Comparing the teaching quality of alternatively certified teachers and traditionally certified teachers: findings from a large-scale study

Christin Lucksnat[1] · Eric Richter[1] · Sofie Henschel[2] · Lars Hoffmann[2] ·
Stefan Schipolowski[2] · Dirk Richter[1]

Abstract
The present study investigates whether alternatively and traditionally certified teachers differ in their teaching quality. We conducted doubly latent multilevel analysis using a large-scale dataset with student ratings on the quality of instruction provided by 1550 traditionally and 135 alternatively certified secondary school mathematics teachers in Germany. Findings show no evidence of differences in teaching quality between alternatively and traditionally certified teachers. Independent of the type of certification, novice teachers scored lower on classroom management than experienced teachers. Findings suggest that beginning teachers—whether alternatively or traditionally certified—need to be better prepared for classroom management before starting to teach.

Keywords Alternative certification · Teaching quality · Instructional quality · Lateral entrants · Student ratings

✉ Christin Lucksnat
christin.lucksnat@lk.brandenburg.de

Eric Richter
eric.richter@uni-potsdam.de

Sofie Henschel
sofie.henschel@iqb.hu-berlin.de

Lars Hoffmann
lars.hoffmann@iqb.hu-berlin.de

Stefan Schipolowski
stefan.schipolowski@iqb.hu-berlin.de

Dirk Richter
dirk.richter@uni-potsdam.de

[1] University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany

[2] Institute for Educational Quality Improvement (IQB), Unter den Linden 6, 10099 Berlin, Germany

1 Introduction

Studies show that teachers vary in their success in promoting student learning (Fauth et al., 2019; Hattie, 2009; Kyriakides et al., 2013). Students' learning outcomes depend on the quality of instruction they receive, which in turn depends on teachers' competence (Fauth et al., 2019; Kunter et al., 2013). Teachers' competencies comprise knowledge, attitudes, and motivational variables that form the basis for mastery-specific situations (Kunter et al., 2013). Teachers acquire teaching competence primarily through teacher education programs and through practical experience in the classroom (Kunter et al., 2013). Both teacher training and job experience appear to play crucial roles in the development of teachers' professional competence and in the quality of their teaching.

During the last two decades, growing numbers of individuals without traditional teacher training have entered the teaching profession through alternative certification programs in Europe and the USA (Coppe et al., 2021; Paniagua & Sanchez-Martí, 2018; Redding & Smith, 2016). This development has been driven in large part by teacher shortages (Sutcher et al., 2019). Forty US states recently reported teacher shortages in the areas of math, science, and special education (Sutcher et al., 2019). Most European countries are facing similar shortages (European Comission, 2018). Alternatively certified (AC) teachers offer an important means of overcoming the teacher shortages that have become a worldwide phenomenon (Sutcher et al., 2019; UNESCO, 2016). AC teachers are not, however, distributed equally among schools. Studies show that AC teachers often work at hard-to-staff schools since these schools have difficulties attracting and retaining highly qualified teachers (Podolsky et al., 2016; Richter et al., 2018). To avoid further disadvantaging students at these schools, it is important to ensure that teaching quality does not suffer due to the increased hiring of AC teachers at hard-to-staff schools.

The increasing number of teachers who completed alternative certification programs has raised questions about whether AC teachers provide the same teaching quality as their traditionally certified (TC) colleagues, since alternative certification programs are usually shorter than traditional ones (Darling-Hammond et al., 2005; Swanson & Ritter, 2018). An answer to this question would be of great value, since numerous studies have shown that teaching quality is an important prerequisite for students' motivational and cognitive learning outcomes (Allen et al., 2013; Ruzek et al., 2016). Moreover, the insights gleaned from this question could elucidate the assessment and accountability mechanisms within teacher education programs and inform the evolution of alternative certification pathways.. Prior studies on AC teachers' teaching quality have provided initial insights into how AC teachers organize and design their lessons (Jelmberg, 1996; Mccarty & Dietz, 2011; Miller et al., 1998). Most of these, however, used relatively small samples, focused on individuals coming from selected programs only, and lacked a theoretical framework for teaching quality (Jelmberg, 1996; Mccarty & Dietz, 2011; Miller et al., 1998). The present study addresses these limitations and investigates whether AC teachers provide the same teaching quality as their

traditionally certified counterparts. Using secondary data from a large-scale study conducted in Germany, we examined whether AC and TC mathematics teachers differed in three dimensions of teaching quality (i.e., classroom management, student support, and cognitive activation) using student ratings (Research Question 1). Furthermore, we investigated whether AC and TC teachers differ in teaching quality depending on their teaching experience (Research Question 2). In the following, we shortly describe the context of traditional and alternative teacher certification in Germany. After that, we introduce a precise definition of teaching quality and summarize previous findings on the role of teacher training and teaching experience in teaching quality.

2 Traditional and alternative teacher certification: the German context

In Germany, individuals who want to become teachers must complete two phases of teacher training. The first phase is a university-based teacher training with a bachelor's/master's model generally lasting 5 years, and students choose at least two subjects at the start of their teacher education program (Cortina & Hoover Thames, 2013). The first phase of teacher training consists primarily of content-related courses. However, there are also courses on pedagogical content knowledge and pedagogy as well practical phases in schools. In 2020, most of Germany's federal states implemented a semester-long internship during the master's program (Ulrich et al., 2020). The first university-based phase is followed by an in-school induction program lasting from 1 to 2 years (depending on the federal state). This second phase aims at preparing prospective teachers for their work in schools and focuses on pedagogical knowledge and pedagogical content knowledge (Cortina & Hoover Thames, 2013). The prospective teachers are assigned to a school in which they teach about 10 lessons per week. At the end of the induction phase, prospective teachers are required to pass an exam to obtain certification to teach at a German school (Cortina & Hoover Thames, 2013). This examination typically takes place at the end of the induction phase, often consolidated into a single day. During this session, the prospective teacher is required to conduct a lesson in each respective subject. An evaluative committee, predominantly comprising seasoned educators, appraises the teaching efficacy of the candidate. However, such protocols exhibit variations across federal states. In certain instances, candidates might also be mandated to deliver a presentation or submit a scholarly paper. Thus, there is no standardized test for all prospective teachers in Germany.

In response to the growing teacher shortage in Germany, more accelerated pathways into the teaching profession have been developed in addition to the traditional pathway. In contrast to traditional teacher training, which is highly structured and regulated by standards, alternative certification programs differ widely in length and are not required to meet standards of traditional teacher training (Driesner & Arndt, 2020). There are two general alternative pathways to teacher certification in Germany.

The first alternative pathway is the so-called Seiteneinstieg and is created for individuals who have a completed a master's degree in a field other than teaching but have not received one that qualifies them to teach two school subjects. Candidates receive on-the-job pedagogical training while working as teachers (Richter et al., 2022). Regrettably, the methodologies for on-the-job pedagogical training exhibit substantial variations not only between, but also within, federal states. It is noteworthy to mention that comprehensive data on these trainings cannot be procured for all 16 federal states. Consequently, Germany lacks a harmonized approach for the training of Seiteneinsteiger. The second alternative pathway is called "Quereinstieg" and is designed for individuals who have a completed a master's degree in a field other than teaching that relates to two school subjects. Individuals who meet this prerequisite are allowed to enter the induction phase (second phase of teacher training) together with traditionally trained teachers. The present paper covers AC teachers who enter the teaching profession through both pathways, the Seiteneinstieg and the Quereinstieg. The two pathways cannot be distinguished in our analyses because we do not have any data about the specific pathway teachers have taken. However, teachers coming from both pathways have in common that they did not complete the first phase of initial teacher training. This is the major difference between TC and AC teachers in this sample. Since TC and AC teachers differ in this important aspect, we believe that the type of certification could have an impact on the quality of teaching and thus on student achievement. Therefore, investigating the differences between TC and AC teachers could shed light on the accountability of teacher education programs.

3 Offering high-quality teaching: the role of teacher training

The provision of high-quality teaching, however, does not happen automatically. Teachers need specific competencies to meet the high demands of their profession and to be able to support the diverse students in their classrooms (Kunter et al., 2013). Hence, a central task of teacher education is to provide future teachers with opportunities to acquire the necessary competencies to meet their goals (Darling-Hammond, 2017; Tatto, 2021). Teachers' knowledge can be regarded as a key component of teachers' competence (Kleickmann et al., 2013). Studies have shown that teachers' knowledge has an effect on teaching quality and student learning (Baumert et al., 2010; Hill et al., 2005; Kunter et al., 2013; Fauth et al., 2019). In addition to knowledge, teachers' competencies also comprise attitudes and motivational variables (Kunter et al., 2013).

In light of the importance of teachers' knowledge for student learning, teacher education places great value on fostering teachers' knowledge (Kleickmann et al., 2013). To ensure high-quality teaching, a number of countries have implemented standards that determine what teachers should learn and be able to do (Darling-Hammond, 2017). These standards often refer to teachers' knowledge in the areas of content knowledge, pedagogical content knowledge, and pedagogical knowledge (see Darling-Hammond, 2021 for an overview). Studies showing that both teachers' content knowledge (Hill et al., 2015; Metzler & Woessmann, 2012) and their pedagogical

content knowledge (Keller et al., 2017) positively impact student achievement reflect the relevance of these areas of teacher knowledge. Moreover, pedagogical content knowledge (Kulgemeyer & Riese, 2018) and general pedagogical knowledge (König & Pflanzl, 2016) are also positively associated with teaching quality.

Teachers acquire their knowledge through different sources (Grossman, 1990). Studies highlight the importance of teacher education programs providing different formal learning opportunities (i.e., workshops and lectures) for the acquisition of knowledge (e.g., Kleickmann et al., 2013). Torbeyns et al. (2020), for instance, showed that second- and third-year students in teacher education programs outperformed first-year students in a test of mathematical pedagogical content knowledge. Moreover, in a study of students in teacher education programs preparing to teach French as a foreign language, Evens et al. (2017) showed that second and third-year students outperformed first-year students in tests of content knowledge, pedagogical content knowledge, and pedagogical knowledge. Another study by Morris and Hiebert, (2017) highlights the impact of teacher education programs on teachers' careers: Even after 5 to 6 years of teaching experience, mathematics teachers were more likely to use mathematical concepts in their lesson planning if they had the opportunity to learn these concepts during their teacher training. Hence, teacher training is essential in building a solid knowledge base and in promoting the development of skills that improve teachers' teaching quality and thus also students' learning outcomes.

