Educational Research Review 30 (2020) 100312

Contents lists available at ScienceDirect

Educational Research Review

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

Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness

Lisa Bardach*, Robert M. Klassen

University of York, Department of Education, York, YO10 5DD, United Kingdom

ARTICLE INFO

Keywords:
Teacher
Cognitive ability
Intelligence
Teacher effectiveness
Student achievement
Teacher evaluation
Systematic review

ABSTRACT

This study provides a systematic review of the literature on teachers' cognitive abilities (intelligence test scores and proxies of cognitive abilities such as college entrance exam scores and basic skills test scores) and teacher effectiveness. Twenty-seven studies conducted between 2000 and 2019 constitute the sample for this review. Studies using intelligence test scores were rare, with the results indicating no or negative associations with teacher effectiveness. Studies on proxies of cognitive abilities yielded, at most, small positive relations with teacher effectiveness. However, behind these overall results regarding proxies of cognitive abilities lie interesting heterogeneities, as several studies analyzing different test domains uncover a differentiated pattern of findings. We also identify key limitations related to construct measurement, sampling approaches, statistical analyses and the interpretation and reporting of the included studies, and outline a path for future research on teachers' cognitive abilities and teacher effectiveness.

Teachers have a profound effect on student learning and achievement (e.g., Hattie, 2009), and some teachers are clearly more effective than others in promoting desirable educational outcomes (e.g., Atteberry, Loeb, & Wyckoff, 2015). With the ultimate goal of improving education, identifying the attributes contributing to teacher effectiveness has been and continues to be critical. For this reason, the empirical examination of teacher characteristics potentially linked to teacher effectiveness has spurred considerable interest over the past decades (e.g., Darling-Hammond, 2000; Klassen & Tze, 2014).

However, even though teaching is complex and intellectually demanding (e.g., Rotherham & Mead, 2003; Rowan, 1994; Shulman, 2004), differences in teachers' cognitive abilities—commonly referred to as intelligence—have received little attention in (recent) teacher effectiveness research (e.g., Harris & Rutledge, 2010). In contrast, mounting research highlights the robust relations between intelligence and job performance in a variety of other professions (e.g., manager, clerk, salesperson in Hunter & Hunter, 1984; seven main groups [clerical, engineer, professional, driver, operator, manager, and sales] in Bertua, Anderson, & Salgado, 2005; for a summary of meta-analyses relying on a range of different professions, e.g., clerical jobs, computer programmers, engineers, drivers, petroleum workers, managers, typists, pilots, police officers, see Ones, Dilchert, Viswesvaran, & Salgado, 2012). The fact that intelligence has repeatedly been portrayed as the major determinant of job performance (e.g., Schmidt, 2015) indicates a need to systematically examine the relevance of intelligence for performance in the teaching domain. Importantly, such an endeavor is not about proving (or refuting) the "bright person hypothesis" (Kennedy, Ahn, & Choi, 2008; see also e.g., Kunter et al., 2013) or similar assumptions, which state that the best teachers are smart, and thus teacher selection should heavily rely on cognitive abilities. Without doubt, high-quality teaching requires more than just high cognitive abilities (see e.g., Muijs et al., 2014; Hamre et al., 2013; Kunter et al., 2013; Kleickman et al., 2013). Nonetheless, theorizing on the role of intelligence for teacher effectiveness without a

* Corresponding author.
E-mail address: lisa.bardach@york.ac.uk (L. Bardach).

https://doi.org/10.1016/j.edurev.2020.100312
Received 11 June 2019; Received in revised form 21 November 2019; Accepted 20 January 2020
Available online 28 January 2020
1747-938X/

thorough understanding of the relationship between these two factors arguably hinders progress in theory-building. Hence, it is necessary to summarize and synthesize the available evidence on the association between teachers' intelligence and teacher effectiveness in order to make claims based on evidence, and inform research on teacher effectiveness as well as educational practice.

Compared to research on teachers' intelligence, substantially more studies exist on proxies of cognitive abilities, i.e., college entrance exam scores or basic skills test scores as measures which are typically highly correlated with intelligence (e.g., Frey & Detterman, 2004; Koenig, Frey, & Detterman, 2008; for meta-analyses on the link between proxies of cognitive abilities and teacher effectiveness see Aloe & Becker, 2009; D'Agostino & Powers, 2009). Still, numerous questions regarding proxies of teachers' cognitive abilities have also remained unaddressed so far. For instance, we lack knowledge about the relative importance of different domains (e.g., verbal vs. numerical). Moreover, the quality of studies on proxies of cognitive abilities and the extent to which study flaws limit the trustworthiness of and conclusions drawn from their findings have not yet been subject to systematic research syntheses. Nevertheless, as proxies of cognitive abilities constitute integral components of teacher selection procedures in numerous countries (e.g., Klassen & Kim, 2019), complementing and expanding the current knowledge on how proxies of cognitive abilities are actually related to teacher effectiveness is of pivotal importance.

We therefore conducted a systematic review on empirical studies exploring the relation between teachers' cognitive abilities—both in terms of teachers' intelligence and proxies of cognitive abilities—and teacher effectiveness. Our review focused on the following questions: (a) What measures have been used to map cognitive abilities and what domains of cognitive abilities have been investigated? (b) What are the studies' main findings concerning the relation between cognitive abilities and teacher effectiveness? (c) To what extent do study limitations pose a threat to the substantive conclusions derived from these findings?

1. Teacher effectiveness

In the context of teaching, effectiveness typically refers to the types of action that produce or facilitate learning (Ferguson & Danielson, 2015, see also e.g., Seidel & Shavelson, 2007; Klassen & Tze, 2014). Stated differently, teaching is effective when it enables student learning (Bill & Melinda Gates Foundation, 2013). In this article we define teacher effectiveness as the effects of high-quality teaching on student learning in terms of achievement gains. We regard high quality teaching as the dynamic and interactive process of creating, fostering, adapting, and negotiating learning environments in which all students are supported in activities that have a good chance of improving learning (e.g., Seidel & Shavelson, 2007). Breaking this definition down into tangible and measurable components, teacher effectiveness can thus both be assessed focusing on the outcome itself (achievement gains) as well as teachers' behaviors (teachers' behaviors in terms of high-quality teaching) that should, broadly speaking, contribute to the outcome (e.g., Darling-Hammond, 2010). Accordingly, teacher effectiveness is commonly operationalized and evaluated in terms of gains in students' achievement test scores (value-added, VA, e.g., Ballou, Sanders, & Wright, 2004) or ratings of teachers' performance in the classroom by principals, supervising teachers, or other external parties, with the latter typically involving classroom observations of teachers or pre-service teachers using standardized observation tools (such as the Classroom Assessment Scoring System, CLASS, Hamre et al., 2013).

Certainly, all measures used to assess teacher effectiveness have their limitations. As such, critics urge caution in the use of VA measures as they might be biased by the non-random sorting of better performing (lower performing) students to more effective (less effective) teachers and schools (e.g., Rothstein, 2010)—although a recent systematic review suggests that sorting does not appear to lead to significant bias in effect estimates, especially when researchers include several years of data in the analyses (Everson, 2017). In a study which utilized random-sorting of students to different teachers, it was furthermore shown that teachers who had previously been identified as more effective did also produce higher average student achievement growth following random assignment (Kane, McCaffrey, Miller, & Staiger, 2013). On the other hand, one of the critiques levelled against observations is their failure to differentiate between teachers. As criterion-referenced measures, they do not necessarily lead to a distribution of ratings, and, historically, most teachers have been rated as effective or highly effective (Cohen & Goldhaber, 2016; see also Muijs, 2006). Nonetheless, others have claimed that observations can be valuable measures of effectiveness, given that certain requirements are met (e.g., two or more observations by at least two trained and certified observers, Kane & Staiger, 2012). In our systematic review, we acknowledge the limitations of different measures of teacher effectiveness. Still, we follow other research syntheses in this area (e.g., Aloe & Becker, 2009; D'Agostino & Powers, 2009) and consider all of the assessments outlined above as indicators of the difficult-to-define and even more difficult-to-measure construct of teacher effectiveness (Kane et al., 2013; Klassen & Kim, 2019). Specifically focusing on teachers' cognitive abilities, our review addresses the question "do higher cognitive abilities make a teacher more effective?".

2. Teachers' cognitive abilities and their relation to teacher effectiveness

Intelligence or cognitive ability can be defined as "a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (Gottfredson, 1997, p. 13; for a detailed description of the historical development of the intelligence construct and different theories of cognitive abilities see e.g., Gustafsson, 1984; Sternberg & Grigorenko, 2002). Although there is still disagreement on the exact structure of cognitive abilities (e.g., Nisbett et al., 2012; Sternberg & Grigorenko, 2002), a prominent perspective sees them as hierarchically organized. Following Spearman (1904; see also Carroll, 1993), intelligence can thus be understood in terms of a general cognitive ability (general intelligence factor, "*g*") that pervades all intellectual tasks, and specific abilities unique to each particular intellectual task (e.g., Sternberg, 2012). Thereby, the general intelligence factor ("*g*") occupies the vertex of the hierarchy of cognitive abilities and the specific abilities are grouped underneath. The specific cognitive abilities carry large components of *g* and refer to domains such as quantitative or spatial reasoning (Lubinski, 2004). Another way to characterize the hierarchical nature of intelligence is to differentiate between fluid and crystallized abilities (e.g., Cattell, 1971). Whereas fluid abilities involve the ability to solve novel or abstract problems using general reasoning methods, crystallized abilities refer to one's general store of knowledge relevant to adaptation in life, including skills such as vocabulary (Cattell, 1971; Ones et al., 2012; Sternberg, 2012). However, several researchers have urged that fluid intelligence can simply be equated with *g* (e.g., Gustafsson, 2002; see also Gustafsson & Snow, 1997).

To date, there is overwhelming research evidence showing a strong link between general intelligence (*g*) as well as specific cognitive abilities and job performance in a variety of profession (e.g., clerical employees, sales, manager), with the highest relations between cognitive abilities and performance found for occupations involving greater complexity (e.g., Bertua et al., 2005; Hunter & Hunter, 1984; Ones et al., 2012; Schmidt, 2015). Teaching is highly complex, challenging, and demanding (e.g., Rotherham & Mead, 2003; Shulman, 2004). For instance, comparing 591 occupations, Rowan (1994) reported that about 75% of all occupations were rated lower in complexity than elementary and secondary school teaching. Hence, it seems reasonable to assume that cognitive abilities are likely to be predictive of teacher effectiveness. Moreover, recall the definition of intelligence as ability to, e.g., "reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (Gottfredson, 1997, p. 13) presented above. Most would agree with the assertion that effective teaching requires some—if not all—of the mentioned characteristics: For example, effective teachers have to react to and quickly adapt their teaching to challenging (classroom) situations to ensure that all students have the chance to learn. In addition, effective teachers need to engage in (long-term) planning, e.g., regarding the content and structure of their lessons, in order to sustainably promote student learning and achievement. Furthermore, should highly effective teachers not be most likely those who are able to use the past to inform their current teaching behavior and thus, to learn from prior experiences and continually grow as teachers? In sum, prior research involving other professions and theoretical considerations linking components of intelligence to teachers' job demands provide a rationale for studying the relation between teachers' intelligence and teacher effectiveness.

