AERA Open
January-March 2020, Vol. 6, No. 1, pp. 1–17
DOI: 10.1177/2332858419897884
Article reuse guidelines: sagepub.com/journals-permissions
© The Author(s) 2020. http://journals.sagepub.com/home/ero

The Predictive Validity of Teachers' Personality, Cognitive and Academic Abilities at the End of High School on Instructional Quality in Germany: A Longitudinal Study

Janina Roloff
Uta Klusmann
IPN – Leibniz Institute for Science and Mathematics Education
Oliver Lüdtke
IPN – Leibniz Institute for Science and Mathematics Education
Centre for International Student Assessment
Ulrich Trautwein
Hector Research Institute of Education Sciences and Psychology, University of Tübingen

The present study investigated to what degree teachers' personality traits, cognitive abilities, and academic abilities (high school GPA), which were assessed at the end of high school as well as grades from the first and second state examinations of teacher education, predict later instructional quality. The study combined ratings of instructional quality from 3,768 German secondary school students and their 113 teachers with up to 10 years between assessment of teachers' personality traits, cognitive abilities, as well as academic abilities at the end of high school and student ratings of instructional quality. Multilevel regression analyses revealed that agreeableness, high school GPA, and the second state examination grade predicted teachers' instructional quality.

Keywords: *classroom management, longitudinal studies, school/teacher effectiveness, teacher characteristics*

A GROWING body of research indicates that teacher quality—is the most important in-school factor for student achievement and motivation—and thus their educational attainment (Hattie, 2009; Rivkin, Hanushek, & Kain, 2005). It has been found that high-quality teachers influence their students' achievement and motivation through their ability to provide high instructional quality (Hill, Blazar, & Lynch, 2015; Kunter, Klusmann, et al., 2013). The question of how to improve teachers' instructional quality has recently received much attention from researchers and politicians all around the world (Klassen et al., 2017; Schleicher, 2016; Wayne & Youngs, 2003). One prominent assumption is that the recruitment and selection of teacher candidates with favorable characteristics before entering teacher education in terms of personality traits, cognitive and academic abilities is an important prerequisite in the attempt to raise later instructional quality (Auguste, Kihn, & Miller, 2010; Kelly & Northrop, 2015; Klassen, Durksen, Rowett, & Patterson, 2014). This assumption is supported by empirical findings across various occupational groups, which have shown that personality and cognitive and academic abilities are valid predictors of occupational success even over long time periods (Barrick, Mount, & Judge, 2001; Kuncel, Hezlett, & Ones, 2004; Poropat, 2009; Richardson, Abraham, & Bond, 2012; Schmidt & Hunter, 2004; Spengler et al., 2015; Trapmann, Hell, Hirn, & Schuler, 2007). Despite some attempts to raise teachers' instructional quality through selection of teacher candidates based on specific characteristics before entering teacher education, there is only little empirical evidence for the predictive validity of such characteristics for later instructional quality (Klassen & Kim, 2018; Rimm-Kaufman & Hamre, 2010).

Using data of 3,768 German secondary school students and longitudinal data on their 113 teachers, the main aim of the present study was to investigate the predictive validity of teachers' personality traits as well as cognitive and academic abilities prior to entering teacher education for teachers' instructional quality as rated by their students up to 10 years later.

Instructional Quality as Central Indicator of Teachers' Occupational Performance

Instructional quality has been described as a mediating process between teacher characteristics and student outcomes, for example, student achievement and motivation

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

(Kunter, Klusmann, et al., 2013; Rimm-Kaufman & Hamre, 2010; Seidel & Shavelson, 2007). Thus, favorable student outcomes depend among other factors on teachers' ability to provide high-quality instruction. Reviewing previous international findings on instructional quality, we decided to set the focus on the following aspects of instructional quality, which have been shown to be empirically associated to students learning and motivation: efficient classroom management with low disturbance levels and maximized time use (Emmer & Stough, 2001; Kounin, 1970; Pianta & Hamre, 2009), cognitively challenging instruction offered to the students (Baumert et al., 2010; Hill, Kapitula, & Umland, 2011; Klieme, Pauli, & Reusser, 2009), formation of a supportive social environment in which students feel secure and personally valued (Mainhard, Oudman, Hornstra, Bosker, & Goetz, 2018; Ryan & Powelson, 2014; Wang & Eccles, 2012), and an appropriate instructional tempo that neither bores students nor goes beyond students' capabilities (Berliner, 1990).

Following Kunter, Klusmann, et al. (2013), one line of research on the improvement of teachers' instructional quality highlights the importance of teachers' entry characteristics—cognitive and psychosocial qualities which can already be assessed prior to applicants' entering of teacher education (Kennedy, Ahn, & Choi, 2008; Yeh, 2009). Many researchers and politicians have further argued that teachers' instructional quality and student outcomes can best be raised through recruitment and selection of those individuals who show favorable personality traits, cognitive and academic abilities prior to entering teacher education (Auguste et al., 2010; Klassen et al., 2017).

Teachers' Personality, Cognitive and Academic Abilities, and Instructional Quality

For decades, teachers' personality traits—dispositional explanations of human behavior—have been hypothesized to contribute to teachers' instructional quality and student outcomes (Barr, 1952; Dodge, 1943; Duckworth, Quinn, & Seligman, 2009; Getzels & Jackson, 1963; Haberman, 1993, 1995; Klassen & Tze, 2014; Start, 1966). In contemporary research, the five-factor model of personality is used as a standard model to describe specific personality traits (John, Naumann, & Soto, 2008; McCrae & Costa, 2008). The model comprises five personality traits, often referred to as the Big Five: (1) Neuroticism (people with high scores on this trait can be described as being anxious, hostile, depressive, self-conscious, impulsive, and vulnerable); (2) Extraversion (warm, gregarious, assertive, active, excitement seeking, and positive); (3) Openness to new experiences (curious, creative, inventive, and imaginative); (4) Agreeableness (altruistic, sympathetic, trustworthy, and nurturing); and (5) Conscientiousness (organized, punctual, goal oriented, and honest). Empirical findings have shown that the Big Five (particularly conscientiousness and emotional stability) are valid predictors of occupational success across various occupational groups even over long time periods (Barrick et al., 2001; Spengler et al., 2015). It might reasonably be assumed that the Big Five also predict teachers' occupational behavior, such as instructional quality. For example, in line with findings indicating that agreeableness is related to job performance for jobs that involve interpersonal interactions (Mount, Barrick, & Stewart, 1998), it is likely to expect that a teacher who is agreeable might be more able to establish a supportive social environment in which students feel secure and personally valued. Up to now, only few studies investigated links between teachers' Big Five personality traits and their teaching quality or student outcomes (De Jong et al., 2014; Emmerich, Rock, & Trapani, 2006; see also Kim, Jörg, & Klassen, 2019; Klassen & Tze, 2014). Most recently, Kim et al. (2019) found in their meta-analysis significantly positive associations between student ratings of teaching quality and extraversion ($r = .32$, $p < .05$), conscientiousness ($r = .19$, $p < .05$), and openness ($r = .17$, $p < .05$). Moreover, Bastian, McCord, Marks, and Carpenter (2017) found conscientiousness to be significantly associated with higher teacher value-added estimates and higher evaluation ratings in first-year public school teachers.

Next to personality traits, teachers' cognitive and academic abilities are often presumed to be important prerequisites in the attempt to raise later instructional quality (e.g., Wayne & Youngs, 2003). Kennedy et al. (2008) proposed the hypothesis that "the best teachers are bright, well-educated people who are smart enough and thoughtful enough to figure out the nuances of teaching in the process of doing it" (p. 1250). Kunter, Klusmann, et al. (2013) called this the *bright-person hypothesis*. Consequently, the cognitive abilities (also often referred to as intelligence, see Ackerman & Heggestad, 1997) and general academic ability prior to entering teacher education, as indicated by high school GPA or school achievement test scores (according to Kunter, Klusmann, et al., 2013), are presumed to be important factors for teachers' instructional quality. However, up to now, empirical support for this hypothesis is inconclusive, with international research on cognitive abilities (as measured with standardized ability tests) and general academic ability showing, if at all, weak associations to professional success in the teaching profession (Aloe & Becker, 2009; Rockoff, Jacob, Kane, & Staiger, 2011; Yeh, 2009; Zumwalt & Craig, 2005). For example, in their meta-analysis, Aloe and Becker (2009) found only a weak association between teachers' verbal ability (a specific aspect of cognitive abilities) and teaching effectiveness, primarily indicated by student achievement or in terms of principal or researcher ratings of teacher behavior. Yeh (2009) compared the cost-effectiveness of two approaches to improve student performance in math and reading: (1) employing high-ability teachers (employing

only those teachers with a minimum score of 1,000 in the Scholastic Achievement Test, as an indicator for their general academic ability, and raising teacher salaries) and (2) the implementation of systems that provide formative assessment feedback to students and teachers. He found employing high-ability teachers to be much less cost-effective in improving student achievement. Moreover, Kunter, Klusmann, et al. (2013) found no association between German secondary school teachers' high school GPA and their instructional quality or student outcomes in mathematics. However, it is not yet possible to say whether this also holds for German secondary school teachers with other subjects.