At present, however, a steadily growing number of individuals are entering the teaching profession through alternative certification programs. Since these programs can be completed much more quickly than traditional training programs, they provide reduced opportunities to learn. While the number of AC teachers is rising, only a few studies to date have investigated differences in teaching quality between AC and TC teachers, and these have focused mainly on teachers in the USA and used quantitative data to investigate whether the type of certification was related to teaching quality. Most of these studies did not find significant differences in teaching quality between AC and TC teachers (Hill et al., 2015; Mccarty & Dietz, 2011; Miller et al., 1998), and only one found that TC teachers provided higher-quality teaching than AC teachers (Jelmberg, 1996).

It is noteworthy that previous studies on the teaching quality of AC and TC teachers assessed teaching quality quite differently. Some used teachers' own ratings (Miller et al., 1998), while others used principals' ratings (Jelmberg, 1996), school district administrators' ratings (Mccarty & Dietz, 2011), or experts' ratings of videotape lessons (Hill et al., 2015). Moreover, these studies specified teaching quality differently and did not always refer to established theories. Whereas some focused on specific practices (goal direction, feedback, appropriate constructive criticism, appropriate negative consequences; Miller et al., 1998) as indicators for teaching quality, others used more general specifications (e.g., instructional skills or instructional planning; Jelmberg, 1996). Hill et al. (2015) specified teaching quality with the model by Pianta and Hamre (2009) using the Classroom Assessment Scoring System (CLASS).

In sum, it appears that previous studies have disagreed on how teacher quality should be measured but appear to agree—with the exception of Jelmberg (1996)—that there is little difference in teacher quality between AC and TC teachers.

Moreover, most previous studies were conducted in the United States, and it thus remains unclear whether the findings are also valid for other countries.

4 The role of teaching experience in teaching quality

As noted above, most previous studies have shown that AC and TC teachers do not differ in their teaching quality, despite the shorter duration of AC teachers' training (Hill et al., 2015; Mccarty & Dietz, 2011; Miller et al., 1998). However, these studies did not take teaching experience as a possible moderator into account, which may influence teaching quality over and above the type of teacher training completed (Graham et al., 2020). It seems plausible that the teaching quality of novice teachers is lower than that of experienced teachers, because the first 3 years of teaching are a time of "survival and discovery," characterized by a sense of exhaustion, feeling overwhelmed, and struggling with student discipline (Huberman, 1989).

However, research on this assumption is inconclusive (e.g., Graham et al., 2020). A study by Stuhlman and Pianta (2009) did not find a significant relationship between years of teaching experience and teaching quality among first-grade teachers in the USA. A more recent study by Graham et al. (2020), which investigated whether teaching quality of third-grade teachers in Australia differed in relation to their teaching experience, reports comparable results. However, they found that transitioning teachers, that is, those with 4–5 years of teaching experience, performed significantly worse in classroom organization than novice teachers (0–3 years) and experienced teachers (more than 5 years). The authors explain this finding with the removal of initial support structures paired with an increase in workload and responsibilities (Graham et al., 2020). In contrast to the aforementioned studies, which found no significant differences between experienced and novice teachers, eye-tracking studies have identified differences in some aspects of classroom management. Huang et al. (2021a) measured novice teachers' attention distribution in the classroom and showed that, compared to experienced teachers, novice teachers were scanning narrower classroom areas and focusing more on individual students than on the whole class (Cortina et al., 2015).

In sum, the current state of research only partially supports the hypothesis that novice teachers differ significantly from experienced teachers. But it must also be mentioned that there is an inconsistency in the categorization used to describe categories of experience (Graham et al., 2020). In this study, we follow Huberman's approach and define novice teachers as those with less than 3 years of teaching experience.

Turning the focus to AC teachers, their unique pathways into the profession present a distinct landscape for understanding teaching quality and its development. A recent study by Matsko et al. (2021) showed that AC teachers feel less prepared for teaching than TC teachers. This could possibly be explained by shorter training programs with a lower number of academic classes. Due to the different certification process of AC teachers, they encounter specific challenges during their first years of teaching. Self-reports of AC teachers highlight difficulties in maintaining classroom management and time management and in creating opportunities for differentiation

(for an overview, see Baeten & Meeus, 2016; Haggard et al., 2006). They also report on challenges outside of instructional settings, such as in the organization of daily school routines, understanding educational law, self-management, and collaboration with parents (Richter et al., 2023). Although novice TC teachers also report facing challenges in classroom management (Chaplain, 2008; Voss et al., 2017), AC teachers express higher levels of concern (Baeten & Meeus, 2016). Teaching experience should therefore be also considered when examining differences in teaching quality between AC and TC teachers.

5 What teachers do matters: defining teaching quality

Educational research has searched extensively for characteristics of classroom instruction that positively predict student learning outcomes (e.g., Allen et al., 2013). Empirical studies have shown that scaffolding, teacher feedback, clarity of presentation, and adequate pacing can foster learning in the classroom (Hattie, 2009; Seidel & Shavelson, 2007). Over the last 20 years, research has shifted from looking at individual instructional practices or classroom characteristics to focusing on more general aspects of teaching quality (Borko, 2004; Seidel & Shavelson, 2007). A frequently used empirical framework for teaching quality suggested by Klieme et al. (2009) differentiates between the three dimensions *classroom management*, *student support,* and *cognitive activation.*

The dimension *classroom management* includes all actions taken by teachers to establish order and maximize the time for learning (Baumert et al., 2010; Doyle, 1985; Emmer & Stough, 2001). In order to achieve this goal, teachers first need to identify desirable student behaviors by communicating clear rules and establishing stable routines (Emmer & Stough, 2001; Praetorius et al., 2018). Second, teachers need to prevent disruptions and ensure the efficient use of time by monitoring the classroom (Kounin, 1970) and intervening immediately and effectively if necessary (Emmer & Stough, 2001; Kunter et al., 2007). By following these two key principals, teachers can provide a learning environment in which students can actively engage with subject matter (Brophy, 1999; Kunter et al., 2007). Studies have shown that teachers who are able to successfully manage their classrooms promote student learning significantly better than teachers lacking these skills (Fauth et al., 2014; Kyriakides et al., 2013; Lipowsky et al., 2009).

The second dimension, *student support*, includes all aspects related to the quality of social interactions and relationships between teachers and students, as well as among students (Fauth et al., 2014; Praetorius et al., 2018). Positive interactions and relationships can be encouraged through teacher feedback and dealing constructively with student errors and misconceptions, which positively impacts learning (Baumert et al., 2010; Fauth et al., 2014). By providing support to students, teachers show that they genuinely care about students and make an effort to understand their feelings and points of view (Wallace et al., 2016). Studies have shown that student support is closely related to students' motivational and emotional learning outcomes, such as their

interest, enjoyment, and anxiety (Fauth et al., 2014; Fauth et al., 2019; Kunter et al., 2013; Lazarides & Buchholz, 2019), but also to their achievement (Allen et al., 2013).

Finally, *cognitive activation* describes instruction in which teachers provide challenging tasks or problems that engage students in higher-order thinking processes (Baumert et al., 2010). Tasks should therefore connect students' prior knowledge with the exploration of new concepts and resolution of cognitive conflicts (Baumert et al., 2010; Praetorius et al., 2018). Students taught in classrooms with high cognitive activation were found to develop a deep conceptual understanding (Praetorius et al., 2018; Wallace et al., 2016). Teachers' cognitive activation in the classroom also positively predicts students' motivation (Fauth et al., 2014) as well as student achievement (Allen et al., 2013; Blazar, 2015; Kyriakides et al., 2013; Lipowsky et al., 2009; Seidel & Shavelson, 2007).

Teaching quality can be investigated from different perspectives. The perception of students has been receiving growing attention, and many studies have used student ratings as a measure of teaching quality (e.g., Lazarides & Buchholz, 2019; Lazarides et al., 2021; Kunter et al., 2013; Fauth et al., 2014, 2019; Aldrup et al., 2018; Wagner et al., 2016). These studies have demonstrated that student ratings are valid measures of teaching quality and as suitable as external observers or teachers to assess teaching quality (e.g., De Jong & Westerhof, 2000; Fauth et al., 2014, 2019; Kunter & Baumert, 2006). Studies show that student and teacher ratings are equally informative when assessing classroom management (Kunter & Baumert, 2006). Student ratings, however, are even more reliable than teachers' self-reports with regard to cognitive activation and student support (Kunter & Baumert, 2006). Moreover, student ratings of teaching quality predict students' academic engagement and motivational development better than observer ratings (Clausen, 2002; Maulana & Helms-Lorenz, 2016).

Whereas there are numerous studies using students' perceptions to assess teaching quality, more research is needed on differences in students' perceptions of teaching quality based on background variables (e.g., socioeconomic background, migration background) (Atlay et al., 2019). Studies indicate that students' background information should be included as control variables in the models. The socioeconomic status of the students, for example, is related to students' perceptions of teacher support. In particular, more affluent students tend to rate teachers' support lower compared to less affluent students (Atlay et al., 2019). A study by Wenger et al., (2020) reports similar findings, showing that the socioeconomic background and the migration background of students at the school level is predictive for teaching quality (cognitive activation, classroom management, student support). Moreover, students' prior achievement is related to a more positive perception of student support (Atlay et al., 2019; Wenger et al., 2020), classroom management, and cognitive activation (Wenger et al., 2020). Building on these results, students' background information should be controlled when investigating teaching quality.