The results of early work investigating relations between teachers' cognitive abilities measured via intelligence tests and teacher effectiveness, however, might best be described as inconclusive. Summarizing research in the field, Rostker (1945) pointed towards studies reporting positive as well as (small) negative correlations coefficients. In a later review, Getzels and Jackson (1963; as cited in Harris & Rutledge, 2010) even proposed that inconsistencies in the results of studies linking teachers' cognitive abilities to teacher effectiveness evaluations (here in terms of principal evaluations) led researchers to abandon the study of this topic. Nevertheless, it is difficult to reach confident conclusions due to the limitations of this early body of research especially in terms of methodological concerns such as unreliable measures of student achievement gains and small samples (Rostker, 1945; see also e.g., Harris & Rutledge, 2010). As such, the divergence between inconsistent prior findings, on the one hand, and meta-analytic results on the importance of intelligence for other professions with similar or even lower complexity than teaching (e.g., sales, engineers, clerks, e.g., Hunter & Hunter, 1984; Bertua et al., 2005; Ones et al., 2012) as well as theoretical assumptions on the relation between teacher intelligence and effective teaching, on the other hand, redouble the imperative to learn more about how and whether teachers' intelligence matters for teacher effectiveness.

More recently, researchers have mainly focused on proxies of teachers' cognitive abilities, such as teachers' basic academic skills (e.g., reading, writing, and mathematics as part of the Praxis 1 test for candidates entering teacher education programs) or college entrance exam scores (e.g., Scholastic Aptitude Test [SAT], American College Test [ACT])—in other words, measures that exhibit high associations with intelligence (e.g., Frey & Detterman, 2004; Koenig et al., 2008). Accordingly, numerous studies have investigated the impact of each of these kinds of proxies of teachers' cognitive abilities, and results from this research base have been meta-analytically examined. In one such meta-analysis that narrowed its scope to verbal abilities, Aloe and Becker (2009) found that the effects of scores on verbal college admission tests on teacher effectiveness were not significantly different from zero. A further meta-analysis considering teachers' scores on a basic academic skills test revealed a small positive correlation between basic skills test scores and teacher effectiveness (D'Agostino & Powers, 2009).

Previous work on proxies of teachers' cognitive ability, particularly the meta-analyses by Aloe and Becker (2009) and D'Agostino and Powers (2009), have undoubtedly advanced our knowledge in this area. Meta-analyses quantitatively summarize study effects and are thus well-suited to provide answers to questions like "what is the evidence?" —or, in the case of Aloe and Becker (2009), "where is the evidence?" However, these approaches do not allow for, and are not designed to provide more in-depth insights, such as "how was this evidence derived?", or "what was the context of the study, and to what extent can we trust these findings?" In addition, given that D'Agostino and Powers (2009) analyzed composite scores of teacher basic skills tests and Aloe and Becker (2009) strictly focused on verbal abilities, we know that scores on these specific tests appear to be rather weak predictors of teacher effectiveness, but we do not know whether this holds equally for all domains (e.g., numerical abilities), calling for a research synthesis addressing these issues. Furthermore, gaining a more complete understanding of how cognitive abilities, which are nowadays in the context of teaching mainly assessed by using proxy measures, contribute to teacher effectiveness is crucial, particularly in light of their use in decision-making about who becomes a teacher (e.g., Klassen & Kim, 2019).

3. Aims and research questions

Against the background of the aforementioned research and the increasing awareness that teachers' individual attributes affect their effectiveness and students' learning (e.g., Hill, Charalambous, & Chin, 2018; Klassen & Kim, 2017; Wayne & Youngs, 2003), the present work aims to revisit the relations between teachers' cognitive abilities and teacher effectiveness. Therefore, a systematic review of the relationship between cognitive abilities–both in a strict sense (i.e., intelligence) and proxies of cognitive abilities (i.e., college entrance scores, basic academic skills assessments)–and teacher effectiveness was conducted.[1] Our review is guided by the following research questions: (a) What measures have been used to map cognitive abilities and what domains of cognitive abilities have been investigated? (b) What are the main findings regarding the links between cognitive abilities and teacher effectiveness? and (c) To what extent do study limitations pose a threat to the substantive conclusion derived from these findings?

To minimize the overlap with prior research syntheses and expand and update existing work, we rely on studies from 2000 to 2019 (the meta-analysis by D'Agostino & Powers, 2009, considered studies from 1905 to 2005; Aloe & Becker, 2009, synthesized findings from studies dating from 1960 until 2005; for further reviews on research mainly carried out before 2000, see also Darling-Hammond, 2000; Greenwald, Hedges, & Laine, 1996; Hanushek, 1997; Wayne & Youngs, 2003). Moreover, as existing syntheses have, to the best of our knowledge, exclusively focused on the U.S. context, we furthermore aim to shed light on the relation between cognitive abilities and teacher effectiveness from an international perspective by including studies from a wider range of countries.

4. Method

4.1. Literature search - Phase 1 (first search in databases)

We searched for combinations of the terms *teacher* and *cognitive ability, cognitive skill, cognitive characteristic, intelligence, academic attribute, academic characteristic* and *teacher effectiveness, teaching effectiveness, teaching performance, teacher performance, student achievement, academic performance, academic achievement* in the four databases PsycINFO, Web of Science, ERIC - Education Resources Information Center, and ProQuest Dissertations in January and February 2019. As we focused on studies published between 2000 and 2019 in our review, we restricted the search to exclude work that had been conducted before the turn of the millennium. This search yielded a total of 1688 titles, of which 1562 remained after deduplication.

Inclusion and exclusion criteria. We screened the titles and abstracts of all 1562 articles using a comprehensive set of eight inclusion and exclusion criteria (see Table 1) in order to select studies for the systematic review.

Using these criteria, 1552 of the titles were excluded, leaving us with 10 titles. We screened the full texts of these 10 articles for eligibility, applying the same inclusion and exclusion criteria. Of the 10 articles, five met all criteria for study inclusion and four were deemed ineligible. The remaining study had to be excluded as there was no full text available online and no response to our correspondence with the author asking him to provide the article.

4.2. Literature search – Phase 2 (expanded search)

Due to the small number of hits from the first search, we used the following strategies to increase the number of eligible articles. We conducted a second search in the same databases using different search terms to find studies on proxies of cognitive abilities that the first search might have failed to identify. We thus searched for the terms *college entrance exam test, college admission test, teacher basic skills test, teacher certification test, teacher licensure test* in combination with terms describing teacher effectiveness employed in the first search (e.g., *teacher effectiveness, teacher performance*). This search yielded 426 hits and 403 studies after duplicates were removed. We applied the same inclusion and exclusion criteria as in the first search and identified three additional studies that could be included.

Furthermore, we checked the reference lists of meta-analyses and reviews on closely related topics (e.g., Aloe & Becker, 2009; D'Agostino & Powers, 2009; Harris & Rutledge, 2010; Wayne & Youngs, 2003) that had been conducted after 2000 to identify studies published after 2000. This led to three not yet identified studies. We also searched Google Scholar using the same terms as in the database searches and found one further study meeting our inclusion criteria. In addition, given that measures of cognitive abilities are commonly employed as part of teacher selection procedures, we reviewed the references of a recently published meta-analysis on teacher selection practices and teacher effectiveness (Klassen & Kim, 2019). Three studies cited in this meta-analysis were eligible for inclusion in our review. We also performed citation tracking for all included studies from the data base searches and the expanded search not relying on data bases (e.g., studies found on Google Scholar, studies cited in meta-analyses), which allowed us to identify 11 additional studies. Importantly, one of these studies (P. C. Hall & West, 2011) relied on the same data set as a study identified in the first database search (P. C. Hall, 2009). Thus, we deleted the latter and considered only the former to avoid data dependencies. To identify further relevant studies, we screened the references of all studies that were included. This yielded two additional studies. The additional search yielded 22 new studies, a large number compared to the number retrieved from the first database search. This is mainly due to the fact that most of the 22 studies did not include core search terms from the first search. For instance, studies on teacher selection usually do not refer to ‘cognitive abilities’ or similar terms. All in all, 27 studies could be included in our review. Fig. 1 shows a flowchart of the search process according to PRISMA guidelines.

4.3. Information retrieval process

We first coded descriptive information on all studies (authors, date, year published, journal, location of the study, sample size, measure of teacher effectiveness, and statistical analyses). Then, in order to answer our research questions, the following information was extracted from all studies with respect to each research question(s):

[1] Throughout the manuscript, we use the term “cognitive abilities” as an overarching category, but use the terms “intelligence” and “proxies of cognitive abilities” when referring specifically to these two areas.

Table 1
Inclusion and exclusion criteria.

Criterion	Included	Excluded
1. Measure of teachers' cognitive abilities	Studies measuring teachers' cognitive abilities using a) published intelligence tests, b) college admission tests, such as the SAT, ACT, or GRE, c) basic skills tests that assess basic academic skills in domains such as mathematics, reading, and writing, such as Praxis 1 (or US state-specific tests that can be passed instead of the Praxis 1, for instance the CBEST or WEST-B)	Studies that a) assessed cognitive abilities using instruments that researchers had developed ad hoc for the purpose of their study to ensure a certain degree of comparability across studies, and that the instruments met psychometric standards (e.g., objective and standardized measure, rigorous development), b) conceptualized GPA as cognitive ability, given that GPA, as a very general measure of ability, captures not only cognitive abilities but also personality traits (e.g., Borghans, Golsteyn, Heckman, & Humphries, 2016; Grönqvist & Vlachos, 2016), c) assessed intelligence in terms of teachers' emotional, cultural, social or practical intelligence, d) assessed cognitive abilities as teachers' pedagogical knowledge or subject-specific content knowledge, e.g., the Praxis 2 and Praxis 3 (e.g., Clotfelter, Ladd, & Vigdor, 2010).
2. Measure of teacher effectiveness	Student achievement in terms of achievement growth, studies employing a within-student across-subject estimator (analytical strategy related to a value-added approach in that it controls for average student performance across subjects, see e.g., Grönqvist & Vlachos, 2016); external observer ratings (e.g., principal, supervising teacher) of overall teaching performance	Teachers' self-reported effectiveness, student teachers' performance in terms of GPA in teacher education programs, student achievement (without controlling for prior achievement; no within-student across subject estimator)
3. Research aim	Exploring the relations between teachers' cognitive abilities and teacher effectiveness; when the effects of teachers' cognitive abilities on teacher effectiveness were explored as side effects to other more central research questions, we still included the study as long as sufficient information on the relation between cognitive abilities and teacher effectiveness were reported and focused on the variables relevant for this systematic review (e.g., Corcoran & Tormey, 2013; Harris & Sass, 2011)	Investigating correlates of either teachers' cognitive abilities or teacher effectiveness separately or addressing other research questions (e.g., differences between students and teachers' cognitive abilities)
4. Research design	Quantitative studies	Qualitative studies
5. Match between teacher and student sample	If student data included: Students had to be taught by the teachers in the teacher sample	If student data included: No match between teacher and student sample (e.g., Hanushek, Piopiunik, & Wiederhold, 2014)
6. Sufficient information on relation between cognitive abilities and teacher effectiveness	Results from bivariate correlations or single/multiple regressions with cognitive abilities as separate predictor reported	Studies were excluded that reported estimates only for a set of predictors (including cognitive abilities) together, as this does not make it possible to make statements on the (relative) importance of cognitive abilities in predicting teacher effectiveness (e.g., Atteberry et al., 2015)
7. Language	Published in English or German as the authors of the current review are fluent in these two languages	Published in another language
8. Data overlap	Sample independent of samples of other included studies	Data dependencies*

Note. SAT = Scholastic Aptitude Test; ACT = American College Testing; GRE = Graduate Record Examination; CBEST = California Basic Educational Skills Test; WEST-B = Washington Educator Skills Test-Basic; GPA = Grade Point Average; * In cases of data dependencies or overlap, our preference was to include the published study, the study with the larger sample, and the more recent and more detailed work.