Most of the studies reported up until now have a weakness in their design when it comes to the question of whether teachers' instructional quality can be raised through recruitment and selection of individuals with specific entry characteristics to the teaching profession, as teacher characteristics were assessed after entering teacher education and the teaching profession (e.g., Bastian et al., 2017; De Jong et al., 2014; Emmerich et al., 2006; Kim, Dar-Nimrod, & MacCann, 2018; Rockoff et al., 2011; Tok & Morali, 2009). Although the Big Five describe personality traits that are relatively stable across time (Meeus, Van de Schoot, Klimstra, & Branje, 2011; Roberts & Del Vecchio, 2000) possible effects of socialization might take place during teacher education (cf. a systematic review of personality trait change: Roberts et al., 2017).

Teacher Education and Instructional Quality

Though internationally there are already some attempts to raise teachers' instructional quality through the selection of teacher candidates based on specific characteristics before entering teacher education, in Germany, admission to the teaching profession depends on grades from teacher education. Teacher education in Germany is more than in other European countries or the United States divided into two consecutive phases (for a detailed description on German teacher education, see Cortina & Thames, 2013). The first phase is a teacher education program at university. This university training offers primarily formal and also informal (e.g., peer learning) learning opportunities for subject-specific content knowledge (CK) and subject-specific pedagogical content knowledge (PCK) in two subjects as well as subject-unspecific psychological-pedagogical knowledge (PPK). Teacher candidates' studies for the academic track last about nine semesters and focus on CK in their two teaching subjects. The studies for the nonacademic track last about seven semesters and focus on PCK and PPK. Both groups graduate with the first state examination. The average grade of the first state examination (in the following first exam grade) is based on teacher candidates' performance in oral or written examinations during the university training as well as the grade of a final thesis. Unlike in the United States, teacher education in Germany is highly standardized by the requirements of this examination. Passing the first state examination is obligatory to continue with the second part of teacher education.

The second part of teacher education is a monitored in-school induction program. This part lasts from 18 months to 2 years and ends with the second state examination. During this induction program, teacher candidates attend some courses on subject-specific pedagogical content and subject-unspecific psychological-pedagogical content (about 1 to 1.5 days per week). However, they have to teach lessons and gradually take on responsibility for their classes under the supervision of a mentor teacher for 3.5 to 4 days per week. The average grade from the second state examination (in the following second exam grade) is based on several examination parts: teachers' performance in the lessons during training, a final thesis, at least one examination lesson for each subject taught, oral examinations, and sometimes also evaluations by the school. In Germany, admission to the teaching profession depends on teacher candidates' first and second exam grades.

International findings have already shown the predictive validity of grades from teacher education for teaching quality and favorable student outcomes. In their review on teacher characteristics and student achievement gains, Wayne and Youngs (2003) stated that teachers' college ratings were positively related to student achievement gains and positive effects of grades, coursework, and certification were reported for mathematics. Moreover, D'Agostino and Powers (2009) found in their meta-analysis that preservice teachers' GPA during teacher education predicted later teaching performance (e.g., principal rating, student achievement, student evaluation) with an average effect size of $d = .25$.

At the same time, university education grades were found to be significantly associated with cognitive abilities, general academic ability prior to entering university education, and personality traits across various professions (Kuncel et al., 2004; Poropat, 2009; Richardson et al., 2012; Trapmann et al., 2007; Westrick, Le, Robbins, Radunzel, & Schmidt, 2015). Given these findings, the question remains, in what way a joint consideration of entry characteristics and grades from teacher education affect their predictive validity for teaching quality. To the best of our knowledge, there is no study until now to investigate at the same time associations between teachers' entry characteristics, teachers' grades in the first and second state examination, and instructional quality.

The Present Investigation

Our main aim was to examine the predictive validity of teachers' personality traits as well as cognitive and

academic abilities prior to entering teacher education for their later instructional quality (Research Question 1). Based on Kennedy et al. (2008) and previous research (Aloe & Becker, 2009; Kim et al., 2019; Kunter, Klusmann, et al., 2013; Mount et al., 1998; Yeh, 2009), we expected modest positive predictive validity of cognitive abilities and general academic ability as well as agreeableness, extraversion, conscientiousness, and openness for teachers' instructional quality.

Furthermore, we set out to explore in what way grades from teacher education (Research Question 2a) and a joint consideration of teachers' entry characteristics and grades from teacher education (Research Question 2b) have predictive validity for teachers' instructional quality. We expected teachers' grades from teacher education to have substantial predictive validity for teachers' instructional quality. Furthermore, we expected grades from teacher education to remain statistically significant predictors for instructional quality after controlling for their entry characteristics, whereas we expected the predictive validity of teachers' entry characteristics to decrease after controlling for teacher education grades (Research Question 2b). Finally, we explored whether grades from teacher education act as a mediator for the relationship between teachers' entry characteristics and later instructional quality (see Figure 1).

We decided to use student ratings to assess teachers' instructional quality as teachers' self-reports may be compromised and biased due to teaching ideals and self-serving strategies (Wubbels, Brekelmans, & Hooymayers, 1992), and observer ratings are rather cost and time intensive and observers can only observe very short time periods (Hill et al., 2015; cf. Praetorius, Lenske, & Helmke, 2012). Moreover, research shows that teacher behavior is assessed well when student ratings are aggregated at the classroom level and that the quality of student ratings is related to data from external observers (De Jong & Westerhof, 2001; Downer, Stuhlman, Schweig, Martínez, & Ruzek, 2015).

Method

Sample

This study complements data from the longitudinal study TOSCA (Köller, Watermann, Trautwein, & Lüdtke, 2004; Trautwein, Neumann, Nagy, Lüdtke, & Maaz, 2010) with an add-on study on teachers' instructional quality. TOSCA follows the educational and personality development of two cohorts of students from secondary school to tertiary education and an occupational career. The time frame of the current study is displayed in Figure 2. Starting in 2002, the first cohort consisted of 4,730 students in 149 randomly selected upper secondary schools in Germany. The second cohort began in 2006 and included 6,177 students from 158 schools. The participation rate at the student level was more than 80% for both cohorts. At the first measurement point (Time 1), the students (who were in their final year of upper secondary schooling) completed various achievement tests and personality questionnaires. Participants then completed postal questionnaires every second year, that is, Time 2 (the second measurement point) was in 2004 for the first cohort and in 2008 for the second cohort, including questions about their current vocational situation and personality questionnaires. In the current study, we used data on cognitive abilities, general academic ability, and personality traits from Time 1.

The add-on study used the information on the last known current vocational situation of the first cohort participants who participated at least once in 2006, 2008, and 2010 (N = 2,312), as well as of second cohort participants who participated at least once in 2008 and 2010 (N = 3,048). Subsuming both cohorts, 4,514 TOSCA participants declared at least once that they were enrolled in or had already finished university education. Of these, 464 (10.3%) were identified as becoming or being teachers, which is comparable to the overall proportion of students enrolled in teacher education in Germany (Statistisches Bundesamt, 2004). In 2013, postal questionnaires were sent to them in order to collect information on their professional career paths, grades from teacher education, and to assess their instructional quality as rated by students in two of their classes. The Time 1 assessments were reviewed by the ministry of education and cultural affairs of the state of Baden-Württemberg (the state where the assessment took place). Back then, the ministry took sole responsibility for reviewing research ethics/privacy issues in all state-wide research studies that took place in schools. It also approved the follow-up contacting of participants. Furthermore, we obtained ethical approval/permission to administer the instructional quality questionnaire to students from all ministries of the states in which our participants were teaching in 2013. In sum, 174 participants responded (response rate: 37.7%). Comparing respondents and nonrespondents' personality traits and cognitive and academic abilities, the respondents differed from nonrespondents in neuroticism, openness to new experiences, agreeableness, and conscientiousness. However, the magnitude of these differences was indicative of only moderate selectivity effects ($.19 \leq d \leq .30$). Student ratings of their teachers' instructional quality were sent by 133 teachers. We excluded 20 teachers from elementary and special needs schools, resulting in a total N = 113 secondary school teachers (Cohort 1: n = 69, Cohort 2: n = 44), with 3,768 students (Cohort 1: n = 2,368, Cohort 2: n = 1,400) in 213 classes with diverse school subjects (e.g., mathematics, German, music). On average, 17.69 students per class (SD = 5.98; Min = 2, Max = 30) with a mean grade level of 8.16 (SD = 2.09) took part in the study. Teachers' mean age was 29 years (SD = 2.11; 79.5% female) with a teaching experience of 2 years on average (SD = 2.02). A total of 32.1% of the teachers were still enrolled in the second part of teacher education.