6 The present investigation

In numerous countries, traditional teacher training programs fail to produce an adequate supply of teachers to meet the teacher shortage in schools. AC teachers present a valuable solution to address the global issue of teacher shortages (Sutcher et al., 2019; UNESCO, 2016). However, the increasing

number of AC teachers raises the question of whether they provide the same quality of teaching as TC teachers, given that alternative certification programs are shorter and less comprehensive than traditional teacher education programs. Therefore, the purpose of this study is to investigate the differences between AC and TC teachers in terms of teaching quality. In addition, the study also aims to investigate possible differences between AC and TC teachers based on their teaching experience. In the present study, we used the framework proposed by Klieme et al., (2009), which includes three dimensions of teaching quality (i.e., classroom management, student support, and cognitive activation), as a key model to examine differences in teaching quality between AC and TC teachers. We used dual latent multilevel analyses because we had a clustered data structure with students nested within classes.

Previous research comparing the teaching quality of AC and TC teachers is scarce and provides inconclusive results. Most of these studies have found no significant differences between AC and TC teachers in their teaching quality. The few studies to date that have investigated differences in teaching quality between AC and TC teachers assessed teaching quality differently, referred to small sample sizes, and mainly focused on US teachers, making the findings limited in their generalizability (e.g., Jelmberg, 1996; Mccarty & Dietz, 2011; Miller et al., 1998). To close this gap, we addressed two research questions. The first research question "Do AC and TC teachers differ in their teaching quality measured by classroom management, student support, and cognitive activation?" aims to explore general differences in teaching quality between the two groups of teachers in the dimensions of teaching quality proposed by Klieme et al. (2009). We assumed that students of TC teachers would rate teaching quality higher than students of AC teachers in all three dimensions. This assumption is based on the idea that TC teachers, due to their longer and more comprehensive traditional teacher training, have higher competencies in these three dimensions than AC teachers who have participated in shorter trainings. Furthermore, previous studies did not take teaching experience as a potential moderator of teaching quality into account. Therefore, we wanted to explore this research gap with our second research question "Do AC and TC teachers differ in their teaching quality measured by classroom management, student support, and cognitive activation depending on their teaching experience?" Classroom management has been identified as a major challenge for AC teachers, especially during their first few years of teaching (Baeten & Meeus, 2016; Haggard et al., 2006; Huberman, 1989). Given that novice TC teachers also report difficulties in classroom management (Voss et al., 2017), it seems possible that novice AC teachers might face even greater difficulties since they lack extensive teacher training. Because of this lack, we assumed that novice AC teachers obtain lower ratings of teaching quality than novice TC teachers. However, for experienced teachers, we expected to find no differences between AC and TC teachers since AC teachers might be able to catch up with TC teachers through teaching experience and the uptake of formal learning opportunities (e.g., professional development) as well as informal learning opportunities (sharing material or experiences with colleagues).

7 Method

7.1 Study design and sample

The data used in the present study were collected through the 2018 Institute for Educational Quality Improvement Trends in Student Achievement study (Stanat et al., 2019), a nationally representative large-scale assessment study that surveyed students' achievement in mathematics and science (biology, chemistry, physics) at the end of ninth grade in line with Germany's national educational standards. Accordingly, this is a secondary data analysis. The aim of the Trends in Student Achievement studies is to identify the extent to which students in Germany are achieving the national educational standards in specific subjects (e.g., mathematics, science, German, and English) and the end of fourth or ninth grade. The studies are conducted at periodic intervals linked to the implementation of international school performance studies and are mandatory in all German federal states (Stanat et al., 2019). In 2018, the latest data available at the time this manuscript was written, students in ninth grade were surveyed and tested in mathematics and science (Stanat et al., 2019). The analyses conducted as part of the present study were based on a subsample of 1,685 mathematics teachers teaching 19,004 students at 1150 schools. Therefore, an advantage of this study is that students can be matched with their teachers. The student survey included questions about students' sociodemographic backgrounds and the quality of teaching in their mathematics classes. The teacher survey included questions about teachers' gender, type of certification, and teaching experience.

As the students surveyed were in grade 9, they are between 14 and 15 years old. Among the students in the sample, 51% were male, 17% had a migration background (both parents born in country other than Germany) and the average Highest International Socio-Economic Index of Occupational Status (HISEI; Ganzeboom et al., 1992) was 51.5. The HISEI is a statistic frequently used in international large-scale studies to quantify the socio-economic status of the students, with higher HISEI scores indicating higher socio-economic status (minimum 10 points and maximum 90 points). It is calculated based on the occupations reported by both parents (Mahler & Kölm, 2019).

The teacher sample was selected from all secondary school types in Germany, which include schools in the academic track (*Gymnasium)* and the non-academic track (e.g., *Hauptschule*, *Realschule*). Academic-track schools offer students the opportunity to obtain university entrance qualifications (*Abitur*). Non-academic-track secondary schools prepare students mainly for vocational training (Cortina & Hoover Thames, 2013). Among the teachers in the sample, 72% were employed at non-academic-track schools and 28% at academic-track schools. This corresponds to the distribution of teachers in general education schools across Germany (Statista Research Department, 2023). On average, teachers were 46 years old ($SD = 11.6$) and had been working as teachers for 17 years ($SD = 13.0$). Fifty-five percent were female.

Teachers were categorized as AC and TC teachers based on the teacher training program they had completed. All teachers reporting that they had graduated from

Table 1 Descriptive statistics of sociodemographic and occupational characteristics of AC and TC teachers

	TC teachers		AC teachers		Group comparison		
	%		%		χ^2	p	OR
Gender (proportion female)	59.1		34.4		30.3	< 0.01	0.36
School track (proportion academic-track)	30.5		29.6		< 0.1	0.84	0.96
Novice teachers (≤ 3 years of teaching experience)	16.2		24.2		5.7	0.02	1.66
Experienced teachers (> 3 years of teaching experience)	83.8		75.8				
	M	SD	M	SD	F	p	d
Age (in years)	45.6	11.9	46.3	8.8	0.5	0.49	0.06
Novice teachers	30.0	4.3	39.7	8.6	106.2	< 0.01	1.9
Experienced teachers	48.6	10.4	48.5	7.8	0.1	0.92	0.01
Teaching experience (in years)	17.8	13.1	10.0	8.9	45.5	< 0.01	0.61

$N = 1685$ (AC teachers $n = 135$, TC teachers $n = 1550$)

a traditional teaching program were classified as TC teachers. The German teacher education system differs from other systems such as that in the USA. In Germany, students enroll in teacher education prior to the first semester of their bachelor's degree and must also complete a master's degree prior to obtaining their (traditional) teaching certification. Teachers who reported having graduated from a program that did not lead to a teaching degree were classified as AC teachers. In total, 51 teachers did not report their type of training and were excluded from the analyses. In total, 135 teachers were identified as AC teachers (8%) and 1550 were classified as TC teachers (92%). This proportion is close to the number of AC teachers hired in Germany in 2018 (13%) (KMK, 2019). Unfortunately, Germany does not provide statistics on all AC teachers currently in the education system, but only on newly hired AC teachers per year.

Using univariate analysis of variance and chi-square tests, we found that AC and TC teachers differed significantly in selected sociodemographic and occupational characteristics (Table 1). AC teachers were more likely to be male and less experienced than TC teachers. This is in line with previous research on AC and TC teachers in Germany (Lucksnat et al., 2022). We could also demonstrate that 15.3% of the TC teachers and 25.4% of AC teachers are classified as novices, having less than 3 years of teaching experience. There were no differences regarding school track and age. Upon distinguishing between novice and experienced teachers, we observed distinct age disparities between AC and TC teachers. Specifically, novice AC teachers were, on average, 10 years older than their counterparts who underwent traditional training. This age differential was not evident within the subset of experienced teachers. As a result, we decided to include the type of certification, teachers' gender, teaching experience, and school track as covariates in the analyses. Although there was no difference between AC and TC teachers regarding their school track, we introduced this variable as a covariate since previous studies showed that teachers and students vary across tracks (e.g., Baumert et al., 2010).

7.2 Measures

Teaching quality Students rated the teaching quality of their mathematics teachers in three dimensions: classroom management, student support, and cognitive activation. They rated *classroom management* with three items describing the degree of classroom disruptions (Baumert et al., 2009) (e.g., "In mathematics, classes are often disrupted."). We recoded the items, since we wanted higher ratings to indicate higher teaching quality. For greater precision, we henceforth refer to an *absence of classroom disruptions* instead of classroom management. We evaluated the items regarding their internal consistency (Cronbach's $\alpha = 0.91$). Furthermore, we computed intraclass correlations (ICC) to assess the reliability of the aggregated variables. First, we computed the proportion of variance attributable to the classroom level of analysis ($ICC_1 = 0.37$). Second, we calculated the reliability of the classroom aggregate regarding students' agreement within classes ($ICC_2 = 0.85$; calculated as described in Lüdtke et al., 2009). According to LeBreton and Senter (2008), a value of $ICC_1 > 0.05$ means that individual ratings are attributable to group membership and a value of $ICC_2 > 0.70$ is regarded as acceptable. Moreover, *student support* consisted of five items on individual support and guidance teachers provided in class (Cronbach's $\alpha = 0.87$, $ICC_1 = 0.24$, $ICC_2 = 0.75$, Baumert et al., 2009) (e.g., "Our teacher is interested in the learning progress of each student."). Finally, we used five items to measure cognitive activation (Cronbach's $\alpha = 0.78$, $ICC_1 = 0.13$, $ICC_2 = 0.60$; Baumert et al., 2009) (e.g., "Our teacher more often provides tasks where it's not just the arithmetic that counts, but above all the right approach."). Students rated all items on a 4-point Likert scale ranging from 1 (does not apply) to 4 (applies completely). Please see the Appendix 2 for an overview of all items.

Teacher characteristics Teacher characteristics included years of teaching experience, gender (male vs. female), school track (academic vs. non-academic), and teacher certification (AC teacher vs. TC teacher).

Student characteristics We obtained information on students' characteristics by a student questionnaire. Schools provided the information on students' gender and school track, whereas we assessed migration background based on the student questionnaire. We classified all students with both parents born in a country other than Germany as having a migration background. To quantify the socio-economic status of the students, we used the Highest International Socio-Economic Index of Occupational Status (HISEI; Ganzeboom et al., 1992), with higher HISEI scores indicating higher socio-economic status (minimum 10 points and maximum 90 points). We calculated the HISEI based on the occupations reported by both parents (Mahler & Kölm, 2019). Of the students in the sample, 49% were female, 17% had a migration background (both parents born in a country other than Germany) and the average Highest International Socio-Economic Index of Occupational Status (HISEI; Ganzeboom et al., 1992) was 51.5 (minimum 10 points and maximum 90 points).