For research question (a), we coded information on the instrument used to assess cognitive abilities, e.g., the specific intelligence test used or the type of college entrance exam score. If a study reported both composite scores as well as separate scores for each domain (e.g., SAT verbal and mathematics), we focused on the individual domains (category "measures of cognitive abilities"). For research question (b), we summarized the major findings concerning the associations between cognitive abilities and teacher effectiveness for all studies.[2] To answer research question (c), we turned our attention to study limitations that might affect the

[2] If sufficient information was provided by the authors, we also described the magnitude of the effects rather than solely focusing on statistical significance. Following Cohen's guidelines, we consider correlation coefficients with values over 0.10, 0.30, and 0.50 as small, moderate, and large effect sizes, respectively (Cohen, 1988). Correlations between 0.05 and 0.09 are classified as small-to-trivial relations. While correlation coefficients, which allow for a straightforward interpretation of the magnitude of effects, are commonly reported in educational/educational psychology studies, researchers working in the field of economics tend to provide coefficients from multiple regressions. Hence, in the latter case, we only reported whether cognitive abilities were significantly related to the outcome of interest and noted the direction of the effect.

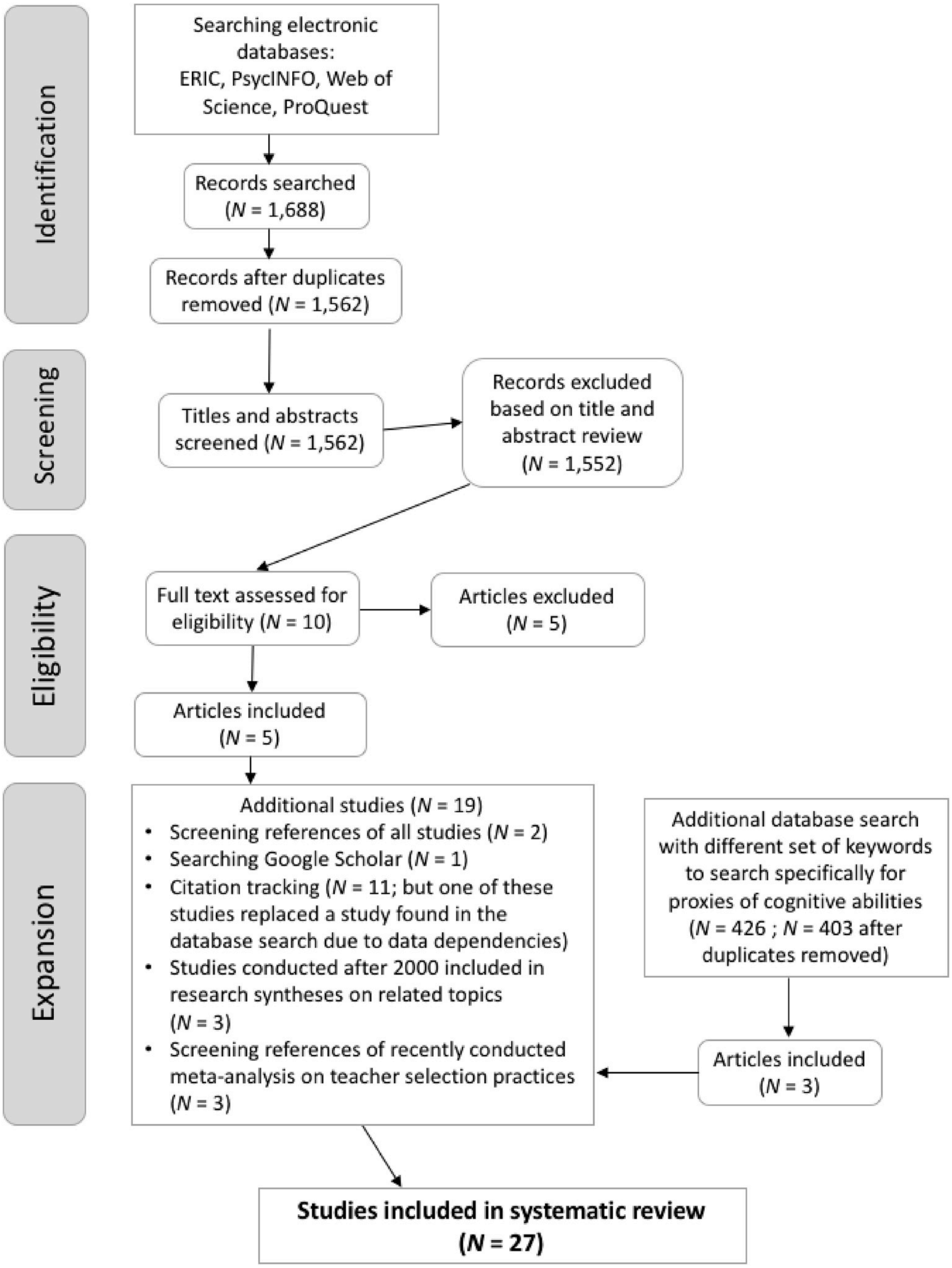

Fig. 1. Flowchart for study inclusion.

interpretation of findings and the level of confidence in any conclusions derived from the findings.

5. Results

5.1. Descriptive findings

The vast majority of the 27 studies consisted of (peer-reviewed) journal articles (19 studies, 70.37%) and 26.63% (8 studies) of the studies were unpublished (e.g., dissertations, working papers). Of the entire pool of studies, 10 (37.04%) had been published (or submitted, in the case of dissertations and conference papers, or last updated, in the case of working papers) before 2010, and 17 (62.96%) between 2010 and 2019. The bulk of research synthesized in the review had been conducted in the U.S. (22 studies,

Table 2
Descriptive information about authors and date, source, location, sample, measure of teacher effectiveness and statistical analyses for all included studies.

Authors & date	Source	Location	Sample	Measure of teacher effectiveness	Statistical analyses
Andrew et al. (2005)	Journal of Teacher Education	U.S.A.	Student teachers (elementary and secondary level); it seems that only 75–76 of 116 could be considered in analyses	Trained teacher supervisor rating of teacher effectiveness in internship	Bivariate correlations
Amato et al. (2018)	Dissertation	U.S.A.	25 teachers (small sample size due to the fact that test scores could not be matched with VA scores for many teachers)	Student achievement (VA); no subject specified	Single regression
Bieri Buschor and Schuler Braunschweig (2018)	Assessment & Evaluation in Higher Education	Switzer-land	334 student teachers (elementary level), 554 candidates, but 182 failed the test; sample further reduced as data provided by mentors was only available for 253 student teachers	Mentors' ratings of teacher effectiveness in “mathematics, German, as well as other subjects”; mentors were experienced lecturers at university with a background in teaching	Hierarchical regression, bivariate correlations
Blue et al. (2002)	Conference presentation	U.S.A.	146 student teachers (elementary and early childhood education level), subsample of 328 student teachers for whom supervisor ratings were available	Supervisor rating of teacher effectiveness (full-time faculty members of the department who supervised internships)	Bivariate correlations
Boyd et al. (2008)	Journal of Policy Analysis and Management	U.S.A.	Teachers (elementary and secondary level); no exact sample size reported; reported models rely on the sample of elementary school students and their teachers	Student achievement (VA) in mathematics	Regression models including a set of control variables; standard errors clustered at the teacher level
Buddin and Zamarro (2009a)	Journal of Urban Economics	U.S.A.	16,412 teachers, but CBEST scores only available for 2738 (elementary level)	Student achievement (VA) in mathematics and reading	Regression models including a set of control variables; level-based VA model (= models achievement controlling for prior inputs), and gains-based VA model (= models gains in achievement controlling for prior inputs), standard errors adjusted for clustering of teachers in schools
Buddin and Zamarro (2009b)	RAND Education Working paper	U.S.A.	Secondary school teachers (high school), 3164 (2017) observations reported for the analyses focusing on student mathematics achievement (English achievement)	Student achievement (VA) in mathematics and English	Regression models controlling for a set of control variables
Buddin and Zamarro (2009c)	RAND Education Working paper	U.S.A.	Secondary school teachers (middle school), different number of observations reported for different specifications (ranging from 2918 to 4941)	Student achievement (VA) in mathematics and English	Regression models controlling for a set of control variables; different specifications: level-based model, gain-based model, gain-based model with Anderson-Hsiao specification; standard errors adjusted for fact that teachers are clustered in schools
Byrnes et al. (2003)	Journal of Teacher Education	U.S.A.	68 student teachers (elementary level)	Evaluation scores given by cooperating teacher and university supervisor	Bivariate correlations, multiple regression
Corcoran and O'Flaherty (2018)	Journal of Education for Teaching	Ireland	400 student teachers (secondary level), but teacher effectiveness data was only available for 330 and 318 in Year 2 and Year 4, respectively; no SAT score equivalents were available for 52 student teachers	Evaluations of effectiveness in school placement (taking place in Year 2 and Year 4)	Ordinal logistic regression, bivariate correlations
Corcoran and Tormey (2013)	Teaching and Teacher Education	Ireland	352 student teachers (secondary level), SAT equivalent scores not available for 28 and observational data not available for 6 students	Evaluations of teacher effectiveness in school placement agreed between at least two trained and experienced supervisors employed by the university	Ordinal logistic regression, bivariate correlations
Gimbert and Chesley (2009)	Journal of School Leadership	U.S.A.	100 of 578 teachers (representing grade levels K–12) for whom Praxis 1 scores were available	Ratings of teachers' teaching ability (employee evaluation)	Single and multiple regression

(*continued on next page*)

Table 2 (*continued*)