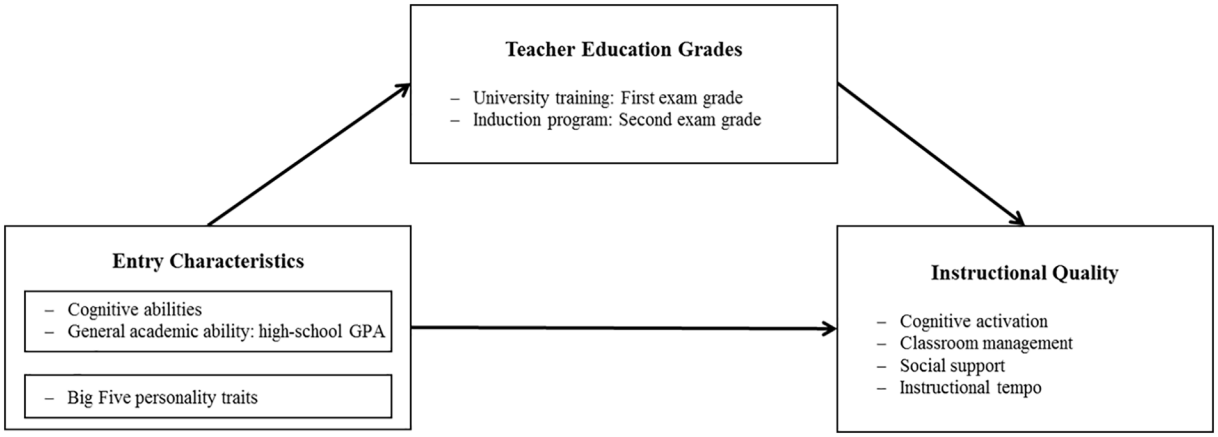

FIGURE 1. *Conceptual overview of the four meditational analyses that were conducted for the four measures of instructional quality.*

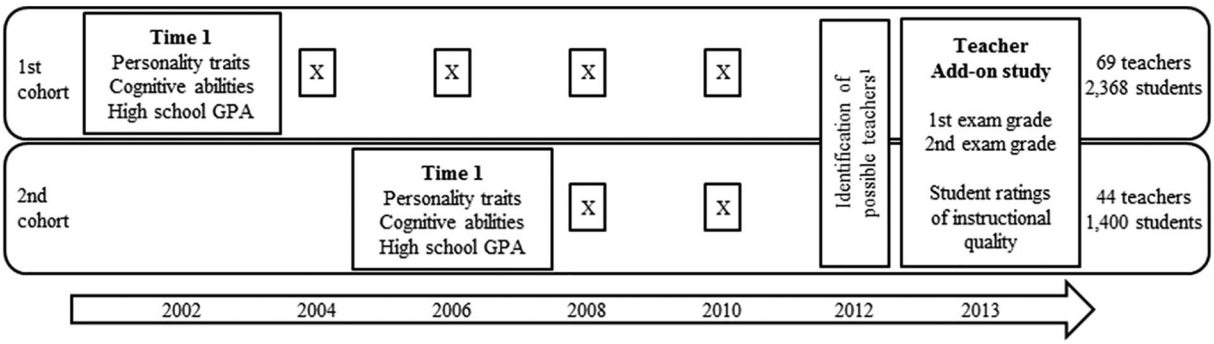

FIGURE 2. *Time frame of the study and final sample sizes.*
Note. X = assessment in the context of TOSCA.
[1]See online Supplemental Figure S1 for further information.

Measures

Entry Characteristics. We used the Kognitiver Fähigkeitstest (KFT 4–12 + R; Heller & Perleth, 2000), a German version of the Cognitive Abilities Test of reasoning skills (Thorndike & Hagen, 1971) to measure prospective teachers' cognitive abilities, implementing the subtests *verbal analogies* (e.g., Creating correlates of relationships between terms: "fire" is to "hot" as "ice" is to . . .) and *figure analogies* (e.g., Creating correlates of relationships between figures: ☐: □ = ○: . . .). To ensure that the mean of the total sample of students who participated in Time 1 was 100 (*SD* = 15) for each cohort, the raw values were transformed (Trautwein, Köller, Lehmann, & Lüdtke, 2007). The mean of cognitive abilities of the teacher sample was 101.17 (*SD* = 12.81).

We used the teachers' high school GPA as a profession-unspecific indicator of teachers' general academic ability prior to entering teacher education. The German high school GPA ranges from 1 = *outstanding* to 6 = *fail*. Information on high school GPA were provided by the schools where the participants finished their final year of secondary schooling. The mean of the high school GPA of the teacher sample in the current study was 2.33 (*SD* = 0.59). To facilitate the interpretation of the results, the high school GPA was reverse coded for all analyses so that higher scores reflected a better GPA.

A German version of the NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) was used to measure the Big Five personality traits: (1) neuroticism, incorporating anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability; (2) extraversion, reflecting warmth, gregariousness, assertiveness, activity, excitement-seeking, and positive emotions; (3) openness to new experiences, incorporating fantasy, aesthetics, feelings, actions, ideas, and values; (4) agreeableness, comprising altruism, sympathy, trust, and nurturance; and (5) conscientiousness, reflecting organization, punctuality, achievement, and honesty. Items were rated on a 4-point Likert-type answering scale (1 = *strongly disagree*; 4 = *strongly agree*). Previous work has shown the reliability, validity, and comparability of the German NEO-FFI (Borkenau & Ostendorf, 1991; Lüdtke, Trautwein, Nagy, & Köller, 2004). The internal consistencies of the scales were acceptable to satisfactory

(neuroticism: $\alpha = .80$; extraversion: $\alpha = .75$; openness to new experiences: $\alpha = .66$; agreeableness: $\alpha = .70$; conscientiousness: $\alpha = .82$).

Grades From Teacher Education. Two grades were used: (1) the final grade from the first state examination (first exam grade; $M = 2.08$, $SD = 0.46$) and (2) the final grade from the second state examination (second exam grade; $M = 1.90$, $SD = 0.46$). Both grades were based on self-reports of the teachers (for the accuracy of self-reported grades, see Sanchez & Buddin, 2016). The grades ranged from 1 = *outstanding* to 6 = *fail* and, to facilitate the interpretation of results, grades were reverse coded for all analyses so that higher scores reflected better grades.

Instructional Quality. Students rated four aspects of their teachers' instructional quality with instruments previously applied in the context of the COACTIV study (Baumert et al., 2009) using a 4-point Likert-type answering scale (1 = *strongly disagree*, 4 = *strongly agree*): cognitive activation (13 items, e.g., "Our teacher lets us use our own strategies to solve difficult problems"; $\alpha = .67$), classroom management (6 items, e.g., "A lot of lesson time is wasted," reverse coded; $\alpha = .80$), perceived social support (19 items, e.g., "Our teacher soon notices if a student has difficulty following the lesson"; $\alpha = .86$), and instructional tempo (6 items, e.g., "We move on so quickly in lessons that a lot of students have difficulty keeping up," reverse coded; $\alpha = .92$). As we had student data on teachers' instructional quality from two classes for most of the teachers (88%), we used three-level multilevel models for all analyses involving instructional quality, with student data being modeled simultaneously at the student, class, and teacher level. Before conducting our analyses, we investigated whether the aggregated student ratings could be generalized across classes and provided reliable indicators of instructional quality (Lüdtke, Robitzsch, Trautwein, & Kunter, 2009). To this end, the total variance of the individual student ratings for each construct was decomposed into the variation between students, classes, and teachers, using an empty three-level model (see Hox, 2010). In the case of a three-level structure (students nested within classes nested within teachers), the intraclass correlation (ICC) indicates the proportion of variance between classes or teachers in relation to the total variance in the individual ratings. The amount of variance for each level as well as ICC values for teacher and class level are displayed in Table 1. The ICC values ranged from 0.12 to 0.19 at the class level and from 0.05 to 0.17 at the teacher level, reflecting substantive variance in the ratings of instructional quality on both levels. The reliability of the aggregated student ratings at the teacher level is a function of these variance components and the number of students within classes and the number of classes (Jeon, Lee, Hwang, & Kang, 2009; see also Hox, 2010).[1] The estimated reliability for the aggregated means of the student ratings at the teacher level was .35 for cognitive activation, .59 for classroom management, .51 for perceived social support, and .44 for instructional tempo. These relatively low reliability values can be explained by the very small number of classes that provided ratings for each teacher (due to our study design, the maximum number of classes per teacher was two). At the class level, the reliability values were .72 for cognitive activation, .84 for classroom management, .82 for perceived social support, and .78 for instructional tempo, indicating that the class-averaged student ratings are sufficiently reliable indicators of instructional quality for classrooms.

Statistical Analyses

Research Question 1 concerned the predictive validity of teachers' personality traits and cognitive and academic abilities for teachers' instructional quality. Our data have a hierarchical structure: teachers' entry characteristics and teacher education grades were assessed at the teacher level and students' ratings of teachers' instructional quality were assessed at the student level. As described above, we used three-level multilevel models for all analyses involving teachers' instructional quality. Although our research question only concerned the teacher level, multilevel models are recommended for this type of data as they account for the hierarchical structure of the data (Hox, 2010; Raudenbush & Bryk, 2002). All multilevel analyses were conducted using the M*plus* 7 software (Muthén & Muthén, 1998–2012) with type = THREE LEVEL to control for the clustering of students within classrooms within teachers. Robust maximum likelihood estimation procedures in M*plus* were used to estimate the models. We investigated models that included teachers' entry characteristics as predictors for cognitive activation (Model I-1), classroom management (Model II-1), social support (Model III-1), and instructional tempo (Model IV-1).

