger, J., & Lind, J. (1980). *Statistically-based tests for the number of common factors* [Conference presentation]. Annual Meeting of the Psychometric Society, Iowa City, IA, United States.

Støle, H., Wagner, Åse Kari H. & Schwippert, K. (2020). The importance of parents' own reading for 10-year-old students' reading achievement in the Nordic countries. In T. S. Frønes, A. Pettersen, J. Radišić, & N. Buchholtz (Eds.), *Equity, Equality and Diversity in the Nordic Model of Education* (pp. 363–384). Springer. https://doi.org/10.1007/978-3-030-61648-9_14

Strati, A. D., Schmidt, J. A., & Maier, K. S. (2017). Perceived challenge, teacher support, and teacher obstruction as predictors of student engagement. *Journal of Educational Psychology, 109*(1), 131–147. https://doi.org/10.1037/edu0000108

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education, 20*, 165–184. https://doi.org/10.1007/s11092-008-9053-z

Taajamo, M. (2016). Teachers' professional development in Nordic countries. In S. Ludvigsen, G. A. Nortvedt, A. Pettersen, A. Pettersson, S. Sollerman, R. F. Ólafsson, M. Taajamo, J. Caspersen, P. Nyström, & J. Braeken (Eds.), *Northern Lights on PISA and TALIS*. Nordic Council of Ministers. TemaNord, 2016:517. http://norden.diva-portal.org/smash/get/diva2:920756/FULLTEXT01.pdf. Accessed 2 May 2022.

Torppa, M., Vasalampi, K., Eklund, K., & Niemi, P. (2022). Long-term effects of the home literacy environment on reading development: Familial risk for dyslexia as a moderator. *Journal of Experimental Child Psychology*, 215. https://doi.org/10.1016/j.jecp.2021.105314

Unrau, N. J. (1992). The TASK of reading (and writing) arguments: A guide to building critical literacy. *Journal of Reading, 35*(6), 436–442.
Van Staden, S., Combrinck, C., Roux, K., Tshele, M., & Palane, N. M. (2019). Moving beyond league table standings: How measures of opportunity to learn can inform educational quality and policy directives. *South African Journal of Childhood Education, 9*(1), a712. https://doi.org/10.4102/sajce.v9i1.712
Van Staden, S., & Zimmerman, L. (2017). Evidence from the Progress in International Reading Literacy Study (PIRLS) and how teachers and their practice can benefit. In V. Scherman, R. J. Bosker, & S. J. Howie (Eds.), *Monitoring the quality of education in schools: Examples of feedback into systems from developed and emerging economies* (pp. 123–138). Sense Publishers. https://doi.org/10.1007/978-94-6300-453-4_9
Vaughn, S., Linan-Thompson, S., Kouzekani, K., Pedrotty Bryant, D., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education, 24*(5), 301–315. https://doi.org/10.1177/07419325030240050501
Volmari, K. (Ed.) (2019). *Basic education in the Nordic region. Similar values, different policies*. Finnish National Agency for Education. Reports and surveys 2019:4. https://www.oph.fi/sites/default/files/documents/30955454_basic_education_in_the_nordic_region_netti_19_04_08_0.pdf. Accessed 10 July 2022.
von Davier, M., Gonzalez, E, & Mislevy, R.J. (2009). What are plausible values and why are they useful? *IERI Monograph Series Volume 2 (October 2009)*. https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf. https://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdfAccessed 29 Aug 2022.
Wantchekon, K., & Kim, J. S. (2019). Exploring heterogeneity in the relationship between reading engagement and reading comprehension by achievement level. *Reading and Writing Quarterly, 35*, 539–555. https://doi.org/10.1080/10573569.2019.1594474
Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology, 93*(2), 243–250. https://doi.org/10.1037/0022-0663.93.2.243
Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2–3), 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005
Yair, G. (2000). Educational battlefields in America: The tug-of-war over students' engagement with instruction. *Sociology of Education, 73*(4), 247–269. https://doi.org/10.2307/2673233
Zhao, Y. (2014). *Who's afraid of the big bad dragon: Why China has the best (and worst) education system in the world.* Jossey-Bass.

Publisher's Note