Authors & date	Source	Location	Sample	Measure of teacher effectiveness	Statistical analyses
Goldhaber et al. (2013)	Economics of Education Review	U.S.A.	8718 elementary school teachers (WEST-B scores only available for 1469)	Student achievement (VA) in mathematics and English	Regression models controlling for a set of control variables
Goldhaber et al. (2017)	Economics of Education Review	U.S.A.	1687 teachers (secondary level), but not all teachers had valid WEST-B test	Student achievement (VA) in middle school math, ninth-grade algebra and geometry, and ninth-grade biology	Regression models controlling for a set of control variables; standard errors clustered at the teacher level
Goldhaber et al. (2018)	CEDR working paper	U.S.A.	Elementary and secondary school teachers; 1044 observations in mathematics and in English, but only 82% of these teachers had valid test scores	Student achievement in mathematics and English	Regression models controlling for a set of control variables (student and teacher covariates as well as mentor covariates, i.e., characteristics of the supervising teacher who mentored the teacher during a placement while student teaching); standard errors clustered at teacher level
Grönqvist and Vlachos (2016)	Labour Economics	Sweden	740 male teachers (secondary level) with cognitive ability data teaching Swedish, English, or mathematics; among those 272 used for identification	Main analyses: Student achievement (composite of standardized test scores in Swedish, English, and mathematics); Auxiliary analyses: Separate subjects investigated	Regression models using a within-student across-subject estimator (analytical strategy related to a VA approach in that authors controlled for average student performance across subjects), controlling for a set of variables; standard errors clustered at the teacher level
Hall and West (2011)	Issues in Educational Research	U.S.A.	74 student teachers (secondary level); 178 student teachers received invitation to participate; rights to 75 emotional intelligence tests had been purchased (relation between emotional intelligence and teaching performance was major research question)	Evaluation of teacher effectiveness by supervisors (full-time faculty members)	Bivariate correlations
Hall (2010)	Dissertation	U.S.A.	30 teachers (elementary level); relationship between cognitive ability and student achievement growth analyzed for 22 (for mathematics) and 23 (for reading) teachers, most likely due to the fact that scores for teachers who taught the same subject in more than one class were averaged	Student achievement growth in mathematics and reading (MAP - Measurement of Adaptive Progress assessment which measures individual student growth using adaptive testing); individual growth index in terms of discrepancies between actual end-of year test achievement and predicted achievement (with predicted achievement based on initial achievement levels) used that was aggregated to the class level; for classes for which data from two years was available: average of two years used	Bivariate correlations
Harris and Sass (2011)	Journal of Public Economics	U.S.A.	Elementary, middle, and high school teachers in the subjects mathematics and reading; 4160 "observations" reported for analyses using college entrance exam scores (does not reflect exact sample size as in elementary school students typically receive all of their instruction from a single teacher in a single "self-contained" classroom)	Student achievement (VA) in mathematics and reading	Regression models, controlling for a set of covariates; standard errors obtained by bootstrapping
Henry et al. (2013)	Journal of Teacher Education	U.S.A.	279 elementary school teachers	Student achievement (VA) in mathematics and reading	Hierarchical linear models controlling for a set of covariates

(*continued on next page*)

Table 2 (*continued*)

Authors & date	Source	Location	Sample	Measure of teacher effectiveness	Statistical analyses
Jacob et al. (2016)	National Bureau of Economic Research Working paper*	U.S.A.	Initial sample: "over 7000 teachers" (elementary and secondary level); only those who were hired could be considered in the analyses focusing on teaching performance as outcome; for analyses including SAT/ACT scores 917 unique teacher observations reported	IMPACT evaluation (score including principal assessment, rating based on classroom observation, and measure of student learning)	Regression models controlling for a set of covariates; standard errors clustered at the teacher level
Memory et al. (2001)	Journal of Personnel Evaluation in Education	U.S.A.	186 student teachers (secondary level programs and '"all-grade" program)	Teacher effectiveness evaluation by university supervisors	Bivariate correlations
Memory et al. (2003)	Journal of Teacher Education	U.S.A.	161 student teachers (elementary level)	Teacher effectiveness evaluation by university supervisors	Bivariate correlations
Preston (2014)	Dissertation	U.S.A.	Secondary school teachers; 986 mathematics teachers and 822 English teachers, but SAT scores only for a portion of the sample available (not further specified)	Student achievement (VA) in mathematics, English, and Algebra	Hierarchical linear models controlling for a set of covariates; standard errors clustered at teacher level
Rockoff et al. (2011)	Education Finance and Policy	U.S.A.	418 new mathematics teachers (elementary and secondary level), 333 completed entire survey, 602 had been invited to participate	a) Student achievement in mathematics (VA) and b) teacher effectiveness evaluation by mentor	Regression models controlling for a set of control variables (different control variables for the two teacher effectiveness outcomes); separate regressions for each predictor; standard errors clustered at the school level
Smith et al. (2019)	Administrative Issues Journal	U.S.A.	104 student teachers in elementary and early childhood education program	Evaluation of teaching competency by faculty members	Multiple regression and bivariate correlations
Walter and Marcel (2013)	Journal of Studies in Education	Israel	100 student teachers	Evaluations of teacher effectiveness by pedagogical instructors in the field	Bivariate correlations, path modelling

Note. VA = value-added; Please note that we classified all school types other than elementary/early childhood education as "secondary education"; *We included the working paper version and not the published version of the work by Jacob and colleagues in our review, as the authors combined measures of cognitive abilities (SAT/ACT) and other indicators (i.e., college GPA, Barron's rank, and Master's degree variables) into an "academics index" and investigated the relation between this index and teacher effectiveness in the published version.

81.48%). A handful of studies (5 studies, 18.52%) had been carried out in other countries (i.e., Ireland, Israel, Sweden, and Switzerland). Table 2 displays descriptions of the samples, the teacher effectiveness measures, and the statistical analyses.

5.2. Measures used to assess teachers' cognitive abilities

Only four studies used some form of intelligence tests to measure teachers' cognitive abilities. In the teacher education admissions test described by Bieri Buschor and Schuler Braunschweig (2018), a German version of the 'Cattell Fluid Intelligence Test' was employed to capture prospective student teachers' fluid intelligence, and thus, their ability to adapt to new situations and reason logically and independently of acquired knowledge (e.g., Cattell, 1963). In their survey study, Rockoff, Jacob, Kane, and Staiger (2011) administered Raven's Progressive Matrices (Standard Version; see e.g., Raven, 2000), a test that also measures fluid intelligence. Grönqvist and Vlachos (2016) relied on an intelligence test used in the Swedish military draft (Swedish Enlistment Battery, SEB). The SEB comprises several subtests assessing logical, verbal, and spatial abilities as well as technical understanding; the results of these subtests are then combined to produce a measure of general intelligence (Grönqvist & Vlachos, 2016; see also e.g., Carlstedt & Mårdberg, 1993; Mårdberg & Carlstedt, 1998). Data for the dissertation by J. D. Hall (2010) was gathered using the Wonderlic Personnel Test. This test, characterized by its developer as a problem-solving test (Wonderlic, 1983), represents a short test of general intelligence that includes verbal, numerical, and spatial content (e.g., Blickle, Kramer, & Mierke, 2010).

The majority of studies (i.e., 14 studies) relied on *college entrance exam scores.* For example, SAT and/or ACT composite scores were used in the studies by Byrnes, Kiger, and Shechtman (2003), P.C. Hall and West (2011), Henry et al. (2013), Preston (2014), Amato, Battles, and Beziat (2018), Jacob, Rockoff, Taylor, Lindy, and Rosen (2016), and Harris and Sass (2011). With the sole exceptions of Bieri Buschor and Schuler Braunschweig (2018) and Grönqvist and Vlachos (2016), who relied on intelligence tests, research conducted outside the U.S. utilized composite standardized test scores described as "SAT-score equivalents" (i.e., the state-administered "leaving certificate" in Ireland; Corcoran & Tormey, 2013; Corcoran & O'Flaherty, 2018, or "psychometric exam grades" in Israel; Walter & Marcel, 2013). Blue, O'Grady, Toro, and Newell (2002), Rockoff et al. (2011), and Boyd, Lankford, Loeb, Rockoff, and Wyckoff (2008) analyzed SAT verbal and mathematics scores separately. The analyses in one study were based on Graduate Record Examination (GRE) verbal, numerical, and analytical scores (Andrew, Cobb, & Giampietro, 2005).

A total of 12 studies reported the use of *basic skills test scores.* Blue et al. (2002), Gimbert and Chesley (2009), Goldhaber, Liddle, and Theobald, (2013), Henry et al. (2013), and Smith, Wageman, Anderson, Duffield, and Nyachwaya (2019) used Praxis 1 composite scores. In all of their three studies, Buddin and Zamarro (2009a,b,c) focused on California Basic Educational Skills Test (CBEST) composite scores. Separate domains were analyzed in the studies by Memory, Antes, Corey, and Chaney (2001) and Memory, Coleman, and Watkins (2003) (Praxis 1 mathematics, reading, and writing scores) and in the studies carried out by Goldhaber, Gratz, and Theobald (2017) and Goldhaber, Krieg, and Theobald (2018) (Washington Educator Skills Test-Basic [WEST-B] mathematics, reading and writing scores). Moreover, two studies employed both college entrance exam scores and basic skills tests scores (Blue et al., 2002; Henry et al., 2013) and one study relied on intelligence test scores as well as college entrance exam scores (Rockoff et al., 2011). Table 3 shows the measures of cognitive abilities used in all studies.

5.3. Main findings

Table 3 provides information regarding the main findings, grouping studies sharing the same cognitive ability measure (intelligence test vs. college entrance exam scores vs. basic skills tests) and the same teacher effectiveness measure (external observer ratings vs. student achievement) together to more effectively present information. Online Supplement S1 includes the same information in text version. In Table 4, we give an overview of combinations of measures of cognitive abilities and teacher effectiveness, separately for student teachers and practicing teachers.

5.4. Study limitations

Like all (empirical) research, the studies included in this review suffered from several limitations that need to be kept in mind when interpreting their findings. In this section, we provide an overview of the issues we deem most important, which may have potentially systematically influenced the findings or simply make it difficult to interpret empirical results of some of the studies in a straightforward and confident fashion.

Samples. If an effect of cognitive abilities exists, but is small, the question becomes whether a given study has enough power to detect this small effect. Unfortunately, a few studies had small sample sizes, raising concerns about low statistical power. Moreover, taking a closer look at the studies that reported information on the initially considered pool of participants reveals that in some studies the sample sizes were drastically reduced, e.g., due to the unavailability of administrative data for large numbers of teachers (e.g., Gimbert & Chesley, 2009). It has long been known that sample size reduction can attenuate the generalizability of findings and represents a potential threat to their validity if the teachers who could not be reached, dropped out, or were excluded differ systematically from those who remained in the sample (e.g., Miller & Hollist, 2007). Some studies explored this possibility (e.g., Rockoff et al., 2011, who report differences between survey responders and non-responders), and a few further studies attempted to account for bias arising from potential non-random teacher attrition in their analyses (e.g., Goldhaber et al., 2017); nevertheless, this remains a critical issue. In addition, participating (prospective) teachers were taken from single institutions, e.g., from one teacher education program (Henry et al., 2013), one university (e.g., Andrew et al., 2005), or one urban high-need and hard-to-staff school district (Gimbert & Chesley, 2009). Most authors were well aware of the non-representativeness of their samples and acknowledged this as a

Table 3
Results for Research Questions 1–2 (measures of cognitive abilities and main findings for all included studies).