Proficiency test Students took a standardized mathematics test that was developed by the Institute for Educational Quality Improvement (IQB) (Mahler et al., 2019). The tasks in the test focused on attainment of the national educational standards in mathematics defining those competencies that students are expected to achieve at the end of compulsory education. Using a multiple matrix sampling design, all students participating in the test received a test booklet including a subset of all 415 mathematics items used in the study. Multiple matrix design means that different test booklets contained some of the same items. The items were thus linked directly within a test booklet or indirectly across several test booklets. The different test booklets were distributed randomly to the students within classes. The items were presented in multiple choice or open response questions. The data processing can only be described in brief since it was a complex procedure (please see Mahler et al., 2019 for a detailed description of the process). Trained coders rated the answers as correct or incorrect. The data were then scaled to generate a difficulty parameter for each item. The item parameters obtained were used to determine students' proficiency scores. We derived proficiency scores for the students from a two-parameter item response theory (IRT) model as described in Becker et al. (2019). We used the first *plausible value* (PV) as an indicator of students' mathematical proficiency as a control variable in our analyses.[1] The reliability for this test was very good (Cronbach's $\alpha = 0.93$). The dataset analyzed during the current study is available from the corresponding author on reasonable request.

7.3 Statistical analysis

We conducted multilevel analyses since we had a clustered data structure, with students nested within classes. We used doubly latent models as proposed by Marsh et al. (2012) and estimated all models in M*plus* 8 (Muthén & Muthén, 1998–2017) using maximum likelihood estimation (MLE; Myung, 2003). To answer the first research question, we conducted a set of three multilevel regression models, with each dimension of teaching quality as a dependent variable (model 1a: absence of disruptions, model 1b: student support, model 1c: cognitive activation). The items for teaching quality were standardized ($M = 0$, $SD = 1$). We included students' gender, migration background, HISEI, and mathematics proficiency as covariates at the student level (level 1) as well as classroom-level aggregates for these covariates at the teacher level (level 2). Mathematics proficiency and HISEI were standardized on both levels and introduced as grand-mean-centered level-2 predictors. We included these student variables since previous studies have shown that students' achievement (Atlay et al., 2019; Sutton et al., 2021), migration background, and socioeconomic status (Wenger et al., 2020) are related to teaching quality. In addition, we introduced teachers' school track, gender, teaching experience (standardized), and type of certification at level 2. To answer the second research question, we slightly changed

---

[1] In order to check the robustness of our results, we estimated the models again with the second and third plausible values. All of the results remained the same. We conclude that using the first plausible value is suitable for indicating students' math proficiency.

Table 2 Descriptive statistics of teaching quality for different groups of teachers (manifest indicators)

	N	Absence of disruptions		Student support		Cognitive activation	
		M	SD	M	SD	M	SD
Total teacher sample	1685	2.68	0.57	2.87	0.42	2.83	0.25
TC teachers	1550	2.69	0.57	2.87	0.43	2.84	0.25
AC teachers	135	2.59	0.58	2.86	0.39	2.83	0.24
Novice TC teachers	244	2.61	0.51	2.94	0.40	2.84	0.26
Experienced TC teachers	1266	2.70	0.58	2.86	0.43	2.84	0.25
Novice AC teachers	32	2.40	0.54	2.84	0.40	2.76	0.25
Experienced AC teachers	100	2.66	0.59	2.87	0.39	2.85	0.24

Teaching quality was rated by students on a scale from 1 = fully disagree to 4 = fully agree. The values represent classroom aggregates. AC = alternatively certified, TC = traditionally certified, novice = up to 3 years of teaching experience, experienced = more than 3 years of teaching experience

the models. Instead of teachers' experience and type of certification, we included three dummy-coded variables according to the career stage model of Huberman, (1989). We had to exclude 43 (2.6%) cases from this grouping since their teaching experience was missing: novice AC teachers with up to 3 years of teaching experience ($n = 32$; 1.9%), novice TC teachers with up to 3 years of teaching experience ($n = 244$; 14.5%), experienced AC teachers with more than 3 years of teaching experience ($n = 100$; 5.9%), and experienced TC teachers with more than 3 years of teaching experience ($n = 1266$; 75.1%). Thus, to answer the second research question, we conducted a second set of models with novice TC teachers as the reference group (model 2a–2c), and in a third set of models, we considered experienced TC teachers as the reference group (model 3a–3c). These two reference groups make it possible to compare novice AC teachers with novice TC teachers as well as experienced AC teachers with experienced TC teachers. We also checked the proportion of missing data for each variable. On average, 2.7% of all responses were missing for teachers and 7.7% for students. To account for the missing data, we opted for full information maximum likelihood (FIML) estimation in all our analyses, a recommended method compared to alternatives such as listwise or pairwise deletion (Graham, 2012). Because FIML uses information from the existing data to obtain realistic estimates for the missing values, aggregates for the classroom-level could be calculated without missing values but with the estimated values. Thus, there were no missing data for classroom-level aggregates of individual student data.

8 Results

Prior to addressing the main research questions, we present a concise summary of the descriptive statistics pertaining to the measures of teaching quality, namely, absence of disruptions, student support, and cognitive activation. These statistics

have been detailed comprehensively in Table 2 and are segmented not just for AC and TC teachers, but further distinguished between novice and experienced teachers within each category. A cursory examination of the data reveals only marginal differences between AC and TC teachers, especially notable in the dimension of absence of disruptions. When examining the subsets of novice and experienced teachers, nuanced differences become apparent. Specifically, there are slight variations in Absence of disruptions for both AC and TC teachers, in Student support among TC teachers, and in Cognitive activation for AC teachers. These initial observations set the stage for a more profound exploration and interpretation of the results that ensue. These preliminary insights lay the groundwork for the subsequent in-depth analyses that follow.

To investigate whether AC and TC teachers differ in their teaching quality (research question 1), we conducted doubly latent multilevel analyses (Table 3). The results show that the teachers' type of certification does not explain a significant amount of variance in the absence of disruptions ($\beta = 0.08$, $p > 0.05$), in student support ($\beta = 0.05$, $p > 0.05$), and in cognitive activation ($\beta = 0.04$, $p > 0.05$). In other words: students of AC teachers rated teaching quality as high as students of TC teachers, controlling for the variables shown in Table 3. Teaching experience is positively related to the absence of disruptions ($\beta = 0.05$, $p < 0.01$) but negatively related to student support ($\beta = -0.07$, $p < 0.01$), and there is no significant effect for cognitive activation ($\beta = -0.01$, $p > 0.05$). This means that students with experienced teachers report fewer disruptions to their mathematics lessons but also feel less individually supported than students of novice teachers. We also found that school track is negatively related to teaching quality (Table 2). Students at non-academic track schools reported lower teaching quality than students in schools with an academic track. We report all regression models without control variables for research question 1 in Appendix 1.

To answer the second research question, we estimated two additional sets of multilevel regression models excluding teachers' type of certification and teaching experience as single variables but introduced dummy variables combining both type of certification and teaching experience. In the first set of regression models, we used novice TC teachers as the reference group. The results in Table 4 show that novice AC teachers obtained significantly lower ratings of student support compared to novice TC teachers ($\beta = -0.20$, $p < 0.05$). There were no significant differences for absence of disruptions ($\beta = -0.21$, $p > 0.05$) or for cognitive activation ($\beta = -0.13$, $p > 0.05$) between novice AC and novice TC teachers. Moreover, students of experienced TC teachers reported feeling less individually supported ($\beta = -0.12$, $p < 0.01$) than students of novice TC teachers. In contrast, students of experienced TC teachers reported fewer classroom disruptions than students of novice TC teachers ($\beta = 0.13$, $p < 0.01$).

In the second set of additional regression models for research question 2, we used experienced TC teachers as the reference group. The results in Table 5 show that there is no difference in teaching quality between experienced AC and experienced TC teachers. Students of novice AC teachers reported more classroom disruptions than students of experienced TC teachers ($\beta = -0.34$, $p < 0.01$).

Table 3 Results of the multilevel regression analyses: teacher characteristics as predictors of teaching quality (unstandardized regression coefficients, standard error, model fit information)

Predictors	Absence of disruptions model 1a	Student support model 1b	Cognitive activation model 1c
	β (*SE*)	β (*SE*)	β (*SE*)
Student level			
Mathematics proficiency	0.02 (0.01)	0.11** (0.01)	0.12** (0.01)
Migration background (1 = both parents born in a country other than Germany)	0.05* (0.02)	0.06** (0.02)	0.01 (0.02)
Socioeconomic status	0.01 (0.01)	−0.02* (0.01)	−0.02** (0.01)
Gender (1 = female)	−0.04** (0.01)	0.02 (0.01)	0.07** (0.01)
Classroom level			
Student characteristics			
Mathematics proficiency	0.23** (0.03)	−0.02 (0.02)	0.04* (0.02)
Migration background	0.09 (0.07)	0.03 (0.06)	0.05 (0.05)
Socioeconomic status	−0.04 (0.02)	−0.02 (0.02)	−0.02 (0.01)
Gender	0.12 (0.07)	−0.04 (0.06)	−0.04 (0.05)
Teacher characteristics			
Gender	−0.01 (0.03)	0.03 (0.06)	−0.03 (0.02)
Teaching experience	0.05** (0.02)	−0.07** (0.01)	−0.01 (0.01)
School track (1 = academic track)	−0.14** (0.05)	−0.20** (0.04)	−0.08* (0.04)
Certification (1 = traditional certification)	0.08 (0.06)	0.05 (0.04)	0.04 (0.03)
R^2 (within)	0.00 (0.00)	0.03 (0.01)	0.03 (0.01)
R^2 (between)	0.11 (0.02)	0.12 (0.02)	0.02 (0.01)
χ^2	158.04	640.92	786.39
Df	26	62	62
p	<0.01	<0.01	<0.01
CFI	0.99	0.98	0.95
RMSEA	0.01	0.03	0.03
SRMR between	0.02	0.06	0.09
SRMR within	0.01	0.02	0.02