To examine Research Question 2a, we included the first and second exam grades as predictors for cognitive activation (Model I-2a), classroom management (Model II-2a), social support (Model III-2a), and instructional tempo (Model IV-2a). In order to investigate Research Question 2b, we simultaneously included teachers' entry characteristics and first and second exam grades as predictors for cognitive activation (Model I-2b), classroom management (Model II-2b), social support (Model III-2b), and instructional tempo (Model IV-2b). Finally, we wanted to explore whether grades from teacher education act as a mediator for the relationship between entry characteristics and later instructional quality and estimated the indirect effect.

Introducing teacher gender or years of professional experience as covariates into the analyses did not change the interpretation of the findings. All significance testing was performed at the .05 level. In the present investigation, data on the second exam grade was available for $N = 78$ teachers.

TABLE 1
Variance Components and Intraclass Correlations for Instructional Quality

	Variance Components			Proportion of Variance Between (ICC)	
	Students (Level 1)	Classes (Level 2)	Teachers (Level 3)	Classes— V2/(V1 + V2 + V3)	Teachers— V3/(V1 + V2 + V3)
Cognitive activation	0.124	0.018	0.007	0.121	0.047
Classroom management	0.343	0.101	0.092	0.188	0.172
Social support	0.195	0.050	0.034	0.179	0.122
Instructional tempo	0.281	0.056	0.030	0.153	0.082

Note. N = 3,768 students in 213 classes with 113 teachers. ICC = intraclass correlation coefficient; V1 = variance on Level 1; V2 = variance on Level 2; V3 = variance on Level 3. Calculations are based on empty three-level models (see Hox, 2010).

TABLE 2
Descriptive Statistics and Intercorrelations of the Entry Characteristics and Teacher Education Grades on Teacher Level

	M	*SD*	1	2	3	4	5	6	7	8
1. Cognitive abilities	101.17	12.76	—							
2. High school GPA	2.33	0.59	**.31**	—						
3. Neuroticism	2.20	0.45	−.06	.13	—					
4. Extraversion	2.99	0.39	−.04	−.20	**−.45**	—				
5. Openness	2.68	0.41	−.03	.01	.19	.12	—			
6. Agreeableness	3.07	0.36	−.09	−.02	−.19	**.35**	.08	—		
7. Conscientiousness	2.96	0.46	−.09	.15	−.12	−.18	−.01	.16	—	
8. First exam grade	2.08	0.47	.10	**.47**	−.03	.06	−.01	.05	.04	—
9. Second exam grade[a]	1.90	0.46	.10	**.31**	−.24	.12	−.01	.10	.22	**.32**

Note. N = 113 teachers. Correlations statistically significant at $p < .05$ are shown in boldface.
[a]N = 78.

Thus, analyses including the second exam grade are based on these teachers and their 2,643 students. Comparing entry characteristics of teachers with and without information on second exam grade revealed statistically significantly higher values in neuroticism for teachers with information on the second exam grade (d = .76). The average percentage of missing data for Big Five personality traits was 19.5%, mainly due to missing by design: The NEO-FFI was administered to all participants of the 2002 starting cohort, but to maximize the number of instruments that were administered to the participants in the TOSCA cohort starting in 2006, two different versions of the questionnaire were used with only one of them containing NEO-FFI items. For this reason, around 50% of the NEO-FFI data in the 2006 starting cohort were missing by design. Concerning all other teacher-level variables, the average percentage of missing data was 2.1%, ranging from no missing data for teachers' cognitive abilities to 6.2% missing data for teachers' first exam grade. Teachers with missing data on gender, high school GPA, and first exam grade did not differ in entry characteristics from those without missing data. In addition, the average percentage of missing data for student ratings of teachers' instructional quality was below 3%. Based on the missing at random assumption (Enders, 2010), we applied the M*plus* software for multiple imputation, using a three-level imputation model. Following the advice of Graham, Olchowski, and Gilreath (2007), we generated 1,000 data sets in which all missing data were replaced by plausible values. Thereby, all information on teachers' gender, cognitive abilities, high school GPA, personality traits (assessed at Time 1 and in 2012) grades from teacher education, as well as student-rated instructional quality were used in the imputation phase. In addition, information on the Big Five personality traits assessed in the 2006 starting cohort through a five-item measure (Rammstedt, Koch, Borg, & Reitz, 2004) which was administered to all participants of this cohort were used as auxiliary variables in the imputation phase. The resulting data sets were analyzed separately and the results for each data set were combined using Rubin's rules, as implemented in the M*plus* software (see Enders, 2010).

Results

Table 2 provides an overview of the descriptive statistics and intercorrelations for teachers' entry characteristics. Correlation analyses revealed statistically significant

positive correlations between cognitive abilities and high school GPA ($r = .31$, $p < .001$). In addition, extraversion was statistically significantly negatively associated with neuroticism ($r = -.45$, $p < .001$) and positively with agreeableness ($r = .35$, $p = .001$). However, analyses of collinearity among the teachers' entry characteristics revealed no serious collinearity problems (cf. additional information online Supplemental Table S5).

Entry Characteristics as Predictors of Instructional Quality

Our first research question concerned the predictive validity of teachers' entry characteristics for teachers' instructional quality. Four three-level multilevel models were estimated, and cognitive activation (Model I-1 in Table 3), classroom management (Model II-1 in Table 3), social support (Model III-1 in Table 4), and instructional tempo (Model IV-1 in Table 4) were regressed on teachers' entry characteristics entered simultaneously. High school GPA was a valid positive predictor of teachers' classroom management ($\beta = .25$, $p = .023$), indicating that teachers with high general academic ability were able to manage time and disturbances in their classroom more effectively after controlling for cognitive abilities and Big Five personality traits. Thereby, all entry characteristics entered simultaneously as predictors explained 21% of the variance between teachers. In addition, teachers' social support was statistically significantly positively predicted by agreeableness, indicating that high school leavers with high agreeableness provided strong social support ($\beta = .30$, $p = .047$) in their lessons, after controlling for cognitive abilities, high school GPA, and other personality traits. All entry characteristics entered simultaneously explained 17% of between-teacher variance in social support. None of these teacher characteristics statistically significantly predicted cognitive activation or instructional tempo. Cognitive abilities, high school GPA, and personality traits entered simultaneously as predictors explained 19% of the variance between teachers in cognitive activation and 11% of the variance between teachers in their instructional tempo.

Entry Characteristics and Grades From Teacher Education as Predictors for Instructional Quality

We included the first and second exam grade in the four multilevel models (Research Question 2a; Tables 3 and 4). The second exam grade statistically significantly positively predicted efficacy in classroom management ($\beta = .44$, $p = .001$) and instructional tempo ($\beta = .34$, $p = .038$), after controlling for the first exam grade. The first and second exam grade entered simultaneously as predictors explained 6% of the variance between teachers in cognitive activation, 16% in classroom management, 6% in social support, and 10% in instructional tempo.

In the next step, we included the first and second exam grade and teachers' entry characteristics simultaneously in the four multilevel models (Research Question 2b; Tables 3 and 4). The results showed that the second exam grade statistically significantly positively predicted efficacy in classroom management ($\beta = .43$, $p = .005$), after controlling for entry characteristics and first exam grade. In addition, the second exam grade was found to statistically significantly positively predict teachers' instructional tempo ($\beta = .41$, $p = .019$) after controlling for teachers' entry characteristics and first exam grade. Whereas high school GPA was not a statistically significant predictor of classroom management after controlling for grades from teacher education, the results showed that agreeableness remained a statistically significant positive predictor of social support even when teachers' grades from teacher education were included in the regression analyses. This indicates that, if we compare teachers with the same grades from teacher education, those teachers who described themselves as altruistic, sympathetic, trustworthy, and nurturing in their last year of high school were able to provide stronger social support to their students several years later. All predictors entered simultaneously explained 23% of between-teacher variance in cognitive activation, 35% of variance in classroom management, 23% of variance in social support, and 24% in instructional tempo.

Finally, we explored whether grades from teacher education act as a mediator for the associations between teachers' entry characteristics and later instructional quality. We set up four mediation models to investigate whether the second exam grade mediated the effect between entry characteristics and teachers' instructional quality. We found a statistically significant positive indirect effect ($\beta = .16$, $p = .029$) of high school GPA mediated through the second exam grade on classroom management. The nonsignificant relationship between high school GPA and classroom management in Model IV-2b indicated a full mediation. Calculating the ratio of the indirect effect to the total effect, 63% of the total effect is explained by the indirect effect (Wen & Fan, 2015). Furthermore, we found a statistically significant positive indirect effect ($\beta = .16$, $p = .028$) of high school GPA mediated through the second exam grade on instructional tempo. The mediation models on cognitive activation and social support did not show significant mediation effects.