Study	Measures of cognitive abilities	Main findings	Effect/Size of effect (size of effect based on correlation coefficients)[a]
1a. Studies using intelligence tests and external observer ratings			
Bieri Buschor and Schuler Braunschweig (2018)	Cattell General Fluid Intelligence Test: CFT-3 (for 2007/2008 tests), CFT 20-R (for 2009–2011 tests)	Cognitive abilities not a significant predictor of student teacher effectiveness in hierarchical regression (positive coefficient); non-significant positive bivariate correlation	Non-significant effect (regression); non-significant positive close-to-zero effect (correlation)
1b. Studies using intelligence tests and student achievement			
Grönqvist and Vlachos (2016)	Cognitive abilities measure from military draft (composite of logical, verbal, and spatial abilities and technical understanding)	Main analyses: Non-significant positive effect of cognitive abilities on student achievement (composite of standardized test scores in Swedish, English, mathematics); results basically unchanged across different specifications; auxiliary analyses revealed asymmetric effects across subjects for cognitive skills: negative effects for English and Swedish and close to zero positive effect for mathematics; important to note that auxiliary analyses and main analyses are not directly comparable as the auxiliary analyses were conducted to test the identifying assumptions of the main analysis and rested on different assumptions	No significant effect in main analyses; indications of negative effects in auxiliary analyses (regression)
J. D. Hall (2010)	Wonderlic Personnel Test (WPT)	Non-significant negative relation between intelligence test scores and student achievement growth in mathematics; non-significant negative relation between intelligence test scores and student achievement growth in reading	Non-significant small-to-medium negative effect for mathematics, non-significant negative effect of medium size for reading (correlations)
1c. Studies using intelligence tests and both external observer ratings and student achievement			
Rockoff et al. (2011) [b]	Raven's Progressive Matrices Standard Version (Raven's test)	Non-significant positive effects of intelligence test scores on both forms of teacher effectiveness	Non-significant effect (regression)
2a. Studies using college entrance exam scores and external observer ratings			
Andrew et al. (2005)	GRE-V (verbal), GRE-Q (quantitative), GRE-A (analytical)	Non-significant positive relations between GRE-V and GRE-Q scores and teacher effectiveness, significant positive relation between GRE-A scores and teacher effectiveness	Non-significant small positive effects for verbal and quantitative domain, significant positive effect of moderate size for analytical domain (correlations)
Blue et al. (2002) [b]	SAT verbal, SAT mathematics	Non-significant positive relation between SAT verbal scores and significant positive relation between SAT mathematics scores and teacher effectiveness	Non-significant small positive effect for verbal domain, significant small positive effect for mathematics (correlations)
Byrnes et al. (2003)	ACT composite score	Bivariate correlations*: Non-significant negative relation between ACT and teacher effectiveness rating by cooperating teacher; significant negative relation between ACT and teacher effectiveness rating by university supervisor	Non-significant negative small effect for cooperating teacher rating, negative significant effect of medium size for university supervisor rating (correlations)
Corcoran and O'Flaherty (2018)	State-administered 'Leaving Certificate' (= SAT score equivalent)	In regressions with teacher effectiveness in year 4 as outcome: significant, positive effect of SAT score equivalent (both when only SAT score equivalents entered as predictor and when SAT score equivalents and teacher effectiveness in year 2 entered as predictors); bivariate correlations: non-significant positive relation with teacher effectiveness in year 2, significant positive relation with teacher effectiveness in year 4	Significant positive effect on year 4 teaching performance (regression); non-significant positive small-to-trivial effect for Year 2 effectiveness outcome and significant positive small effect for year 4 effectiveness outcome (correlations)
Corcoran and Tormey (2013)	State-administered 'Leaving Certificate' (= SAT score equivalent)	SAT score equivalents not a significant predictor in regression, bivariate correlations: non-significant positive relation	Non-significant effect (regression), non-significant positive small-to-trivial effect (correlation)
P. C. Hall and West (2011)	ACT (composite of English, mathematics, reading, and science)	Only reported that ACT scores and teacher effectiveness rating by supervisors were not significantly correlated; no correlation coefficient reported	Non-significant effect

(*continued on next page*)

Table 3 (*continued*)

Study	Measures of cognitive abilities	Main findings	Effect/Size of effect (size of effect based on correlation coefficients)[a]
Walter and Marcel (2013)	Psychometric exam grades (= SAT score equivalent)	Non-significant positive bivariate correlation between psychometric exam grades and teacher effectiveness ratings; non-significant negative effect in path model	Non-significant effect (regression), non-significant positive close-to-zero effect (correlation)
2b. Studies using college entrance exam scores and student achievement			
Amato et al. (2018)	ACT composite scores	Non-significant negative effect	Non-significant effect (regression)
Boyd et al. (2008)	SAT mathematics, SAT verbal	Significant positive effects of SAT mathematics scores and significant negative effects of SAT verbal scores on mathematics achievement gains	Significant positive effect for SAT mathematics, significant negative effect for SAT verbal in regression
Harris and Sass (2011)	SAT scores; ACT and community college placement exam scores converted to SAT score equivalents	Mathematics: non-significant negative estimates for elementary, middle, and high school; reading: non-significant negative estimate for elementary school, non-significant positive estimates for middle and high school	Non-significant effect (regression)
Henry et al. (2013) [b]	SAT/ACT composite scores	Non-significant positive effect of SAT/ACT composite scores on achievement gains in mathematics and reading	Non-significant effect (regression)
Preston (2014)	SAT composite scores	SAT scores not included in main analyses, because SAT scores were only available for a portion of the sample and it is argued that including this covariate reduces sample sizes drastically; results of supplementary analyses including SAT scores indicate no significant effect on achievement in any of the investigated subjects (mathematics, English, and Algebra); no coefficients reported	Non-significant effect (regression)
2c. Studies using college entrance exam scores and both external observer ratings and student achievement			
Jacob et al. (2016)	SAT (composite of verbal and mathematics) or ACT equivalent	Composite of student learning measure, classroom observation rating and principal evaluation used; significant positive effect in regression with SAT/ACT scores as single predictor (plus control variables); to account for potential selection bias due to hiring, additional specifications included in single regression: effect still significant and positive when including a recommended pool by year fixed effect and a school fixed effect; when multiple predictors considered, effects of SAT/ACT scores non-significant and positive (only for specification including a school fixed effect, non-significant and negative estimate)	Significant positive effect in regression with SAT scores as single predictor, non-significant effect in multiple regression
Rockoff et al. (2011) [b]	SAT mathematics, SAT verbal	Non-significant positive effects of SAT mathematics score on mathematics achievement and subjective teacher evaluation, non-significant negative effect of SAT verbal score on mathematics achievement, non-significant positive effect of SAT verbal score on subjective teacher evaluation	Non-significant effect (regression)
3a. Studies using basic skills test scores and external observer ratings			
Blue et al. (2002) [b]	Praxis 1 composites	Significant positive relation between Praxis 1 scores and teacher effectiveness	Significant small positive effect (correlation)
Gimbert and Chesley (2009)	Praxis 1 (composite score of reading, writing, mathematics)	Praxis 1 scores do not significantly predict teacher effectiveness in single or multiple regression (negative coefficients)	Non-significant effect (regression)
Memory et al. (2001)	Praxis 1 reading, writing mathematics	Non-significant positive relation between Praxis 1 mathematics scores and teacher effectiveness; non-significant negative relation between Praxis 1 reading scores and teacher effectiveness; non-significant positive relation between Praxis 1 writing scores and teacher effectiveness	Non-significant small positive effect for mathematics, non-significant small-to-trivial positive effect for writing, non-significant close-to-zero negative effect for reading (correlations)

(*continued on next page*)

Table 3 (*continued*)

Study	Measures of cognitive abilities	Main findings	Effect/Size of effect (size of effect based on correlation coefficients)[a]
Memory et al. (2003)	Praxis 1 reading, writing mathematics	Non-significant positive relations between Praxis 1 scores in all domains and teacher effectiveness	Non-significant small-to-trivial positive effects for all three domains, i.e., mathematics, reading, and writing (correlations)
Smith et al. (2019)	Praxis 1 composite or Core Academic Skills for Educators composite (Praxis 1 transitioned to Core Academic Skills test in 2013)	Non-significant negative effect in multiple regression; non-significant positive correlation	Non-significant effect (regression), non-significant positive close-to-zero effect (correlation)
3b. Studies using basic skills test scores and student achievement			
Buddin and Zamarro (2009a)	CBEST composite score	Level-based VA model: Non-significant negative effect of CBEST on student achievement in reading, significant negative effect of CBEST scores on student achievement in mathematics; in model with other predictors (other licensure test scores) significant negative effect in both mathematics and reading; gains-based VA model: Non-significant negative effects of CBEST scores on student achievement gains in mathematics and reading; in model with other predictors non-significant negative effect in both mathematics and reading; when re-estimating the models and controlling for student and not teacher heterogeneity: significant negative effect in both subjects for level-based model and non-significant negative effect in both subjects for gain-based model	Varying depending on specification
Buddin and Zamarro (2009b)	CBEST composite score	Non-significant negative effects on student achievement (VA) in both mathematics and English	Non-significant effect (regression)
Buddin and Zamarro (2009c)	CBEST composite score	Non-significant effects in all models (level-based models, gain-based models, gain-based model with Anderson-Hsiao specification) and in both subjects (reading and mathematics) and irrespective of whether CBEST scores are investigated separately or jointly with other licensure test scores	Non-significant effect (regression)
Goldhaber et al. (2013)	WEST-B composite	WEST-B coefficients positive for both subjects (mathematics and reading) and statistically significant in mathematics	Significant positive effect for mathematics, non-significant effect for reading
Goldhaber et al. (2017)	WEST-B mathematics, WEST-B reading, WEST-B writing	WEST-B mathematics: significant positive effects for middle school mathematics; non-significant, positive effects for high school algebra and geometry; significant positive effects for high school biology (strongest effect for biology); effects quite robust across specification (e.g., significant effects still significant when school-year-grade-track fixed effects included in which teachers are compared only with teachers in the same school, year, grade, and track); mentioned that in model with all three basic skills test scores, WEST-B mathematics remained a significant predictor for student math performance, no effects for other WEST-B scores; WEST-B mathematics effect for biology also robust to controlling for other basic skills scores, but here, WEST-B writing also a significant predictor in some specifications	WEST-B mathematics: Significant positive effect for VA in mathematics, non-significant effect for VA in algebra and geometry, significant positive effect for VA in biology
Goldhaber et al. (2018)	WEST-B mathematics, WEST-B reading, WEST-B writing	None of the effects significant; negative coefficients for WEST-B mathematics and WEST-B reading in mathematics and English; negative coefficient for WEST-B writing in mathematics and positive coefficient in English	Non-significant effect (regression)

(*continued on next page*)

Table 3 (*continued*)

Study	Measures of cognitive abilities	Main findings	Effect/Size of effect (size of effect based on correlation coefficients)[a]
Henry et al. (2013) [b]	Praxis 1 reading, Praxis 1 writing, Praxis 1 mathematics	Non-significant negative effects of Praxis 1 reading and writing scores on student VA achievement in mathematics and in reading, non-significant positive effects of Praxis 1 mathematics score on student VA achievement in mathematics and reading	Non-significant effect (regression)

Note. GRE = Graduate Record Examination scores; SAT = Scholastic Aptitude Test; ACT = American College Testing; CBEST = California Basic Educational Skills Test; WEST-B = Washington Educator Skills Test-Basic; VA = value added; [a] If sufficient information was provided by the authors, we also described the magnitude of the effects rather than solely focusing on statistical significance; Following Cohen's guidelines, we consider correlation coefficients with values over 0.10, 0.30, and 0.50 as small, moderate, and large effect sizes, respectively (Cohen, 1988); Correlations between 0.05 and 0.09 are classified as small-to-trivial relations; If no correlation coefficients were reported in a study, we only note whether a significant or non-significant effect was found; [b] Indicates that the study used more than one measure of cognitive abilities and is thus listed in more than one category, *For the study by Byrnes and colleagues, we only report correlational results and not regression results due to inconsistencies in the reporting of the latter results.