*$p < 0.05$, **$p < 0.01$

9 Discussion

Two research questions guided this paper. First, we investigated whether AC and TC teachers differ in their teaching quality. Second, we explored whether experienced and novice AC as well TC teachers differ in their teaching quality. It is important to note that our data focus on teachers and students as they teach and learn mathematics. Regarding the first research question, we found that AC and TC teachers did not differ significantly in any of the three dimensions. Regarding the second research

Table 4 Results of the multilevel regression analyses: interaction of teaching experience and certification as predictors of teaching quality with novice TC teachers as reference group (unstandardized regression coefficients, standard error, model fit information)

Predictor	Absence of disruptions model 2a	Student support model 2b	Cognitive activation model 2c
	β (SE)	β (SE)	β (SE)
Student level			
Mathematics proficiency	0.02 (0.01)	0.11** (0.01)	0.12** (0.01)
Migration background (1 = both parents born in a country other than Germany)	0.05* (0.02)	0.06** (0.02)	0.01 (0.02)
Socioeconomic status	0.01 (0.01)	−0.02* (0.01)	−0.02** (0.01)
Gender (1 = female)	−0.04** (0.01)	0.02 (0.01)	0.07** (0.01)
Classroom level			
Student characteristics			
Mathematics proficiency	0.22** (0.03)	−0.01 (0.02)	0.04* (0.02)
Migration background	0.06 (0.07)	0.08 (0.06)	0.06 (0.05)
Socioeconomic status	−0.04* (0.02)	−0.01 (0.02)	−0.02 (0.01)
Gender	0.12 (0.07)	−0.04 (0.06)	−0.04 (0.05)
Teacher characteristics			
Gender	0.00 (0.03)	0.01 (0.02)	−0.04 (0.02)
School track (1 = academic track)	−0.12* (0.05)	−0.22** (0.04)	−0.08* (0.03)
Novice AC teachers	−0.21 (0.11)	−0.20* (0.08)	−0.13 (0.08)
Experienced AC teachers	0.07 (0.06)	−0.08 (0.05)	−0.03 (0.04)
Experienced TC teachers	0.13** (0.04)	−0.12** (0.03)	0.02 (0.03)
R^2 (within)	0.00 (0.00)	0.03 (0.00)	0.03 (0.01)
R^2 (between)	0.11 (0.02)	0.10 (0.02)	0.03 (0.01)
χ^2	146.46	636.43	781.09
df	28	66	66
p	< 0.01	< 0.01	< 0.01
CFI	0.99	0.98	0.95
RMSEA	0.02	0.02	0.03
SRMR between	0.02	0.05	0.09
SRMR within	0.01	0.02	0.02

AC = alternatively certified, TC = traditionally certified, novice = up to 3 years of teaching experience, experienced = more than 3 years of teaching experience.

question, differential findings emerged. First, we found that students of novice TC teachers feel more individually supported than students of novice AC but also experienced TC teachers. Second, students of novice AC and TC teachers reported more frequent classroom disruptions than students of experienced TC teachers. Experienced AC and TC teachers did not differ in the absence of classroom disruptions.

Table 5 Results of the multilevel regression analyses: interaction of teaching experience and certification as predictors of teaching quality with experienced TC teachers as reference group (unstandardized regression coefficients, standard error, model fit information)

Predictor	Absence of disruptions model 3a	Student support model 3b	Cognitive activation model 3c
	β (SE)	β (SE)	β (SE)
Student level			
Mathematics proficiency	0.02 (0.01)	0.11** (0.01)	0.12** (0.01)
Migration background (1 = both parents born in a country other than Germany)	0.05* (0.02)	0.06** (0.02)	0.01 (0.02)
Socioeconomic status	0.01 (0.01)	−0.02* (0.01)	−0.02** (0.01)
Gender (1 = female)	−0.04** (0.01)	0.02 (0.01)	0.07** (0.01)
Classroom level			
Student characteristics			
Mathematics proficiency	0.22** (0.03)	−0.01 (0.02)	0.04* (0.02)
Migration background	0.06 (0.07)	0.08 (0.06)	0.06 (0.05)
Socioeconomic status	−0.04* (0.02)	−0.01 (0.02)	−0.02 (0.01)
Gender	0.12 (0.07)	−0.04 (0.06)	−0.04 (0.05)
Teacher characteristics			
Gender	0.00 (0.03)	0.01 (0.02)	−0.04 (0.02)
School track (1 = academic track)	−0.12* (0.05)	−0.22** (0.04)	−0.08* (0.03)
Novice AC teachers	−0.34** (0.11)	−0.08 (0.08)	−0.10 (0.08)
Experienced AC teachers	−0.06 (0.06)	0.04 (0.05)	0.00 (0.04)
Novice TC teachers	−0.13** (0.04)	0.12** (0.03)	0.02 (0.03)
R^2 (within)	0.00 (0.00)	0.03 (0.01)	0.03 (0.01)
R^2 (between)	0.11 (0.02)	0.10 (0.02)	0.03 (0.01)
χ^2	146.46	636.43	781.09
df	28	66	66
p	< 0.01	< 0.01	< 0.01
CFI	0.99	0.98	0.95
RMSEA	0.02	0.02	0.03
SRMR between	0.02	0.05	0.09
SRMR within	0.01	0.02	0.02

AC = alternatively certified teachers, TC = traditionally certified teachers, novice = up to 3 years of teaching experience, experienced = more than 3 years of teaching experience

*p < 0.05, **p < 0.01

9.1 Teacher training does not make a difference in teaching quality

In most previous studies, the type of certification did not make a difference in teaching quality (Hill et al., 2015; Mccarty & Dietz, 2011; Miller et al., 1998). The present study's findings are in line with most of the research since we also found no differences between AC and TC teachers in the quality of their teaching. That is, the

AC teachers in our sample provided similar teaching quality to TC teachers. From our point of view, there are two possible explanations for this finding.

First, traditional teacher education programs do not prepare prospective teachers in a way that would bring differences in teaching quality between AC and TC teachers to light. This result challenges the accountability and the quality of today's design of traditional teacher education programs. Our findings provide a starting point for answering questions about the effectiveness of traditional teacher education programs. However, they are not sufficient to provide a comprehensive answer. We need more evidence from future studies on the quality of traditional teacher education programs. Grossman and Pupik Dean (2019) noted that university teacher education programs in the USA have been criticized for focusing more on building knowledge than on teaching core practices. This also applies to the German context (Terhart, 2019). This argument is further substantiated by studies reporting a reality shock experienced by novice TC teachers when optimistic ideals developed during university teacher education collapse and the reality of teaching in the classroom sets in (Dicke et al., 2015; Voss & Kunter, 2020). In response to these and similar findings, there has been a shift in traditional teacher education programs from focusing solely on knowledge building to placing more attention on core practices (Grossman & Pupik Dean, 2019). According to Ball and Forzani (2009), core practices are tasks and activities that are fundamental for beginning teachers to carry out important instructional responsibilities (e.g., leading a discussion of solutions to a problem, probing students' answers). However, implementing core practices in teacher education programs brings with it an array of challenges. One major problem is developing common ground and reaching a consensus around the practice of teaching (Grossman & Pupik Dean, 2019). Teacher education programs attempt to address the importance of core practices by providing opportunities to practice in school, yet they often fail to focus on certain core practices (Forzani, 2014). Teacher education programs tend to be structured more around broader domains of teaching (e.g., content methods or educational psychology), which means that aspiring teachers have few opportunities to try out certain core practices (Forzani, 2014). According to Forzani, (2014), the length of time that trainees spend in the field plays a minor role, since beginning teachers "might spend months in student teaching […] and never learn how to lead a productive class discussion, for example, because this practice has not been clearly identified as something to learn" (p. 358).

A second explanation for the equivalent levels of teaching quality might be that the measures we used in this study were not able to shed light on differences in teaching quality caused by completing or not completing traditional teacher training. As Grossman & Pupik Dean, (2019) have pointed out, most teacher education programs mainly focus on building knowledge, and several studies have shown the positive link between teacher education programs and the development of knowledge (e.g., Torbeyns et al., 2020; Evens et al., 2017; Kleickmann et al., 2013). However, our measures did not focus on teachers' knowledge but on their behavior in the classroom or the consequences of this behavior.

Despite the fact that there is no difference in teaching quality between AC and TC teachers, we were able to show a strong relationship between the absence of classroom disruptions and students' mathematical performance at the classroom level, which highlights the importance of classroom management as a necessary and core skill for teachers. More importantly, the current results show that this relationship is stronger when compared to cognitive activation and student support. In light of our findings, classroom management courses should be developed and made a standard part of teacher education programs for both TC and AC teachers. Universities should empirically evaluate these courses to make their effects visible so that other universities can compare their results. Schools could also support the improvement of classroom management skills by creating opportunities for novice and experienced teachers to work together.

9.2 Teaching experience does make a difference in teaching quality

Although we could not find significant differences in teaching quality regarding teachers' type of certification, we did find that teaching experience makes a difference in the absence of classroom disruptions and in the provision of student support. We found that students of experienced TC teachers reported fewer classroom disruptions than students of novice teachers. This result applies to AC teachers as well as to TC teachers. We see different possible explanations for these findings.

The difficulties in managing a classroom appropriately may be a result of the neglect of teaching classroom management techniques in teacher education programs (Thiel et al., 2020). In addition, there are limited opportunities for prospective teachers to practice classroom management before they take over their own classroom as a novice teacher. While internships are available for prospective teachers, these environments may not always replicate authentic classroom situations. The presence of mentor or supervising teachers can alter the classroom dynamics, potentially limiting the hands-on experiences available to preservice teachers.