Discussion

The main aim of the present study was to investigate the link between teachers' entry characteristics and instructional quality. Our first research question addressed the question of whether teachers' Big Five personality traits, cognitive abilities, and general academic ability (indicated by the high school GPA) can predict secondary school teachers' instructional quality, as rated by their students up to 10 years later.

TABLE 3

Entry Characteristics and Teacher Education Grades Predicting Cognitive Activation and Classroom Management

	Cognitive Activation								Classroom Management							
			Model I-1		Model I-2a[a]		Model I-2b[a]				Model II-1		Model II-2a[a]		Model II-2b[a]	
	r	*p*	β (*SE*)	*p*	β (*SE*)	*p*	β (*SE*)	*p*	*r*	*p*	β (*SE*)	*p*	β (*SE*)	*p*	β (*SE*)	*p*
Entry characteristics																
Cognitive abilities	.00	.986	.07 (.19)	.692			.06 (.18)	.753	.18	.132	.10 (.12)	.435			.08 (.11)	.494
High school GPA	−.20	.161	−.18 (.16)	.258			−.12 (.21)	.552	**.29**	.003	**.25** (.11)	.023			.20 (.14)	.160
Neuroticism	−.19	.246	−.18 (.20)	.368			−.16 (.20)	.419	.00	.995	−.09 (.17)	.585			−.02 (.16)	.919
Extraversion	.20	.218	.03 (.21)	.902			.05 (.22)	.815	−.08	.542	−.01 (.16)	.971			.00 (.17)	.997
Openness	.25	.134	.27 (.17)	.125			.27 (.18)	.131	.20	.145	.24 (.14)	.095			.24 (.14)	.079
Agreeableness	.16	.318	.11 (.17)	.522			.10 (.17)	.561	−.13	.341	−.17 (.14)	.223			−.20 (.13)	.133
Conscientiousness	.00	.984	.01 (.19)	.949			.00 (.20)	.987	.16	.237	.15 (.15)	.298			.10 (.15)	.496
Teacher education grades																
First exam grade	−.21	.236			−.26 (.18)	.149	−.22 (.22)	.301	.05	.735			−.11 (.13)	.396	−.21 (.15)	.148
Second exam grade	.05	.770			.16 (.20)	.420	.15 (.22)	.509	**.43**	<.001			**.44** (.12)	<.001	**.43** (.15)	.005
R^2 (teacher level)			.19		.06		.23				.21		.16		.35	

Note. The table presents intercorrelations (*r*), standardized regression coefficients (β), and standard errors (*SE*) of three-level multilevel models (N = 3,768 students with 113 teachers). Coefficients statistically significant at $p < .05$ are shown in bold print.

[a]N = 2,643 students with 78 teachers.

TABLE 4

Entry Characteristics and Teacher Education Grades Predicting Social Support and Instructional Tempo

	Social Support								Instructional Tempo							
			Model I-1		Model I-2a[a]		Model I-2b[a]				Model II-1		Model II-2a[a]		Model II-2b[a]	
	r	*p*	β (*SE*)	*p*	β (*SE*)	*p*	β (*SE*)	*p*	*r*	*p*	β (*SE*)	*p*	β (*SE*)	*p*	β (*SE*)	*p*
Entry characteristics																
Cognitive abilities	.00	.982	.06 (.15)	.705			.05 (.15)	.766	.08	.571	.10 (.16)	.539			.08 (.16)	.631
High school GPA	−.09	.468	−.09 (.14)	.519			−.16 (.15)	.294	.01	.965	.01 (.16)	.972			−.06 (.19)	.737
Neuroticism	−.10	.503	.00 (.18)	.988			.06 (.18)	.734	−.08	.612	−.06 (.20)	.752			.02 (.20)	.906
Extraversion	.20	.185	.08 (.20)	.698			.08 (.20)	.711	.08	.625	−.06 (.21)	.971			−.06 (.21)	.791
Openness	.09	.582	.06 (.16)	.709			.04 (.15)	.776	.03	.866	.02 (.17)	.912			.00 (.17)	.988
Agreeableness	**.32**	.028	**.30** (.15)	.047			**.28** (.14)	.046	.23	.133	.27 (.16)	.098			.25 (.15)	.098
Conscientiousness	.01	.932	.01 (.16)	.967			−.02 (.16)	.913	−.05	.776	−.10 (.18)	.582			−.14 (.18)	.440
Teacher education grades																
First exam grade	−.03	.834			−.14 (.13)	.280	−.09 (.14)	.513	−.04	.771			−.17 (.15)	.236	−.16 (.15)	.295
Second exam grade	.19	.198			.27 (.16)	.084	.31 (.16)	.056	**.31**	.038			**.34** (.17)	.038	**.41** (.18)	.019
R^2 (teacher level)			.17		.06		.23				.11		.10		.24	

Note. The table presents intercorrelations (*r*), standardized regression coefficients (β), and standard errors (*SE*) of three-level multilevel models ($N = 3{,}768$ students with 113 teachers. Coefficients statistically significant at $p < .05$ are shown in boldface.

[a]$N = 2{,}643$ students with 78 teachers.

Drawing on a longitudinal study, our findings indicate that some entry characteristics of teachers have predictive validity for their later instructional quality. However, the size of these effects can be rated as small to moderate. Furthermore, we found the second exam grade to have predictive validity for teachers' instructional quality.

Which Teacher Characteristics Lead to High-Quality Instruction?

Entry Characteristics and Instructional Quality. In our first research question, we investigated the predictive validity of teachers' entry characteristics for teachers' instructional quality, as rated by their students. We found that teachers who enter teacher education with high values in agreeableness are able to provide a supportive social environment in which students feel secure and personally valued. This is in line with Mount et al. (1998) who already found that agreeableness is associated with job performance in jobs with interpersonal interactions. In addition, teachers' high school GPA was predictive of classroom management. However, cognitive abilities were not predictive of instructional quality. Instead, high school GPA and cognitive abilities were—though not statistically significantly—negatively related to cognitive activation. This finding is intriguing, raising questions about the appropriateness of students' ratings concerning cognitive activation as well as underlying processes explaining these findings. Summarizing, our findings give only limited support for the bright person hypothesis which proposed that the best teachers are bright, well-educated people (Kennedy et al., 2008). These findings are also contrary to previous research across various professions showing that cognitive abilities are valid predictors of occupational success (Schmidt & Hunter, 2004). One possible explanation for this difference might be that those studies used other indicators for occupational success, for example, salary or status of the profession. However, these indicators are less suitable to differentiate between German teachers as teachers' salary is not performance-related but depends to a large extent on school type and years on the job. Another possible explanation for this difference is that there are characteristics of a teachers' workplace that differ largely from those of other occupations: the complexity of teachers' work mainly results from their interactions with students (Doyle, 1986; Rowan, 1994). Teaching therefore seems to require other characteristics, for example, agreeableness, as discussed before. Research on other entry characteristics which are more closely related to teachers' interactions with students, such as the ability to perceive, express, and manage emotions (Salovey, Bedwell, Detweiler, & Mayer, 2000)—might be promising. Another explanation for the finding that conscientiousness was not predictive for instructional quality might be a strong restriction in the variance in our sample of teachers for two reasons. First, our sample might have been positively selected as only those teacher candidates who had finished the first part of teacher education without delay (e.g., failed exams) were included in our sample. However, a comparison of our sample to all students who wanted to become a teacher two years after finishing high school revealed no significant selection effect. Second, people who choose teacher education might differ in their personality characteristics from those who choose other professions. However, using the data of the TOSCA study, Roloff Henoch, Klusmann, Lüdtke, and Trautwein (2015) and Klusmann, Trautwein, Lüdtke, Kunter, and Baumert (2009) found no differences in conscientiousness between teacher candidates and other students.

The focus of this study was on teachers' instructional quality, which seems reasonable as instruction is the "core business of teaching" (Baumert et al., 2013). Nevertheless, the work of a teacher comprises more than instruction—aspects which are not included in student ratings of instructional quality—and we suggest that teachers' entry characteristics, such as conscientiousness, are important predictors of some other aspects of teaching (e.g., school development, engagement in further training, counseling). We expect, for example, a highly conscientious teacher to be more engaged at preparing lessons, teaching additional courses (e.g., if other teachers are sick), and engaging in further training. Future research might benefit from including other measures of teachers' job performance.