Table 4

Information on the combinations of measures of cognitive abilities and measures of teacher effectiveness of the included studies, separately for teachers and student teachers.

Population: Teachers			
Measure of cognitive ability	Intelligence test	College entrance exam test score	Basic skills test score
Measure of teacher effectiveness			
Student achievement	Grönqvist and Vlachos (2016); Hall (2010); Rockoff et al. (2011) [a, b] (3)	Amato et al., (2018); Boyd et al. (2008); Harris & Sass (2011); Henry et al. (2013)[a]; Preston (2014); Rockoff et al. (2011)[a, b] (6)	Buddin and Zamarro (2009a,b,c); Goldhaber et al. (2013); Goldhaber et al. (2017); Goldhaber et al. (2018); Henry et al. (2013)[a] (7)
External observer ratings	Rockoff et al. (2011) [a, b] (1)	Rockoff et al. (2011) [a, b] (1)	Gimbert and Chesley (2009) (1)
Population: Student teachers			
Measure of cognitive ability	Intelligence test	College entrance exam test score	Basic skills test score
Measure of teacher effectiveness			
Student achievement	(0)	(0)	(0)
External observer ratings	Bieri Buschor and Schuler Braunschweig (2018) (1)	Andrew et al. (2005); Blue et al. (2002)[a]; Byrnes et al. (2003); Corcoran and O'Flaherty (2018); Corcoran and Tormey (2013); Hall and West (2011); Walter and Marcel (2013) (7)	Blue et al. (2002)[a]; Memory et al. (2001); Memory et al. (2003); Smith et al. (2019) (4)

Note. Numbers in parentheses indicate the number of studies relying on the combination of teacher effectiveness measure and cognitive ability measure represented by the category; [a] Indicates that the study used more than one measure of cognitive abilities and is thus listed in more than one category; [b] Indicates that the study used more than one measure of teacher effectiveness and is thus listed in more than one category; The measure used by Jacob et al. (2016) contains a mixture of several effectiveness measures and is therefore not included in the table.

limitation of their research. However, it clearly undermines the generalization of some of the research findings.

A related problem specifically weakened the rigor of studies focusing on the predictive validity of measures involved in selection procedures (e.g., Bieri Buschor & Schuler Braunschweig, 2018). The problem of the absence of data from unselected applicants plagues selection research –outcome data are typically not available for rejected applicants and, as an effect of the selection, the sample of selected applicants is not random and therefore not representative of the applicant population (range restriction problem, see e.g., Berk, 1983; Gross & McGanney, 1987; Thorndike, 1949; Schmidt, Oh, & Le, 2006). Importantly, the relation obtained from a range-restricted data set underestimates the relation that would be obtained from the (not available) unrestricted data set. This bias must be corrected to provide a more valid population estimate (e.g., Pfaffel, Kollmayer, Schober, & Spiel, 2016) —which none of the studies included here did.

Measurement of constructs. We further noted a set of measurement-related issues that might muddy interpretation of the impact teachers' cognitive abilities have on teacher effectiveness. For example, Blue et al. (2002) described their teacher effectiveness measure as "very preliminary efforts to develop an instrument to assess teaching" (p. 2). Hence, we do not know whether the absence of an effect—or the presence of an artificial effect—might stem from weaknesses in the predictor or outcome measure. Moreover, although most studies collected test score information from administrative data, in some instances, self-reported scores were used (e.g., Rockoff et al., 2011). Of course, we cannot know whether faulty measures biased the results, but it is a problem worth bearing in mind when interpreting and comparing findings.

Another minor issue that might nevertheless be worth noting is that Bieri Buschor and Schuler Braunschweig's (2018) study applied two different versions of an intelligence test, as the latter version was not available when the admission test was designed. In order to combine the data for the two test versions, the authors categorized the overall test scores into five groups. While we are certainly not suggesting that the statistically insignificant relation between cognitive abilities and teacher effectiveness was caused by creating these categories instead of using interval scale scores, it is important to point out that this approach led to a loss of information and loss of statistical power. It is not advisable to collapse interval scale scores into categories as this converts scores which allow for a fine-grained assessment of cognitive abilities into a less informative ordinal scale. Relatedly, without continuous information, the ability to detect a relationship decreases.

Moreover, we noticed large differences in the nature of the teacher effectiveness evaluations by external observers —e.g., in terms of the assessment tools (e.g., scales vs. rubrics, multiple vs. single item measures) and their content, the rating scales (e.g., Andrew et al., 2005, employed a 4-point rating scale ranging from “acceptable” to “outstanding”, while in Bieri Buschor & Schuler Braunschweig, 2018, student teachers were rated on a 5-point scale from “low teaching performance” to “high teaching performance”), assessors (e.g., full-time university members—mostly professors—in the study by P.C. Hall & West, 2011, vs. “pedagogical instructors in the field” in the work by Walter & Marcel, 2013), and timing of the assessment (e.g., summative assessment vs. multiple evaluations over the course of an internship). While we do not believe that this should be regarded as a limitation of the studies per se, we wish to make readers aware of this issue, as we believe that embracing and unravelling these differences and their implications are nontrivial research topics in their own right.

Statistical analyses. Two main problems influenced the quality of the statistical analyses: missing data and failure to account for measurement error. First, most studies did not report the amount of missing data at the item level (see e.g., Henry et al., 2013 for an exception with regard to missing data on Praxis scores). Moreover, information about how the authors dealt with missing data was largely lacking, and some of the approaches taken were either inappropriate (listwise deletion without further clarification) or the authors did not provide enough information to evaluate their appropriateness. Second, we observed few attempts to control for measurement error in the analyses. Obviously, a considerable number of studies used single item measures and had sample sizes too small to employ techniques such as latent variable modelling. Nevertheless, even the studies based on a larger number of participants and employing multiple indicators to assess constructs such as intelligence or teacher effectiveness neglected this issue. In addition, while studies relying on student achievement gains as measures of teacher effectiveness tended to control for a range of factors to rule out (some) of the potential threats to the validity of the results (e.g., Goldhaber et al., 2018; Jacob et al., 2016), we noted two exceptions (Amato et al., 2018; Hall, 2010) where this was not the case.

Interpretational ambiguities. Failure to provide information on the assessed constructs flawed several studies. Walter and Marcel (2013), for instance, did not provide information on how teacher effectiveness was scored or which features of teaching were evaluated. A further issue concerns inadequate reporting of results, or more precisely, the tendency of some authors to let statistical significance guide their reporting. P. C. Hall and West (2011), for instance, indicated that the correlation between cognitive abilities and teacher effectiveness was not significant but failed to provide effect size indicators, making it impossible for readers to judge the magnitude of the relation. Uncertainties in matching the student and teacher samples raise concerns about the trustworthiness of some of the findings in the review. For example, Grönqvist and Vlachos (2016) relied on a sample of teachers and their students and state that in Sweden, the same subject teacher is usually responsible for a subject throughout middle school. However, they also admit that there might be a certain degree of turnover in student-teacher matches, and, as no records were kept prior to the final year of compulsory school, they had no way to determine the number of years students and teachers had actually been in class together.

6. Discussion

The present systematic review is, to the best of our knowledge, the first in the teacher effectiveness domain to focus on teachers' cognitive abilities both in terms of intelligence as determined by intelligence tests and proxies of cognitive abilities captured by college entrance exam scores and basic skills test scores. With regard to the latter, our review stands out for including both composite scores (see e.g., D'Agostino & Powers, 2009, for a meta-analysis on composite scores, but solely for basic skills test scores), and, if available, scores in different individual domains, enabling us to draw differentiated conclusions (see Aloe & Becker, 2009, for a meta-analysis exclusively focusing on the verbal domain). Melding educational (psychology) research with research from the field of economics, examining studies from different world regions, and updating and expanding existing work by focusing on the period between 2000 and 2019, our work aimed to answer a set of questions concerning (a) the measurement of teachers' cognitive abilities, (b) findings on the role of teachers' cognitive abilities in predicting teacher effectiveness, and (c) the trustworthiness of these findings.

6.1. Teachers’ intelligence – should we not care or have we not cared enough?

Intelligence has been shown to be a valid predictor of job performance in numerous occupations (e.g., clerical employees, sales, manager, e.g., Ones et al., 2012; Bertua et al., 2005). In our review, however, we found only a small number of studies examining teachers' intelligence, suggesting that in the past two decades research on teacher effectiveness has largely ignored intelligence as a potential predictor of how well teachers perform at their job. We can only speculate on the reasons for this lack of research interest, but it has, for instance, been proposed that teacher effectiveness researchers might choose not to focus on intelligence due to a reluctance to propose that teachers are "born not made" (Harris & Rutledge, 2010)—despite the fact that scholars' understanding of intelligence as a fixed entity has been replaced by a more dynamic perspective viewing intelligence as susceptible to environmental influences (see Ritchie & Tucker-Drob, 2018, for a recent meta-analysis). On the other hand, the four studies on teacher intelligence in this review were published in or after 2010. In conjunction with the fact that we found even fewer studies on teacher intelligence conducted between 1980 and 2000 in an informal literature research, this might indicate a modest positive change in researchers' attitudes towards or interest in this topic.