Secondly, novice teachers have difficulties recognizing disruptions in the classroom. Previous studies have shown that experienced teachers are able to monitor the whole classroom more evenly and are able to prevent disruptions in the classroom (Cortina et al., 2015; Huang et al., 2021a; Wolff et al., 2016). In contrast, novice teachers tend to focus more on individual students and narrower areas (Emmer & Gerwels, 2006; Huang et al., 2021a; Thiel et al., 2020; van den Bogert et al., 2014). Moreover, they either recognize the disruptive behaviors too late or not at all, and if they do, they have difficulties focusing on both the source of the disruption and the rest of the classroom (van den Bogert et al., 2014).

Another possible explanation for the finding that experienced teachers can create a learning environment with fewer disruptions than novice teachers is that teachers acquire pedagogical knowledge and skills on the job over time. A study by Kyndt et al. (2016) showed that informal learning activities, such as sharing ideas, experimenting, and collaborating with colleagues, have a positive impact on the development of classroom management strategies. Over time, teachers acquire more sophisticated and contextualized knowledge of the events taking place in the classroom, and thus learn to better apply their pedagogical knowledge to specific types of events and students (Wolff et al., 2016). This seems to apply equally to AC and TC teachers.

Moreover, we found that novice TC teachers had significantly higher ratings for student support compared to experienced TC teachers but also compared to novice AC teachers. This finding can possibly be explained by the age of the different groups. On average, novice TC teachers are 30 years old. In contrast, novice AC teachers are 40 years old, and experienced AC as well as TC teachers are 49 years old on average. It can be assumed that students feel more supported by younger teachers because the age difference is smaller. Studies in higher education have shown that students perceive younger educators to be more helpful and interested in their work than older educators (Clayson, 2020; Wilson et al., 2014). These finding may support our hypothesis that younger teachers, in our case the novice TC teachers, offer more comprehensive student support than older teachers. Another way of looking at this difference is to use Huberman's (1989) model of the career stages. In this model, Huberman (1989) distinguishes five different stages, of which stage four may be the most important in explaining the lower level of social support for experienced teachers. Stage four, which includes teachers with 19 to 30 years of experience, is a phase in which teachers have reached a level of competence and stability in their careers. They focus more on refining their teaching practice rather than trying new approaches in the classroom. This can result in a less dynamic teaching style, which can affect their ability to engage and support pupils who could benefit from more diverse teaching methods. In addition, the increasing intergenerational differences between teachers and their students become more pronounced at this stage (Huberman, 1989).

9.3 School tracking does make a difference in teaching quality

In addition to identifying a significant disparity in teaching quality based on teachers' experience, we also found that teaching quality differs between school tracks. Students at non-academic track schools reported significantly lower teaching quality than students at academic track schools. This is in line with previous research that showed that academic track schools are more successful than non-academic track schools in teaching mathematics (Becker et al., 2007). Furthermore, studies highlight the differences in teaching quality between tracks. A study by Donaldson et al. (2017) showed that student receive less support in lower tracks than in higher tracks.

9.4 Strengths and limitations of the study

The present paper has several methodological and content-related strengths. First, we used large-scale data with a sample of secondary schools from all of Germany's federal states including not only novice but also experienced AC teachers of a high-need subject. Second, we utilized different sources of data, such as teacher and student surveys as well as proficiency tests. To answer our research questions, we applied multilevel analyses using a doubly latent modelling approach and took control variables such as teachers' and students' characteristics as well as student achievement into account. In addition to these methodological aspects, our paper identified possible predictors of teaching quality, thus expanding the existing body of research not only on teaching quality but also on AC teachers. In addition to previous studies on teaching quality of AC teachers, we expanded our focus and took both certification and teaching experience into account.

Despite these strengths of our study, it also has some limitations that should be considered when interpreting the results. First, our data on teaching quality were only reported by students about their mathematics teachers. It is therefore reasonable to assume that the results may only have implications for mathematics teachers. However, we believe that the results of this study could have possible implications for other subjects, as we used a framework that covers teaching quality in general. Even though our framework covers more general aspects of teaching quality, it can be regarded as a specific framework of teaching quality that has been used widely in other studies in the German context (e.g., Fauth et al., 2019) since it often considers only selected aspects of the three dimensions of teaching quality. For example, for classroom management, we only had data on the absence of disruptions in the classroom. However, other aspects of classroom management such as effective time use or monitoring could be taken into account in future studies (Voss et al., 2022). To get a more detailed look at teaching quality, future research could draw on other theoretical frameworks. For example, Praetorius et al. (2020) developed a new framework combining models of the three dimensions of teaching quality (Klieme et al., 2009) with models that are frequently used in other countries. Their framework proposes seven dimensions of teaching quality including additions of subject-specific as well as generic aspects of teaching. Since this framework also includes aspects of practice as well as the selection and thematization of content, it goes beyond the three dimensions of teaching quality. However, this framework must be tested for validity in future studies.

Third, we have no information about the teachers' previous knowledge or beliefs, making it impossible to explain the non-existent differences with reference to other teacher aspects besides the ones we used. We believe that it is not unlikely that AC teachers already acquired relevant skills in previous university programs that are relevant for teaching quality. This leads us to the third limitation, namely that we do not have any information about the

contents of the training AC teachers had completed before starting to teach. Although this information would provide more detailed insight into the extent to which specific programs promote teaching quality, it would be difficult to compile an overview of the programs' contents since the certification programs for AC teachers in Germany differ between and within federal states, and the content of the various programs is not always publicly accessible (Driesner & Arndt, 2020). Finally, we cannot make causal claims since our research design is cross-sectional, and we do not know how long the teachers had taught the classes in our sample. Future research on AC teachers' teaching quality should therefore also ask how long teachers had spent teaching the specific class or learning group.

9.5 Implications for research and practice

Based on our findings, we can draw several implications for both research and practice. Firstly, future studies should identify possible reasons for the non-existent differences in classroom management and cognitive activation between novice AC and TC teachers. One possible starting point could be to investigate how and to what extent teacher education programs promote competencies in these dimensions of teaching quality. Another approach could be to identify other prior experience or knowledge of AC teachers (e.g., work as a substitute teacher) that might have a positive impact on their teaching quality.

Our research shed light on how AC and TC teachers differ in their teaching quality. However, little is known about teachers who are not certified and therefore teach without any pedagogical training. Consequently, a subsequent avenue for research would be the examination of these "lay teachers." An exploration into how they contrast with TC and AC teachers in facets such as competence, motivation, and teaching quality is warranted.

Thirdly, we found differences between novice teachers and experienced teachers in classroom management, and also found that TC novice teachers provide better student support. Previous studies have highlighted the positive relationship between student achievement and classroom management as well as cognitive activation and the positive relationship between student interest and student support (e.g., Fauth et al., 2014). Based on these findings, we first argue that research should focus more on what students learn when taught by novice teachers compared to experienced teachers, since the two groups differed in their teaching quality. These studies should be longitudinal and take students' prior knowledge as well as AC teachers' knowledge and motivation into account. In addition to cognitive student outcomes, studies could also focus on students' motivational (e.g., interest in the subject taught) or emotional outcomes (e.g., anxiety) when taught by AC teachers. These studies would give more detailed insight into the effect of AC teachers on cognitive and motivational as well as emotional student outcomes.

In terms of practical implications, a relevant question that needs to be addressed is whether current teacher education programs in Germany provide the desired added value in terms of classroom management, given that the novice teachers in our sample received lower classroom management ratings than experienced teachers. The findings from our study can serve as a starting point for examining the effectiveness of these programs and identifying areas where improvements may be needed. Lessons learned from experienced teachers' classroom management practices should be incorporated into teacher education curricula to better prepare beginning teachers. This could include linking of effective classroom management strategies of experienced teaches to the development of specific training modules for prospective teachers. Thus, we suggest that teacher education programs and programs for alternative certification should include more opportunities to learn classroom management skills, since classroom management is known to be difficult for beginning teachers (Thiel et al., 2020). Teacher education programs, however, often lack training to promote these specific skills (Thiel et al., 2020). However, some universities have decided to offer training using different approaches to support classroom management skills (for an overview of possible approaches, see Christofferson & Sullivan, 2015). Whereas the video-based training developed by Thiel et al. (2020) has the goal of strengthening "preservice teachers' skills in noticing, reasoning, and generating strategies to deal with disruptions in the classroom" (p. 2), other forms of training use virtual reality classrooms to teach classroom management skills (Huang et al., 2021b).

10 Conclusion

The present study investigated differences in teaching quality between alternatively and traditionally certified teachers. We identified two overarching results. First, we found no differences in the absence of classroom disruptions, in the provision of student support, or in cognitive activation in relation to teachers' certification status. Second, with a more detailed approach, we showed that students of novice AC and TC teachers reported more classroom disruptions compared to students of more experienced TC teachers. Therefore, both traditionally and alternatively certified beginning teachers need more opportunities to acquire classroom management skills, both during their certification programs and on the job. Finally, our results challenge the accountability of traditional teacher education programs and thus call for an investigation of their quality.

Appendix 1

Table 6 Results of the multilevel regression analyses: teacher certification as predictor of teaching quality without covariates (unstandardized regression coefficients, standard error, model fit information)

Predictors	Absence of disruptions model 1a	Student support model 1b	Cognitive activation model 1c
	β (SE)	β (SE)	β (SE)
Classroom level			
Teacher characteristics			
Certification (1 = traditional certification)	0.12* (0.05)	0.01 (0.04)	0.03 (0.03)
R^2 (within)	-	-	-
R^2 (between)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
χ^2	51.36	320.24	315.43
Df	4	18	18
p	< 0.01	< 0.01	< 0.01
CFI	0.99	0.99	0.98
RMSEA	0.03	0.03	0.03
SRMR between	0.03	0.05	0.04
SRMR within	0.01	0.02	0.02

*$p < 0.05$, **$p < 0.01$

Appendix 2

Items for the three dimensions of teaching quality.

Classroom management

What is it like in your mathematics lessons?

a) In mathematics, classes are often disrupted.
b) In mathematics, there's constant loud chatter.
c) In mathematics class, nonsense is continually done.

Student support

How much do the following statements apply to your mathematics lessons?

a) Our teacher is interested in the learning progress of each student.
b) Our teacher gives us extra support when we need help.
c) Our teacher supports us in our learning.
d) Our teacher explains something until we understand it.
e) Our teacher gives us the opportunity to express our opinions.