Many researchers and politicians have argued that recruitment and selection of individuals who show favorable entry characteristics might raise teachers' instructional quality (Auguste et al., 2010; Kim et al., 2018; Klassen et al., 2014). Although the predictive validity of teachers' entry characteristics on instructional quality was, if at all, moderate in size, the current study showed that teachers with a good high school GPA had an advantage for their later classroom management. This finding is especially important as student discipline problems (which might result from a lack of classroom management) are one of the most important reasons for teacher turnover (Ingersoll, 2001). In addition, teachers who scored high on agreeableness provided higher social support to their students. Considering current research findings, showing that personality traits change through intervention (Roberts et al., 2017), it might reasonably be assumed that relevant personality traits, such as agreeableness, can be supported in a favorable way through adequate programs during teacher education (e.g., training on social skills). We performed some additional analyses on the stability of personality during teacher education as well as predictive validity of personality assessed in the year 2010 for instructional quality. The results show that stability of personality traits is moderate with correlations ranging from .50 to .62 (see online Supplemental Table S2). More important in the present context, associations between personality traits assessed in 2010 and student ratings on their teachers'

instructional quality are quite similar compared with the relationship between personality assessed in Time 1 and instructional quality (see online Supplemental Tables S3 and S4)—37 of the 40 associations have identical significant/nonsignificant associations with instructional quality. Still unknown is whether the associations between teacher characteristics and instructional quality vary throughout their professional career, depending on other teacher characteristics. For example, most beginning teachers are highly motivated when entering the teaching profession and this motivation may suppress, for example, effects of conscientiousness. The initial motivation might decrease after a few years of teaching and effects of conscientiousness on teachers' behavior in the classroom may then become more apparent (e.g., see Trautwein et al., 2015, for compensatory effects of interest and conscientiousness in students). Further research is needed on how the phase of the professional career (e.g., beginning vs. experienced teachers) moderates the relevance of teacher characteristics throughout the professional careers of teachers (see also Richter, Kunter, Klusmann, Lüdtke, & Baumert, 2011, for associations between the phase of teachers' professional career and their uptake of learning opportunities).

Moreover, based on the *offer/take-up model* (Göbel & Helmke, 2010), Kunter, Kleickmann, Klusmann, and Richter (2013) suggested that teachers' entry characteristics might have an indirect effect on teachers' instructional quality through their impact on the use of learning opportunities during teacher education and in-service teacher training. This, in turn, influences the development of professional knowledge and skills, which are assumed to be a key factor in teachers' instructional quality (Baumert et al., 2010; Krauss, Baumert, & Blum, 2008; Krauss, Brunner, et al., 2008; Kunter, et al., 2007). Further research is needed on these indirect effects of teachers' entry characteristics through the development of professional knowledge and skills.

Grades From Teacher Education and Instructional Quality. In recent years, a growing body of research has focused on the predictive validity of characteristics of teachers' education, such as grades from teacher education, having a major in the field to be taught, or teachers' certification status,[2] for teachers' instructional quality and student outcomes (Darling-Hammond, 2000; Hill et al., 2015; Wayne & Youngs, 2003). Our study contributes to this research in that it investigated the relative predictive validity of grades from German teacher education and teachers' entry characteristics for teachers' instructional quality. We found the second exam grade to positively predict classroom management and instructional tempo when teachers' entry characteristics were controlled for, whereas the first exam grade had no predictive validity for teachers' instructional quality. Interestingly—though not statistically significantly—the first exam grade related negatively to student ratings of instructional quality. Due to the division of German teacher education in two separate phases, the first exam grade mainly reflects the amount of declarative knowledge (e.g., knowledge of how to react to disturbances in lessons) acquired during university education, whereas the second exam grade may be an indicator of teacher candidates' procedural knowledge (e.g., the ability to make use of this theoretical knowledge). Our findings point in the direction that procedural knowledge, which builds on declarative knowledge from university education, is more proximal to instructional behavior. Nonetheless, it is important to note that grades are only distal indicators of the acquired knowledge. However, when interpreting our results on the first exam grade it needs to be kept in mind that university grades might vary in their meaning due to different grading practices between different subjects (e.g., German, English, mathematics) and universities. Thus, teachers' first exam grade might not be comparable across subjects and universities which might limit its predictive validity. The second exam grade might be more comparable across subjects, as the focus of this phase is on practical training and pedagogical content. When interpreting the predictive validity of the second exam grade for instructional quality, it should be noted that there is some kind of overlap between these constructs, as the second exam grade is partly based on ratings of teachers' instructional quality. However, there are considerable differences in time of assessment, context, and rater (examination board vs. students).

Despite some overlap between the second exam grade and the instructional quality ratings used in our study, our findings on the predictive validity of the second exam grade are very interesting for policy makers. In Germany, the decision concerning whether teacher candidates actually become teachers mainly depends on their second exam grade and we found this grade to be a valid predictor of teachers' instructional quality. Therefore, our findings confirm the suitability of this selection method. The current study shows that there are differences in the predictive validity of teachers' grades from teacher education between the investigated aspects of instructional quality. It might be assumed that these differences are due to variations in the amount and quality of learning opportunities during teacher education for the aspects of instructional quality, for example, few high-quality learning opportunities for cognitive activation and social support. A stronger integration of learning opportunities on social support into teacher education might also lead to a less pronounced association between applicants' agreeableness and the social support they provide later in the classroom as teachers.

Limitations and Future Research

The present study complements previous findings on teaching quality in several ways. Nonetheless, some main

limitations need to be considered. First, we used student ratings of teachers' instructional quality to identify high-quality teachers, as we argued that student reports are a valid source of information on teachers' instructional quality. Therefore, we avoided shared-method variance, which is present in studies using teachers' self-reports on the quality of their instruction (Biermann, Karbach, Spinath, & Brünken, 2015) or using student ratings' of teaching quality together with student ratings of the teachers' personality (Kim et al., 2018). However, the ratings of students might be influenced by their opinions on what high-quality teachers are, instead of actually on ratings of their teachers' instructional quality (cf. Duckworth et al., 2009). The use of student achievement gains or observer ratings might solve this problem. Nevertheless, we believe teachers' instructional quality as rated by their students to be an important criterion of teacher efficacy, which affects not only student achievement but also students' motivation and emotions (Kunter, Klusmann, et al., 2013).

Second, reviewing previous empirical research on instructional quality, we decided to set the focus of the present investigation on efficient classroom management, cognitively challenging instruction, formation of a supportive social environment, and an appropriate instructional tempo. However, other aspects of instructional quality might also be important to investigate, such as enhancing students' learning difficulties or adaption of instruction (e.g., Corno, 2008).

Third, as we described in the Method section, our sample was selected in terms of entry characteristics due to a self-selection process into teacher education as well as due to external selection processes (e.g., only student teachers who successfully passed the first exam were able to participate in the second part of teacher education) during teacher education. Furthermore, 32.1% of the teachers in our sample were still enrolled in the second part of teacher education and we do not know whether all of them will stay in the teaching profession. Moreover, we found a medium-sized selectivity bias concerning neuroticism between teachers still enrolled in the second part of teacher education and those who already finished teacher education. In addition, due to the data collection procedure, the response rate on teacher level was 37.7% and there is no information on the response rate on student level available. Reasons for attrition may be due to either teachers' and students' unwillingness to respond, teachers' inability to incorporate the data collection into the time frame of their lessons and students' absence from the lessons of data collection. Selectivity analyses concerning teachers' entry characteristics of respondents and nonrespondents were possible and described in the methods section. However, a conceivable high level of selectivity bias concerning instructional quality as well as grades from teacher education of respondents compared with nonrespondents could not be investigated.

Fourth, there was considerable variance on the class level (1.8%–10.1%), indicating that class characteristics may influence the aggregated student ratings on the class level, such as subject taught, gender distribution, or age of the students. For example, teachers' agreeableness might be more important for younger students than for older ones. Further research is needed to investigate the influence of class characteristics in the context of the present research questions and that includes more than two classes per teacher.

Fifth, results of the current study may point to the view that personality traits are not particularly important for teachers' academic and job success. However, it needs to be kept in mind that we investigated rather objective criteria of teachers' academic and job success. Previous research has shown that personality traits are more important for subjective criteria of academic and job success, such as job satisfaction (Judge, Heller, & Mount, 2002; Ng, Eby, Sorensen, & Feldman, 2005) and affective well-being (Alarcon, Eschleman, & Bowling, 2009; Klusmann, Kunter, Voss, & Baumert, 2012).

Sixth, we tested for numerous associations and in order to avoid important effects to go undetected, we decided not to adjust the *p* values for multiple testing, which results in an increased risk of Type I errors (see Keselman, Miller, & Holland, 2011). Moreover, due to the three-level structure of our data (instructional quality ratings of students nested within classes which are nested within teachers), we were not able to apply bootstrapping to the mediation models which might limit the reliability of our findings. We want to caution the reader that more research in diverse contexts is required to investigate whether our results can be replicated.

Conclusions

The question of which teacher characteristics explain differences in the quality of teachers' instruction is central to research on teaching quality and students' educational attainment. The present study contributes to the existing literature on teachers' instructional quality by investigating the role of teachers' entry characteristics and teachers' grades from teacher education based on longitudinal data regarding German secondary school teachers. Our findings indicate that the entry characteristics of teacher candidates might be predictive for their instructional quality which should be further investigated in future research. Summarizing, although there are some important limitations, for example, concerning possible selectivity bias due to the nonresponse rates, this investigation can be seen as an important starting point for further research on teacher characteristics which predict teachers' instructional quality.

Acknowledgments

The research in this paper was funded by the German Federal Ministry of Education and Research (Grant no. 01PK11011B, 01PK11011C, 01PK11011D). Any opinions expressed are those of the authors.