With regard to the types of intelligence tests used, the research considered here was evenly divided into two studies relying on measures of non-verbal fluid intelligence capacities (Bieri Buschor & Schuler Braunschweig, 2018; Rockoff et al., 2011) and two studies employing measures of teachers' general intelligence by combining scores from several domains (Grönqvist & Vlachos, 2016; Hall, 2010). At first glance, the pattern of findings is relatively straightforward, as none of the studies found that higher cognitive abilities in the sense of intelligence make teachers more effective. The analyses by Rockoff et al. (2011) yielded no significant effect of teachers' fluid intelligence on mathematics achievement gains and teacher effectiveness evaluations by external observers. The same was true of the work by Bieri Buschor and Schuler Braunschweig (2018) examining the relation between student teachers' fluid intelligence and teacher effectiveness (external observer ratings) in a range of subjects, including mathematics and German. Interestingly, Grönqvist and Vlachos' (2016) auxiliary analyses, in which the authors investigated subjects separately, indicated a negative statistically significant effect of male teachers' general intelligence on student achievement in English, and a significantly less negative (i.e., close to zero) effect in mathematics. The other study applying a measure of general intelligence showed a small-to-medium (non-significant) negative relation for mathematics achievement growth and a moderate (non-significant) negative relation for reading achievement growth (Hall, 2010). All in all, half of the studies did not show any statistically significant effect of intelligence, whereas the other half reported negative effects. The same pattern emerges if we only consider the two studies reporting correlation coefficients which allow for a straightforward interpretation of the size of the effects (close-to zero positive effect for Bieri Buschor & Schuler Braunschweig, 2018, and moderate [reading] and small-to-moderate [mathematics] negative effects for J.D. Hall, 2010).

Does higher intelligence thus potentially make teachers even *less* effective? We clearly advise against overstating the indications of detrimental effects of intelligence, due to (a) the scarcity of research on teacher intelligence and (b) the fact that the study of teachers' cognitive abilities and teacher effectiveness is complicated by the variety of features involved (e.g., effectiveness outcome: student achievement vs. external observer ratings, investigated subjects, student teachers vs. practicing teachers, type of intelligence test, contents of subtests etc.), and (c) the presence of study limitations. In this review, we have attempted to disentangle some of these features and to identify a range of study limitations, and we aim to discuss the findings in light of these attempts. Inspecting the combinations of teacher effectiveness and cognitive ability measures reveals that negative effects surfaced only in two of the studies that used student achievement as outcome (Grönqvist & Vlachos, 2016; Hall, 2010), and these studies necessarily relied on practicing teachers given that data on student achievement is rarely available for student teachers: As student teachers mostly teach during relatively short-term placements, it is arguably difficult to gather and less useful or trustworthy to confide in student achievement test scores. Of course, we also have to keep in mind that the paucity of research on intelligence in the teacher domain and the few studies filling each "cell" of the effectiveness-cognitive abilities matrix currently make it impossible to draw firm conclusions. For example, only one study with practicing teachers and student teachers, respectively, investigated relations between external observer ratings and intelligence test scores (Bieri Buschor & Schuler Braunschweig, 2018; Rockoff et al., 2011). Hence, we do not know which picture would have emerged with more studies in these categories and more research will be necessary to refine and further elaborate our findings.

Another commonality of the studies finding negative effects vs. no (statistically significant) effects relates to the fact that the former analyzed separate subjects (Grönqvist & Vlachos, 2016 at least in the auxiliary analyses). On the other hand, Rockoff et al. (2011) restricted their focus to mathematics and Bieri Buschor and Schuler Braunschweig (2018) did not differentiate between evaluations made in different subjects. While we generally see a need for empirical studies systematically filling and crossing different cells of the cognitive ability-teacher effectiveness matrix and aiming to minimize potential bias in their designs, analyses, and reporting, we believe that future research should systematically examine potential subject-specific patterns of relations.

Moreover, we want to stress that with regard to study limitations that might have biased the findings, we are particularly concerned about the study by J. D. Hall (2010). As an ambitious Ph.D. dissertation project, the work addressed highly relevant questions concerning the relations between teacher selection tools and student achievement gains; however, in addition to using a different measure of student achievement growth than the other studies included in this review,[3] the small sample size, inconsistencies in the statistical analyses and reporting, and the lack of control for potentially confounding factors are worth noting. Accordingly, and echoing the need for more external observer-studies, we also envision that future studies with sound research designs focusing on student achievement gains and relying on VA models should take up the question of whether teachers' intelligence is predictive of student outcomes. As VA measures are a common feature of economics studies, we particularly encourage researchers in this domain to pay increased attention to teachers' intelligence in their studies on observable teacher characteristics. On a related note, researchers might also consider student surveys as measures of teacher effectiveness, and thus add a further cell to

[3] The author relied on scores from adaptive tests and used an individual growth index in terms of discrepancies between actual end-of year test achievement and predicted achievement (with predicted achievement based on initial achievement levels), that was aggregated to the class level (see Table 2 in the results section).

the teacher effectiveness-cognitive ability-matrix. For example, research in the MET (Measures of Effective Teaching) project has shown that student surveys of teachers' effective practices are predictive of student achievement gains and can produce even more consistent results than classroom observations or achievement gain measures (Bill & Melinda Gates Foundation, 2012; Kane & Staiger, 2012), hence pointing towards the value of student surveys in future research following up on our systematic review.

Lastly, can we offer practical implications of our findings on the role of teachers' and student teachers' intelligence for their performance in the classroom? The mixed findings of our review (e.g., close-to-zero and moderate negative effect) and the open questions that remain mainly due to the small number of studies let us at this point advise against the use of intelligence tests in teacher selection and the selection of candidates for teacher education programs, at least for decision making. This recommendation thus sharply contradicts those for other professions (see e.g., Schmidt, 2002, claiming to 'select on intelligence'); however, thousands of studies form a valid evidence base for other professions whereas only four studies could be summarized in the present work. For now, we thus see the major future task related to 'teacher intelligence' in the research domain, with multiple implications for future research, and believe it would be premature and even misleading to demand an increased focus on teachers' intelligence in applied settings.

6.2. Proxies of cognitive abilities – the promise of a differentiated view?

Compared to studies employing intelligence tests, a larger number of studies assessed proxies of teachers' cognitive abilities, with most studies relying on college entrance exam scores. Studies on college entrance exam scores and basic skills tests tended to focus on composite scores and fewer studies investigated separate domains, such as verbal versus numerical (i.e., only 4 out of 15 for college entrance exam test scores and 5 out of 12 for basic skills test scores). For college entrance exam test scores, more than half of the studies did not find any effects on teacher effectiveness. The studies identifying some effects and a) reporting effect size indicators in terms of correlation coefficients revealed mainly small and positive effects (Andrew et al., 2005; Blue et al., 2002; Corcoran & O'Flaherty, 2018; but see Byrnes et al., 2003, for a moderate negative effect), b) reporting regression results yielded either indications of statistically significant positive findings (Jacob et al., 2016, at least in the single regression) or contrasting results (statistically significant positive vs. negative for numerical vs. verbal test scores, Boyd et al., 2008). The majority of studies with basic skills test scores did not show any effect on teacher effectiveness either (but see Blue et al., 2002, and the numerical domain in Memory et al., 2001, for small positive effects size indicators in terms of correlation coefficients; see Goldhaber, Liddle, & Theobald, 2013, and Goldhaber et al., 2017, for regression results yielding some positive statistically significant findings). All in all, the non-existent to, at best, small positive effects of college entrance exam test scores and basic skills test scores square well with the findings of prior quantitative syntheses (Aloe & Becker, 2009; D'Agostino & Powers, 2009).

At this juncture, readers might worry about the credibility of conclusions based on the studies included in this review, given that we have pointed out that most of them seem to be flawed and some seem to be seriously flawed. Still, across countries, levels of schooling, stages of career, and despite the limitations of single studies, such as a lack of generalizability, the message is clear: cognitive abilities—or at least cognitive abilities as assessed in our studies—do not seem to predict teacher effectiveness very well. In terms of practical implications, our findings indicate that non-cognitive predictors evaluated in other studies, such as motivation, might be more relevant (e.g., Kim, Jörg & Klassen, 2019; Klassen & Tze, 2014). This further suggests that teacher selection procedures should probably attach more weight to (prospective) teachers' non-cognitive rather than cognitive characteristics.

Nevertheless, the overall results for proxies of cognitive abilities might conceal some interesting heterogeneities, and we now turn our attention to studies examining separate domains instead of composite scores. For college entrance exam test scores, the preponderance of studies employing this method found that the (strength of the) relations to teacher effectiveness varied across domains (Andrew et al., 2005; Blue et al., 2002; Boyd et al., 2008). The picture is less distinct for basic skill test scores, as some of the studies using separate domains showed domain-dependent differences (e.g., Goldhaber et al., 2017), but others did not (e.g., Goldhaber et al., 2018; Memory et al., 2003).

Can we identify domains that are likely to be (more) important for teacher effectiveness? There is some indication that future studies would do well to specifically explore cognitive abilities in the numerical domain, although the available evidence synthesized here does not provide a definitive answer: In sum, small positive effect sizes in terms of correlation coefficients were obtained in the studies by Blue et al. (2002), Andrew et al. (2005) and Memory et al. (2001), and statistically significant positive effects for teachers' numerical abilities were found in the regression studies by Goldhaber et al. (2017) and Boyd et al. (2008). On the other hand, the regression results of Henry et al. (2013), Rockoff et al. (2011), and Goldhaber et al. (2018) indicate non-significant findings, and Memory et al. (2003) report small-to-trivial correlations for all investigated domains.

Under which circumstances might numerical abilities be more likely to produce an effect? An intuitively appealing expectation would be that teachers with strong numerical abilities contribute to student performance in mathematics. Of the studies that investigated students' mathematics achievement gains relying on VA measures, two support this notion (Goldhaber et al., 2017, with a significant effect for middle school mathematics, but no significant effect for high school algebra and geometry; Boyd et al., 2008), but three do not (Henry et al., 2013; Goldhaber et al., 2018; Rockoff et al., 2011, who used self-reported SAT scores). Research on other subjects is too scarce to make comparisons; for example, only one study included biology (Goldhaber et al., 2017). Noteworthy, in the work of Goldhaber et al. (2017), the positive effect of WEST-B mathematics for student performance in biology exceeded those for mathematics as well as algebra and geometry.

Comparing different domains of proxies of cognitive abilities, such as numerical vs. verbal abilities, might uncover further interesting insights. In several of the examined studies, numerical abilities appeared to be more strongly related to teacher effectiveness than verbal abilities. For instance, in the study by Boyd et al. (2008), SAT mathematics scores turned out to be statistically significant positive predictors of student achievement growth in mathematics, whereas SAT verbal scores had a statistically significant negative impact. Goldhaber et al. (2017) and Blue et al. (2002) found larger positive effects and/or more significant effects for numerical abilities than verbal abilities. However, the correlation coefficients differed negligibly in the studies by Andrew et al. (2005) and Memory et al. (2001, 2003). Similarly, Henry et al. (2013) and Rockoff et al. (2011) reported non-significant regression results for both verbal and numerical scores.