Cognitive activation

How does your math teacher go about teaching?

a) Our teacher assigns various tasks on a topic to ascertain if we've truly grasped the mathematical concept.
b) Our teacher modifies tasks to show what we've understood.
c) Our teacher more often provides tasks where it's not just the arithmetic that counts, but above all the right approach.
d) When we practice in math, we often apply what we have learnt to other things.
e) Among the practice tasks in math class, there are often tasks that genuinely indicate whether we've understood something.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aldrup, K., Klusmann, U., Lüdtke, O., Göllner, R., & Trautwein, U. (2018). Social support and classroom management are related to secondary students' general school adjustment: A multilevel structural equation model using student and teacher ratings. *Journal of Educational Psychology, 110*(8), 1066–1083. https://doi.org/10.1037/edu0000256

Allen, J. P., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—secondary. *School Psychology Review, 42*(1), 76–98. https://doi.org/10.1080/02796015.2013.12087492

Atlay, C., Tieben, N., Fauth, B., & Hillmert, S. (2019). The role of socioeconomic background and prior achievement for students' perception of teacher support. *British Journal of Sociology of Education, 40*(7), 970–991. https://doi.org/10.1080/01425692.2019.1642737

Baeten, M., & Meeus, W. (2016). Training second-career teachers: A different student profile, a different training approach? *Educational Process: International Journal, 5*(3), 173–201. https://doi.org/10.12973/edupij.2016.53.1

Ball, D., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511. https://doi.org/10.1177/0022487109348479

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., Krauss, S., Kunter, M., Löwen, K., Neubrand, M., & Tsai, Y.-M. (2009). Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV). Dokumentation der Erhebungsinstrumente. Max-Planck-Institut für Bildungsforschung.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Becker, M., Lüdtke, O., Trautwein, U., & Baumert, J. (2007). Leistungszuwachs in Mathematik Evidenz für einen Schereneffekt im mehrgliedrigen Schulsystem? [Achievement Gains in Mathematics: Evidence for Differential Achievement Trajectories in a Tracked School System?]. *Zeitschrift Für Pädagogische Psychologie, 20*(4), 223–242. https://doi.org/10.1024/1010-0652.20.4.233

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*(3), 682–699. https://doi.org/10.1037/a0027608

Becker, B., Weirich, S., Mahler, N., & Sachse, K. A. (2019). Testdesign und Auswertung des IQB-Bildungstrends 2018: Technische Grundlagen [test design and evaluation of the IQB Trends in Student Achievement study 2018: Technical Basics]. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich, & S. Henschel (Eds.), *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe im zweiten Ländervergleich* (pp. 411–426). Waxmann.

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review, 48*(4), 16–29. https://doi.org/10.1016/j.econedurev.2015.05.005

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Brophy, J. (1999). Perspectives of classroom management. In J. H. Freiberg (Ed.), *Beyond behaviorism: Changing the classroom management paradigm* (pp. 43–56). Alyn & Bacon.

Chaplain, R. P. (2008). Stress and psychological distress among trainee secondary teachers in England. *Educational Psychology, 28*(2), 195–209. https://doi.org/10.1080/01443410701491858

Christofferson, M., & Sullivan, A. L. (2015). Preservice teachers' classroom management training: A survey of self-reported training experiences, content coverage, and preparedness. *Psychology in the Schools, 52*(3), 248–264. https://doi.org/10.1002/pits.21819

Clausen, M. (2002). Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstukt - und Kriteriumsvalidität [Teaching quality: A question of perspective? Empirical analyses of agreement, construct and criterion validity]. Waxmann.

Clayson, D. E. (2020). Student perception of instructors: The effect of age, gender and political leaning. *Assessment & Evaluation in Higher Education, 45*(4), 607–616. https://doi.org/10.1080/02602938.2019.1679715

Coppe, T., März, V., Coertjens, L., & Raemdonck, I. (2021). Transitioning into TVET schools: An exploration of second career teachers' entry profiles. *Teaching and Teacher Education, 101* https://doi.org/10.1016/j.tate.2021.103317.

Cortina, K. S., & Hoover Thames, M. (2013). Teacher Education in Germany. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), Mathematics teacher education: Vol. 8. Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project (pp. 49–62). Springer.

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education, 13*(2), 389–403. https://doi.org/10.1007/s10763-014-9610-5

Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education, 40*(3), 291–309. https://doi.org/10.1080/02619768.2017.1315399

Darling-Hammond, L. (2021). Defining teaching quality around the world. *European Journal of Teacher Education*, 1–14 https://doi.org/10.1080/02619768.2021.1919080.

Darling-Hammond, L., Holtzmann, D. J., Gatlin, S. J., & Vasquez Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*(42), 1–51.

De Jong, R., & Westerhof, K. J. (2000). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*(1), 51–85. https://doi.org/10.1023/A:1011402608575

Dicke, T., Elling, J., Schmeck, A., & Leutner, D. (2015). Reducing reality shock: The effects of classroom management skills training on beginning teachers. *Teaching and Teacher Education, 48*, 1–12. https://doi.org/10.1016/j.tate.2015.01.013

Donaldson, M. L., LeChasseur, K., & Mayer, A. (2017). Tracking instructional quality across secondary mathematics and English Language Arts classes. *Journal of Educational Change, 18*(2), 183–207. https://doi.org/10.1007/s10833-015-9269-x

Doyle, W. (1985). Recent research on classroom management. *Journal of Teacher Education, 36*(3), 31–35. https://doi.org/10.1177/002248718503600307

Driesner, I., & Arndt, M. (2020). Die Qualifizierung von Quer- und Seiteneinsteiger*innen. Konzepte und Lerngelegenheiten im bundesweiten Überblick. [The qualification of teachers without a teaching license]. *DDS Die Deutsche Schule, 2020*(4), 414–427. https://doi.org/10.31244/dds.2020.04.05

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103–112. https://doi.org/10.1207/S15326985EP3602_5

Emmer, E. T., & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In M. C. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 407–437). Lawrence Erlbaum Associates.

European Comission. (2018). Teaching careers in Europe: Access, progression and support. *Eurydice report*. Publications Office of the European Union.

Evens, M., Elen, J., & Depaepe, F. (2017). Effects of opportunities to learn in teacher education on the development of teachers' professional knowledge of French as a foreign language. *Journal of Advances in Education Research*, *2*(4). https://doi.org/10.22606/jaer.2017.24007

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fauth, B., Decristan, J., Decker, A.-T., Büttner, G., Hardy, I., Klieme, E., & Kunter, M. (2019). The effects of teacher competence on student outcomes in elementary science education: The mediating role of teaching quality. *Teaching and Teacher Education, 86*, 102882. https://doi.org/10.1016/j.tate.2019.102882

Forzani, F. M. (2014). Understanding "core practices" and "practice-based" teacher education. *Journal of Teacher Education, 65*(4), 357–368. https://doi.org/10.1177/0022487114533800

Ganzeboom, H. B., de Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), 1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Graham, J. W. (2012). Missing data. Analysis and design. Springer. https://doi.org/10.1007/978-1-4614-4018-5.

Graham, L. J., White, S. L., Cologon, K., & Pianta, R. C. (2020). Do teachers' years of experience make a difference in the quality of teaching? *Teaching and Teacher Education, 96*, 103190. https://doi.org/10.1016/j.tate.2020.103190

Grossman, P. L. (1990). *The making of a teacher*. Teachers College Press.

Grossman, P., & Pupik Dean, C. G. (2019). Negotiating a common language and shared understanding about core practices: The case of discussion. *Teaching and Teacher Education, 80*, 157–166. https://doi.org/10.1016/j.tate.2019.01.009

Guill, K., Lüdtke, O., & Köller, O. (2017). Academic tracking is related to gains in students' intelligence over four years: Evidence from a propensity score matching study. *Learning and Instruction, 47*, 43–52. https://doi.org/10.1016/j.learninstruc.2016.10.001

Haggard, C., Slostad, F., & Winterton, S. (2006). Transition to the school as workplace: Challenges of second career teachers. *Teaching Education, 17*(4), 317–327. https://doi.org/10.1080/1047621060 1017410

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

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406. https://doi.org/10.3102/00028312042002371

Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching. *AERA Open, 1*(4), 233285841561770. https://doi.org/10.1177/2332858415617703

Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2021a). Teachers' professional vision in action. *Zeitschrift Für Pädagogische Psychologie*, 1–18. https://doi.org/10.1024/1010-0652/a000313

Huang, Y., Richter, E., Kleickmann, T., Wiepke, A., & Richter, D. (2021b). Classroom complexity affects student teachers' behavior in a VR classroom. *Computers & Education, 163*, 104100. https://doi.org/10.1016/j.compedu.2020.104100

Huberman, M. (1989). The professional life cycles of teachers. *Teacher College Record, 91*(1), 31–58.

Jelmberg, J. (1996). College-based teacher education versus state-sponsored alternative programs. *Journal of Teacher Education, 47*(1), 60–66. https://doi.org/10.1177/0022487196047001010

Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching, 54*(5), 586–614. https://doi.org/10.1002/tea.21378

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge. *Journal of Teacher Education, 64*(1), 90–106. https://doi.org/10.1177/0022487112460398

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.

KMK. (2019). Einstellung von Lehrkräften 2018. [Recruitment of Teachers 2018]. KMK. https://www.kmk.org/fileadmin/Dateien/pdf/Statistik/Dokumentationen/Dok_218_EvL_2018.pdf.

König, J., & Pflanzl, B. (2016). Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality. *European Journal of Teacher Education, 39*(4), 419–436. https://doi.org/10.1080/02619768.2016.1214128

Kounin, J. (1970). *Discipline and group management in classrooms (Discipline and group management in classrooms)*. Rinehart & Winston.