Notes

1. The reliability (ρ^2) of the mean of the student ratings aggregated at the teacher level is given as follows (Jeon et al., 2009):

$$\rho^2 = \frac{V3}{V3 + \frac{V2}{\overline{n}_c} + \frac{V1}{\overline{n}_c \overline{n}_{jc}}}$$

where V1 is the variance between teachers, V2 is the variance between classes within teachers, V3 is the variance between students within classes and teachers, $\overline{n}_c$ is the average number of classes for teachers, and $\overline{n}_{jc}$ is the average number of students within classes. As can be seen, the reliability strongly depends on the average number of classes for teachers

2. Standard certificate means that a teacher has been prepared in a state-approved teacher education program at the undergraduate or graduate level and has a major or minor in the field to be taught, a certain number of credits in education, and a few weeks of student teaching experience.

References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245.

Alarcon, G., Eschleman, K. J., & Bowling, N. A. (2009). Relationships between personality variables and burnout: A meta-analysis. *Work & Stress*, *23*, 244–263. doi:10.1080/02678370903282600

Aloe, A. M., & Becker, B. J. (2009). Teacher verbal ability and school outcomes: Where is the evidence? *Educational Researcher*, *38*, 612–624. doi:10.3102/0013189X09353939

Auguste, B. G., Kihn, P., & Miller, M. (2010). *Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching: An international and market research-based perspective*. New York, NY: McKinsey.

Barr, A. S. (1952). The measurement of teacher characteristics and prediction of teaching efficiency. *Review of Educational Research*, *22*, 169–174.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9–30. doi:10.1111/1468-2389.00160

Bastian, K. C., McCord, D. M., Marks, J., & Carpenter, D. (2017). A temperament for teaching? Associations between personality traits and beginning teacher performance and retention. *AERA Open*, *3*, 1–17. doi:10.1177/2332858416684764

Baumert, J., Blum, W., Brunner, M., Dubberke, T., Jordan, A., Klusmann, U., . . . Tsai, Y.-M. (2009). *Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz (COACTIV): Dokumentation der Erhebungsinstrumente. Materialien aus der Bildungsforschung: Nr. 83* [Professional competence of teachers, cognitively activating instruction, and the development of students' mathematical literacy (COACTIV): Documentation of data collection instruments]. Berlin, Germany: Max-Planck-Institut für Bildungsforschung.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*, 133–180.

Baumert, J., Kunter, M., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). Professional competence of teachers, cognitively activating instruction, and the development of students' mathematical literacy (COACTIV): A research program. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 1–21). New York, NY: Springer.

Berliner, D. C. (1990). What's all the fuss about instructional time? In M. Ben-Peretz, & R. Bromme (Eds.), *The nature of time in schools: Theoretical concepts, practitioner perceptions* (pp. 3–35). New York, NY: Teachers College Press.

Biermann, A., Karbach, J., Spinath, F. M., & Brünken, R. (2015). Investigating effects of the quality of field experiences and personality on perceived teaching skills in German pre-service teachers for secondary schools. *Teaching and Teacher Education*, *51*, 77–87. doi:10.1016/j.tate.2015.06.005

Borkenau, P., & Ostendorf, F. (1991). Ein Fragebogen zur Erfassung fünf robuster Persönlichkeitsfaktoren [A questionnaire for the assessment of five robust personality traits]. *Diagnostica*, *37*, 29–41.

Corno, L. (2008). In teaching adaptively. *Educational Psychologist*, *43*, 161–174. doi:10.1080/00461 52080 21784 66

Cortina, K. S., & Thames, M. H. (2013). Teacher Education in Germany. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 49–62). New York, NY: Springer.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*, 5–13. doi:10.1037/1040-3590.4.1.5

D'Agostino, J. V., & Powers, S. J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. *American Educational Research Journal*, *46*, 146–182. doi:10.3102/0002831208323280

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, *8*, 1. doi:10.14507/epaa.v8n1.2000

De Jong, R., Mainhard, T., van Tartwijk, J., Veldman, I., Verloop, N., & Wubbels, T. (2014). How pre-service teachers' personality traits, self-efficacy, and discipline strategies contribute to the teacher-student relationship. *British Journal of Educational Psychology*, *84*, 294–310. doi:10.1111/bjep.12025

De Jong, R., & Westerhof, K. J. (2001). The quality of student ratings of teacher behavior. *Learning Environments Research*, *4*, 51–85.

Dodge, A. (1943). What are the personality traits of the successful teacher? *Journal of Applied Psychology*, *27*, 325–337. doi:10.1037/h0062404

Downer, J. T., Stuhlman, M., Schweig, J., Martínez, J. F., & Ruzek, E. (2015). Measuring effective teacher-student interactions from a student perspective: A multi-level analysis. *Journal of Early Adolescence*, *35*, 722–758. doi:10.1177/0272431614564059

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 393–431). New York, NY: Macmillan.

Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *Journal of Positive Psychology*, *4*, 540–547. doi:10.1080/17439760903157232

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

Emmerich, W., Rock, D. A., & Trapani, C. S. (2006). Personality in relation to occupational outcomes among established teachers. *Journal of Research in Personality*, *40*, 501–528. doi:10.1016/j.jrp.2005.04.001

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Getzels, J. W., & Jackson, P. W. (1963). The teachers' personality and characteristics. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 506–582). Chicago, IL: Rand McNally.

Göbel, K., & Helmke, A. (2010). Intercultural learning in English as foreign language instruction: The importance of teachers' intercultural experience and the usefulness of precise instructional directives. *Teaching and Teacher Education*, *26*, 1571–1582. doi:10.1016/j.tate.2010.05.008

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213. doi:10.1007/s11121-007-0070-9

Haberman, M. (1993). Predicting the success of urban teachers (the Milwaukee trials). *Action in Teacher Education*, *15*, 1–5.

Haberman, M. (1995). Selecting "star" teachers for children and youth in urban poverty. *Phi Delta Kappan*, *76*, 777–781.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.

Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4.–12. Klassen–Revision (KFT 4-12 + R)* [Cognitive Abilities Test for grades 4 to 12, revision]. Göttingen, Germany: Hogrefe.

Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching: Examining personal and institutional predictors of high-quality instruction. *AERA Open*, *1*(4), 1–23. doi:10.1177/2332858415617703

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*, 794–831. doi:10.3102/0002831210387916

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed., Quantitative Methodology series). New York, NY: Routledge.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*, 499–534. doi:10.3102/00028312038003499

Jeon, M.-J., Lee, G., Hwang, J.-W., & Kang, S.-J. (2009). Estimating reliability of school-level scores using multilevel and generalizability theory models. *Asia Pacific Education Review*, *10*, 149–158. doi:10.1007/s12564-009-9014-3

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the intergrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York, NY: Guilford Press.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, *87*, 530–541.

Kelly, S., & Northrop, L. (2015). Early career outcomes for the "best and brightest": Selectivity, satisfaction, and attrition in the beginning teacher longitudinal survey. *American Educational Research Journal*, *52*, 624–656.

Kennedy, M. M., Ahn, S., & Choi, J. (2008). The value added by teacher education. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education* (pp. 1249–1273). New York, NY: Routledge.

Keselman, H. J., Miller, C. W., & Holland, B. (2011). Many tests of significance: New methods for controlling type I errors. *Psychological Methods*, *16*, 420–431. doi:10.1037/a0025810

Kim, L. E., Dar-Nimrod, I., & MacCann, C. (2018). Teacher personality and teacher effectiveness in secondary school: Personality predicts teacher support and student self-efficacy but not academic achievement. *Journal of Educational Psychology*, *110*, 309–323. doi:10.1037/edu0000217

Kim, L. E., Jörg, V., & Klassen, R. M. (2019). A meta-analysis of the effects of teacher personality on teacher effectiveness and burnout. *Educational Psychology Review*, *31*, 163–195. doi:10.1007/s10648-018-9458-2

Klassen, R. M., Durksen, T. L., Kim, L. E., Patterson, F., Rowett, E., Warwick, J., . . . Wolper, M.-A. (2017). Developing a proof-of-concept selection test for entry into primary teacher education programs. *International Journal of Assessment Tools in Education*, *4*, 96–114.

Klassen, R. M., Durksen, T. L., Rowett, E., & Patterson, F. (2014). Applicant reactions to a situational judgement test used for selection into initial teacher training. *International Journal of Educational Psychology*, *3*, 104–124. doi:10.4471/ijep.2014.07

Klassen, R. M., & Kim, L. E. (2018). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review*, *26*, 32–51. doi:10.1016/j.edurev.2018.12.003

Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, *12*, 59–76. doi:10.1016/j.edurev.2014.06.001

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). Münster, Germany: Waxmann.