Kulgemeyer, C., & Riese, J. (2018). From professional knowledge to professional performance: The impact of CK and PCK on teaching quality in explaining situations. *Journal of Research in Science Teaching, 55*(10), 1393–1418. https://doi.org/10.1002/tea.21457

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research, 9*(3), 231–251. https://doi.org/10.1007/s10984-006-9015-7

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction, 17*(5), 494–509. https://doi.org/10.1016/j.learninstruc.2007.09.002

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

Kyndt, E., Gijbels, D., Grosemans, I., & Donche, V. (2016). teachers' everyday professional development. *Review of Educational Research, 86*(4), 1111–1150. https://doi.org/10.3102/0034654315627864

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Lazarides, R., & Buchholz, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction, 61*, 45–59. https://doi.org/10.1016/j.learninstruc.2019.01.001

Lazarides, R., Fauth, B., Gaspard, H., & Göllner, R. (2021). Teacher self-efficacy and enthusiasm: Relations to changes in student-perceived teaching quality at the beginning of secondary education. *Learning and Instruction, 73*, 1–10. https://doi.org/10.1016/j.learninstruc.2020.101435

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19*(6), 527–537. https://doi.org/10.1016/j.learninstruc.2008.11.001

Lucksnat, C., Richter, E., Schipolowski, S., Hoffmann, L., & Richter, D. (2022). How do traditionally and alternatively certified teachers differ? A comparison of their motives for teaching, their well-being, and their intention to stay in the profession. *Teaching and Teacher Education, 117*, 1–13. https://doi.org/10.1016/j.tate.2022.103784

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology, 34*(2), 120–131. https://doi.org/10.1016/j.cedpsych.2008.12.001

Mahler, N., & Kölm, J. (2019). Soziale Disparitäten. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich, & S. Henschel (Eds.), *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe im zweiten Ländervergleich* (pp. 265–294). Waxmann.

Mahler, N., Schipolowski, S., & Weirich, S. (2019). *Anlage, Durchführung und Auswertung des IQB-Bildungstrends 2018* [Creation, implementation and evaluation of the IQB Education Trend 2018]. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich, & S. Henschel (Eds.), *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe im zweiten Ländervergleich* (pp. 99–124). Waxmann.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Matsko, K. K., Ronfeldt, M., & Nolan, H. G. (2021). How different are they? Comparing teacher preparation offered by traditional, alternative, and residency pathways. *Journal of Teacher Education*, 002248712110159 https://doi.org/10.1177/00224871211015976.

Maulana, R., & Helms-Lorenz, M. (2016). Observations and student perceptions of the quality of preservice teachers' teaching behaviour: Construct representation and predictive quality. *Learning Environments Research, 19*, 335–357. https://doi.org/10.1007/s10984-016-9215-8

Mccarty, W. L., & Dietz, D. (2011). Alternative teacher certification: The case for transition to teaching. *Journal of Applied Learning in Higher Education, 3*, 45–57.

Metzler, J., & Woessmann, L. (2012). The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. *Journal of Development Economics, 99*(2), 486–496. https://doi.org/10.1016/j.jdeveco.2012.06.002

Miller, J. W., McKenna, M. C., & McKenna, B. A. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education, 49*(3), 165–176. https://doi.org/10.1177/0022487198049003002

Morris, A. K., & Hiebert, J. (2017). Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *American Educational Research Journal, 54*(3), 524–567. https://doi.org/10.3102/0002831217695217

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (7th ed.). Muthén & Muthén.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*(1), 90–100. https://doi.org/10.1016/S0022-2496(02)00028-7

Paniagua, A., & Sanchez-Martí, A. (2018). Early career teachers: Pioneers triggering innovation or compliant professionals? *OECD Education Working Papers*(190), 1–52.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. https://doi.org/10.3102/0013189X09332374

Podolsky, A., Kini, A., Bishop, J., & Darling-Hammond, L. (2016). Solving the teacher shortage: How to attract and retain excellent educators. Learning Policy Institute.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of Three Basic Dimensions. *ZDM Mathematics Education, 50*(3), 407–426. https://doi.org/10.1007/s11858-018-0918-4

Praetorius, A.-K., Rogh, W., & Kleickmann, T. (2020). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft, 48*(3), 303–318. https://doi.org/10.1007/s42010-020-00072-w

Redding, C., & Smith, T. M. (2016). Easy in, easy out: Are alternatively certified teachers turning over at increased rates? *American Educational Research Journal, 53*(4), 1086–1125. https://doi.org/10.3102/0002831216653206

Richter, D., Marx, A., & Zorn, D. (2018). *Lehrkräfte im Quereinstieg: sozial ungleich verteilt?* [Teachers in lateral entry: socially unequal distribution?]. Bertelsmann Stiftung. https://doi.org/10.11586/2018048

Richter, E., Lucksnat, C., Redding, C., & Richter, D. (2022). Retention intention and job satisfaction of alternatively certified teachers in their first year of teaching. *Teaching and Teacher Education, 114.* https://doi.org/10.1016/j.tate.2022.103704

Richter, E., Klusmann, U., & Richter, D. (2023). Berufliche Herausforderungen und Beanspruchungserleben von Lehrkräften im Seiteneinstieg: Die Rolle der Mentorinnen und Mentoren [Professional Challenges and Perceived Strain of Alternatively Certified Teachers: The Role of Mentors]. *Erziehung und Unterricht, 173*(7/8), 605–615.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction, 42*, 95–103. https://doi.org/10.1016/j.learninstruc.2016.01.004

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Stanat, P., Schipolowski, S., Mahler, N., Weirich, S., & Henschel, S. (Eds.). (2019). *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe im zweiten Ländervergleich* [IQB Education Trend 2018: Maths and science skills at the end of secondary school in the second country comparison]. Waxmann.

Statista Research Department. (2023). Anzahl der Lehrkräfte an allgemeinbildenden Schulen in Deutschland im Schuljahr 2021/2022 nach Schulart. [Number of teachers at general education schools in Germany in the school year 2021/2022 by school type]. https://de.statista.com/statistik/daten/studie/162263/umfrage/anzahl-der-lehrkraefte-nach-schularten/.

Stuhlman, M. W., & Pianta, R. C. (2009). Profiles of educational quality in first grade. *The Elementary School Journal, 109*(4), 323–342. https://doi.org/10.1086/593936

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2019). Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Education Policy Analysis Archives, 27*(35), 1–40. https://doi.org/10.14507/epaa.27.3696

Sutton, E., Brown, J. L., Lowenstein, A. E., & Downer, J. T. (2021). Children's academic and social-emotional competencies and the quality of classroom interactions in high-needs urban elementary schools. *Contemporary Educational Psychology, 66*, 101975. https://doi.org/10.1016/j.cedpsych.2021.101975

Swanson, E., & Ritter, G. (2018). Using classroom observations and student surveys to evaluate alternatively certified teachers: A case study of the Arkansas Teacher Corps. *Journal of School Leadership, 28*, 642–670.

Tatto, M. T. (2021). Professionalism in teaching and the role of teacher education. *European Journal of Teacher Education, 44*(1), 20–44. https://doi.org/10.1080/02619768.2020.1849130

Terhart, E. (2019). Teacher education in Germany. In E. Terhart (Ed.), Oxford Research Encyclopedia of Education. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.377.

Thiel, F., Böhnke, A., Barth, V. L., & Ophardt, D. (2020). How to prepare preservice teachers to deal with disruptions in the classroom? Differential effects of learning with functional and dysfunctional video scenarios. *Professional Development in Education*, 1–15. https://doi.org/10.1080/19415257.2020.1763433.

Torbeyns, J., Verbruggen, S., & Depaepe, F. (2020). Pedagogical content knowledge in preservice preschool teachers and its association with opportunities to learn during teacher training. *ZDM Mathematics Education, 52*(2), 269–280. https://doi.org/10.1007/s11858-019-01088-y

Ulrich, I. Klingbiel, F., Bartels, A., Staab, R., Scherer, S & Gröschner, A. (2020). How does the practical semester in the teacher training program affect students? A systematic review. In I. Ulrich & Gröschner, A. (eds.), Praxissemester im Lehramtsstudium in Deutschland: Wirkungen auf

Studierende [The practical semester in the teacher training program in Germany: effects on students] (pp. 1–66). Springer VS.

UNESCO. (2016). The world needs almost 69 million new teachers to reach the 2030 education goals (UIS Fact Sheet n. 39).

van den Bogert, N., van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education, 37*, 208–216. https://doi.org/10.1016/j.tate.2013.09.001

Voss, T., & Kunter, M. (2020). "Reality shock" of beginning teachers? Changes in teacher candidates' emotional exhaustion and constructivist-oriented beliefs. *Journal of Teacher Education, 71*(3), 292–306. https://doi.org/10.1177/0022487119839700

Voss, T., Wagner, W., Klusmann, U., Trautwein, U., & Kunter, M. (2017). Changes in beginning teachers' classroom management knowledge and emotional exhaustion during the induction phase. *Contemporary Educational Psychology, 51*, 170–184. https://doi.org/10.1016/j.cedpsych.2017.08.002

Voss, T., Zachrich, L., Fauth, B., & Wittwer, J. (2022). The same yet different? Teaching quality differs across a teacher's classes, but teachers with higher knowledge make teaching quality more similar. *Learning and Instruction, 80*, 1–11. https://doi.org/10.1016/j.learninstruc.2022.101614

Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology, 108*(5), 705–721. https://doi.org/10.1037/edu0000075

Wallace, T. L., Kelcey, B., & Ruzek, E. (2016). What can student perception surveys tell us about teaching? Empirically testing the underlying structure of the tripod student perception survey. *American Educational Research Journal, 53*(6), 1834–1868. https://doi.org/10.3102/0002831216671864

Wenger, M., Gärtner, H., & Brunner, M. (2020). To what extent are characteristics of a school's student body, instructional quality, school quality, and school achievement interrelated? *School Effectiveness and School Improvement, 31*(4), 548–575. https://doi.org/10.1080/09243453.2020.1754243

Wilson, J. H., Beyer, D., & Monteiro, H. (2014). Professor age affects student ratings: Halo effect for younger teachers. *College Teaching, 62*(1), 20–24. https://doi.org/10.1080/87567555.2013.825574

Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science, 44*(3), 243–265. https://doi.org/10.1007/s11251-016-9367-z