Klusmann, U., Kunter, M., Voss, T., & Baumert, J. (2012). Berufliche Beanspruchung angehender Lehrkräfte: Die Effekte von Persönlichkeit, pädagogischer Vorerfahrung und professioneller Kompetenz [Emotional exhaustion and job satisfaction of beginning teachers: The role of personality, educational experience and professional competence]. *Zeitschrift für Pädagogische Psychologie*, *26*, 275–290. doi:10.1024/1010-0652/a000078

Klusmann, U., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2009). Eingangsvoraussetzungen beim Studienbeginn: Werden die Lehramtskandidaten unterschätzt? [Cognitive and psychosocial characteristics upon entry into teacher training: Are teacher candidates underestimated?] *Zeitschrift für Pädagogische Psychologie*, *23*, 265–278. doi:10.1024/1010-0652.23.34.265

Köller, O., Watermann, R., Trautwein, U., & Lüdtke, O. (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA: Eine Untersuchung an allgemein bildenden und beruflichen*

Gymnasien. [Educational pathways to college in Baden-Württemberg. TOSCA: A study of traditional and vocational Gymnasium schools]. Opladen, Germany: Leske + Budrich.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rinehart and Winston.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *International Journal on Mathematics Education*, *40*, 873–892.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*, 716–725.

Kuncel, N. R., Hezlett, S. A., & Ones, D. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.

Kunter, M., Kleickmann, T., Klusmann, U., & Richter, D. (2013). The development of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers* (pp. 63–77). New York, NY: Springer.

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

Kunter, M., Klusmann, U., Dubberke, T., Baumert, J., Blum, W., Brunner, M., . . . Tsai, Y.-M. (2007). Linking aspects of teacher competence to their instruction: Results from the COACTIV project. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG Priority Programme* (pp. 32–52). Münster, Germany: Waxmann.

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*, 120–131. doi:10.1016/j.cedpsych.2008.12.001

Lüdtke, O., Trautwein, U., Nagy, G., & Köller, O. (2004). Eine Validierungsstudie zum NEO-FFI in einer Stichprobe junger Erwachsener: Effekte des Itemformats, faktorielle Validität und Zusammenhänge mit Schulleistungsindikatoren [A validation of the NEO-FFI in a sample of young adults: Effects of the response format, factorial validity, and relations with indicators of academic achievement]. *Diagnostica*, *50*, 134–144.

Mainhard, T., Oudman, S., Hornstra, L., Bosker, R. J., & Goetz, T. (2018). Student emotions in class: The relative importance of teachers and their interpersonal relations with students. *Learning and Instruction*, *53*, 109–119.

McCrae, R. R., & Costa, P. T. (2008). The Five-Factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). New York, NY: Guilford Press.

Meeus, W., Van de Schoot, R., Klimstra, T., & Branje, S. (2011). Personality types in adolescence: Change and stability and links with adjustment and relationships: A five-wave longitudinal study. *Developmental Psychology*, *47*, 1181–1195. doi:10.1037/a0023816

Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-Factor Model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, *11*, 145–165. doi:10.1080/08959285.1998.9668029

Muthén, L. K., & Muthén, B. (1998–2012). *MPlus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, *58*, 367408. doi:10.1111/j.1744-6570.2005.00515.x

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

Poropat, A. E. (2009). A meta-analysis of the Five-Factor Model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338.

Praetorius, A.-K., Lenske, G., & Helmke, A. (2012). Observer ratings of instructional quality: Do they fulfill what they promise? *Learning and Instruction*, *22*, 387–400. doi:10.1016/j.learninstruc.2012.03.002

Rammstedt, B., Koch, K., Borg, I., & Reitz, T. (2004). Entwicklung und Validierung einer Kurzskala für die Messung der Big Five Persönlichkeitsdimensionen in Umfragen [Development and validation of a short scale for measuring the Big Five personality dimensions in surveys]. *ZUMA Nachrichten*, *55*, 5–28.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed., Advanced Quantitative Techniques in the Social Sciences series 1). Thousand Oaks, CA: Sage.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*, 353–387. doi:10.1037/a0026838

Richter, D., Kunter, M., Klusmann, U., Lüdtke, O., & Baumert, J. (2011). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education*, *27*, 116–126.

Rimm-Kaufman, S. E., & Hamre, B. K. (2010). The role of psychological and developmental science in efforts to improve teacher quality. *Teachers College Record*, *112*, 2988–3023.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417–458.

Roberts, B. W., & Del Vecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*, 117–141.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*, 43–74. doi:10.1162/EDFP_a_00022

Roloff Henoch, J., Klusmann, U., Lüdtke, O., & Trautwein, U. (2015). Who becomes a teacher? Challenging the "Negative Selection" hypothesis. *Learning and Instruction*, *36*, 46–56. doi:10.1016/j.learninstruc.2014.11.005

Rowan, B. (1994). Comparing teachers' work with work in other occupations: Notes on the professional status of teaching. *Educational Researcher*, *23*, 4–17. doi:10.3102/0013189X023006004

Ryan, R. M., & Powelson, C. L. (2014). Autonomy and relatedness as fundamental to motivation and education. *Journal of Experimental Education*, *60*, 49–66. doi:10.1080/00220973.1991.10806579

Salovey, P., Bedwell, T., Detweiler, J. B., & Mayer, J. D. (2000). Current directions in emotional intelligence research. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 504–520). New York, NY: Guilford Press.

Sanchez, E., & Buddin, R. (2016). *How accurate are self-reported high school courses, course grades, and grade point average?* (ACT Research Report 2016-3). Iowa City, IA: ACT.

Schleicher, A. (2016). *Teaching excellence through professional learning and policy reform*. Paris, France: OECD.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*, 162–173.

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*, 454–499. doi:10.3102/0034654307310317

Spengler, M., Brunner, M., Damian, R. I., Lüdtke, O., Martin, R., & Roberts, B. W. (2015). Student characteristics and behaviors at age 12 predict occupational success 40 years later over and above childhood IQ and parental socioeconomic status. *Developmental Psychology*, *51*, 1329–1340. doi:10.1037/dev0000025

Start, K. B. (1966). Personality measures and their relation to teaching ability. *British Journal of Educational Psychology*, *36*, 158–165.

Statistisches Bundesamt. (2004). *Datenreport 2004. Zahlen und Fakten über die Bundesrepublik Deutschland* [Data report 2004. Numbers and facts about the Federal Republic of Germany] (2nd ed.). Bonn, Germany: Bundeszentrale für politische Bildung.

Thorndike, R. L., & Hagen, E. (1971). *Cognitive Abilities Test (CAT)*. Boston, MA: Houghton-Mifflin.

Tok, S., & Morali, S. (2009). Trait emotional intelligence, the Big Five personality dimensions and academic success in physical education teacher candidates. *Social Behavior and Personality*, *37*, 921–931. doi:10.2224/sbp.2009.37.7.921

Trapmann, S., Hell, B., Hirn, J.-O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Journal of Psychology*, *215*, 132–151.

Trautwein, U., Köller, O., Lehmann, R., & Lüdtke, O. (2007). *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten* [*School achievement in Gymnasium schools: Regional, school-based and social disparities*]. Münster, Germany: Waxmann.

Trautwein, U., Lüdtke, O., Nagy, N., Lenski, A., Niggli, A., & Schnyder, I. (2015). Using individual interest and conscientiousness to predict academic effort: Additive, synergistic, or compensatory effects? *Journal of Personality and Social Psychology*, *109*, 142–162.

Trautwein, U., Neumann, M., Nagy, G., Lüdtke, O., & Maaz, K. (2010). *Schulleistungen von Abiturienten: Die neugeordnete gymnasiale Oberstufe auf dem Prüfstand* [*School achievement of Abitur students: Evaluating the reform of upper secondary education*]. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Wang, M.-T., & Eccles, J. S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development*, *83*, 877–895. doi:10.1111/j.1467-8624.2012.01745.x

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

Wen, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure. *Psychological Methods*, *20*, 193–203. doi:10.1037/met0000029

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*, *20*, 23–45. doi:10.1080/10627197.2015.997614

Wubbels, T., Brekelmans, M., & Hooymayers, H. P. (1992). Do teacher ideals distort the self-reports of their interpersonal behavior? *Teaching and Teacher Education*, *8*, 47–58. doi:10.1016/0742-051X(92)90039-6

Yeh, S. S. (2009). The cost-effectiveness of raising teacher quality. *Educational Research Review*, *4*, 220–232. doi:10.1016/j.edurev.2008.06.002

Zumwalt, K., & Craig, E. (2005). Teachers' characteristics: Research on the demographic profile. In M. Cochran-Smith, & K. M. Zeichner (Eds.), *Studying teacher education* (pp. 111–155). Mahwah, NJ: Lawrence Erlbaum.

Authors

JANINA ROLOFF is affiliated to the IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany. Her research interests are teachers' professional competence, teachers' instructional quality, and the measurement of change.

UTA KLUSMANN is affiliated to the IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany. Her research interests are teachers' professional competence, occupational well-being in schools, and career choice.

OLIVER LÜDTKE is affiliated to the IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany, as well as to the Centre for International Student Assessment, Germany. His research interests are multilevel analysis in educational and psychological research, national and international large-scale educational assessment, personality development in adolescence and young adulthood, application of Bayesian methods (e.g., missing data, multilevel modeling), and estimation of causal effects with nonexperimental data.

ULRICH TRAUTWEIN is affiliated to the Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Germany. His research interests are the development of self-referent cognitions in the school context, school management, and the influence of homework on school achievement.