Based on the current state of research summarized here, we clearly do not want to over-emphasize the value of numerical abilities and instead, we suggest future research to carefully examine whether they can truly add something—and maybe something more than verbal abilities—to the prediction of teacher effectiveness. Still, in combination with Aloe and Becker's (2009) meta-analytical findings indicating average effects of college entrance exam scores in the verbal domain on teacher effectiveness not significantly different from zero, we wish to highlight the *potential* usefulness of further exploring relations between numerical abilities and teacher effectiveness. The reasons why numerical abilities might exert (some) effects on teacher performance are hard to establish from the studies included here. We cautiously propose that teachers with higher numerical abilities might also possess specific qualities or be better able to develop specific skills, such as the ability to organize and structure information, present complex ideas clearly, or explain subject matter in a logical way (e.g., Van de Cavey & Hartsuiker, 2016). These qualities and skills then, in turn, might influence their effectiveness–maybe at least under certain circumstances, in certain subjects, with certain groups of students and with regard to certain areas of teaching performance and performance assessment etc. Testing these speculations and investigating whether and how other cognitive as well as non-cognitive attributes transmit, i.e., mediate, effects of numerical abilities on teacher effectiveness could be an interesting direction for future research. We moreover suggest that further research on proxies of cognitive abilities building on our review would particularly benefit from (a) continuing research on separate domains instead of composite scores, (b) examining and contrasting effects for different subjects, and (c) widening the scope to include other subjects than those typically investigated in studies relying on VA student achievement (i.e., mathematics and English or reading), such as biology (Goldhaber et al., 2017), (d) fill all—and especially the currently under-represented—"cells" of the effectiveness-cognitive ability matrix, e.g., external observer studies with practicing teachers and domains of basic skill test scores (see Online Supplement S2 for further suggestions for future research on teachers' cognitive abilities not directly related to our findings).

7. Limitations

This review is limited in several respects. First, although we specifically searched for grey literature (e.g., in databases indexing dissertations and conference presentations) and included working papers, our sample mainly comprised published work. Nonetheless, a large number of the published studies reported statistically nonsignificant results, making it unlikely that publication bias distorted our findings. Second, readers might feel some unease about the fact that more studies were identified in the expanded search than in the initial database search. However, following the authors of another recent review (Schrijvers, Janssen, Fialho, & Rijlaarsdam, 2018), we argue that this demonstrates the importance of conducting a search via a variety of sources, such as citation tracking and screening the references of included studies and reviews on related topics. Third, our work focused on the relation between cognitive abilities and teacher effectiveness. This generated new and vital insights into a not-well understood topic in educational research, but inevitably precluded an examination of other pivotal educational outcomes, such as students' motivation, perceptions of instructional practices, or teachers' relatedness with students (e.g., Bardach, Lüftenegger, Yanagida, Spiel, & Schober, 2019; Bardach, Yanagida, Schober, & Lüftenegger, 2018; Klassen, Perry, & Frenzel, 2012; Schweder, 2019). Fourth, we have to be aware that test scores on cognitive ability measures do not only reflect individual differences in cognitive abilities, but also motivation —and particularly in low-stakes testing situations, test motivation can act as a confounding variable (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011). Fifth, this review centered on 'raw' intelligence and its proxies. It is thus restricted to abilities and skills covered by these constructs and assessed by their measurement vehicles, but necessarily excludes other cognitive features, e.g., teachers' content knowledge and pedagogical content knowledge (e.g., Kleickmann et al., 2013), alternative understandings of intelligence (e.g., Sternberg's theory of successful intelligence, e.g., Sternberg, 2015), and related constructs (e.g., creativity, Davidovitch & Milgram, 2006). While we think the focus on the types of cognitive abilities covered in this review is still an important and worthy one, future research in the teacher effectiveness domain could expand the scope of research on teachers' cognitive abilities and move beyond intelligence and its proxies to embrace other possible perspectives on and explore further components of cognitive abilities.

Funding

This work is supported by the European Research Council (grant #647234 SELECTION).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2020.100312.

References

Aloe, A. M., & Becker, B. J. (2009). Teacher verbal ability and school outcomes: Where is the evidence? *Educational Researcher, 38*, 612–624.
*Amato, J., Battles, J., & Beziat, E. (2018). *Teacher preparation in Tennessee: Examining the relationship between teachers' ACT scores and first-year teacher effectiveness in three teacher education programs*(Doctoral Dissertation). Available from: ProQuest Dissertations and Theses database UMI No. 10814029.

*Andrew, M. D., Cobb, C. D., & Giampietro, P. J. (2005). Verbal ability and teacher effectiveness. *Journal of Teacher Education, 56*, 343–354.
Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open, 1*, 1–23.
Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*, 37–65.
Bardach, L., Lüftenegger, M., Yanagida, T., Spiel, C., & Schober, B. (2019). Achievement or agreement - Which comes first? Clarifying the temporal ordering of achievement and within-class consensus on classroom goal structures. *Learning and Instruction, 61*, 72–83.
Bardach, L., Yanagida, T., Schober, B., & Lüftenegger, M. (2018). Within-class consensus on classroom goal structures - Relations to achievement and achievement goals in mathematics and language classes. *Learning and Individual Differences, 67*, 68–90.
Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American Sociological Review, 48*, 386–398.
Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology, 78*, 387–409.
*Bieri Buschor, C., & Schuler Braunschweig, P. (2018). Predictive validity of a competence-based admission test - mentors' assessment of student teachers' occupational aptitude. *Assessment & Evaluation in Higher Education, 43*, 640–651.
Bill, & Melinda Gates Foundation (2012). *Measures of Effective Teaching (MET) Project. Asking students about teaching: Student perception surveys and their implementation.* Seattle, WA: Bill & Melinda Gates Foundation.
Bill, & Melinda Gates Foundation (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET project's three-year study.* Seattle, WA: Bill & Melinda Gates Foundation.
Blickle, G., Kramer, J., & Mierke, J. (2010). Telephone-administered intelligence testing for research in work and organizational psychology. *European Journal of Psychological Assessment, 26*, 154–161.
*Blue, T. W., O'Grady, R. J., Toro, J. A., & Newell, E. A. (2002). How do we find the best teachers? A study of the relationships among SAT, GPA, Praxis series test scores, and teaching ratings. *Paper presented at the annual meeting of the association of teacher educators* (Denver, Colorado).
Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences, 113*, 13354–13359.
*Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management, 27*, 793–818.
*Buddin, R., & Zamarro, G. (2009a). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics, 66*, 103–115.
*Buddin, R., & Zamarro, G. (2009b). *Teacher effectiveness in urban high schools.* RAND Education Working Paper No. 693-IES.
*Buddin, R., & Zamarro, G. (2009c). *Teacher qualifications and middle school student achievement.* RAND Education Working Paper No. 671-IES.
*Byrnes, D. A., Kiger, G., & Shechtman, Z. (2003). Evaluating the use of group interviews to select students into teacher-education programs. *Journal of Teacher Education, 54*, 163–172.
Carlstedt, B., & Mårdberg, B. (1993). Construct validity of the Swedish enlistment battery. *Scandinavian Journal of Psychology, 34*, 353–362.
Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, UK: Cambridge University Press.
Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1.
Cattell, R. B. (1971). *Abilities: Their structure and growth.* Boston, MA: Houghton Mifflin.
Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources, 45*, 655–681.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.
Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher, 45*, 378–387.
*Corcoran, R. P., & O'Flaherty, J. (2018). Factors that predict pre-service teachers' teaching performance. *Journal of Education for Teaching, 44*, 175–193.
*Corcoran, R. P., & Tormey, R. (2013). Does emotional intelligence predict student teachers' performance? *Teaching and Teacher Education, 35*, 34–42.
Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives, 8*, 1–44.
Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching.* Washington, D.C.: Centre for American Progress.
Davidovitch, N., & Milgram, R. M. (2006). Creative thinking as a predictor of teacher effectiveness in higher education. *Creativity Research Journal, 18*(3), 385–390.
Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences, 108*, 7716–7720.
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.
Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Research, 87*, 35–70.
Ferguson, R. F., & Danielson, C. (2015). How framework for teaching and tripod 7Cs evidence distinguish key components of effective teaching. In T. Kane, K. Kerr, & R. Pianta (Eds.). *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* (pp. 98–143). San Francisco, CA: John Wiley & Sons.
Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*, 373–378.
*Gimbert, B. G., & Chesley, D. (2009). Predicting teacher success using teacher selection practices and classroom performance assessment. *Journal of School Leadership, 19*, 49–80.
*Goldhaber, D., Gratz, T., & Theobald, R. (2017). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student STEM achievement and course-taking. *Economics of Education Review, 61*, 112–129.
Goldhaber, D., Krieg, J., & Theobald, R. (2018). Effective like me? Does having a more productive mentor improve the productivity of mentees? *CEDR Working* Paper No. 11232018-1-1.
*Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review, 34*, 29–44.
Gottfredson, L. S. (1997). Intelligence and social policy. *Intelligence, 24*, 1–12.
Greenwald, R., Hedges, L., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361–396.
*Grönqvist, E., & Vlachos, J. (2016). One size fits all? The effects of teachers' cognitive and social abilities on student achievement. *Labour Economics, 42*, 138–150.
Gross, A. L., & McGanney, M. L. (1987). The restriction of range problem and nonignorable selection processes. *Journal of Applied Psychology, 72*, 604–619.
Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.
Gustafsson, J. E. (2002). Measurement from a hierarchical point of view. In H. L. Braun, D. G. Jackson, & D. E. Wiley (Eds.). *The role of constructs in psychological and educational measurement* (pp. 73–95). Mahwah, NJ: Erlbaum.
Gustafsson, J. E., & Snow, R. E. (1997). Ability profiles. In R. F. Dillon (Ed.). *Handbook on testing* (pp. 107–135). Westport, CT: Greenwood Press.
Hall, P. C. (2009). *Potential predictors of student teaching performance: Considering emotional intelligence*(Doctoral dissertation). Available from: ProQuest Dissertations and Theses database UMI No. 3353948.
*Hall, J. D. (2010). *Fourth and fifth grade student achievement and possible relationship to three teacher selection tools*(Doctoral dissertation). Available from: ProQuest Dissertations and Theses database UMI No. 3353948.
*Hall, P. C., & West, J. H. (2011). Potential predictors of student teaching performance: Considering emotional intelligence. *Issues in Educational Research, 21*, 145–161.
Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., et al. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*, 461–487.
Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*, 141–164.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). *The value of smarter teachers: International evidence on teacher cognitive skills and student performance. Program on Education Policy and Governance Working Papers Series PEPG 14-06*.
Harris, D. N., & Rutledge, S. A. (2010). Models and predictors of teacher effectiveness: A review of the literature with lessons from (and for) other occupations. *Teachers College Record, 112*, 914–960.
*Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*, 798–812.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Park Square, OX: Rutledge.
*Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., & Lys, D. B. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education, 64*, 439–453.
Hill, H. C., Charalambous, C. Y., & Chin, M. J. (2018). Teacher characteristics and student learning in mathematics: A comprehensive assessment. *Educational Policy,* 1–32 00(00).
Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.
*Jacob, B., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2016). *Teacher applicant hiring and teacher performance: Evidence from DC public schools.* National Bureau of Economic Research Working Paper No. 22054.
Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill and Melinda Gates Foundation.
Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Bill & Melinda Gates Foundation.
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). (3rd ed.). New York, NY: Routledge.
Kim, L., 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.
Klassen, R. M., & Kim, L. E. (2017). Assessing critical attributes of prospective teachers: Implications for selection into initial teacher education programs. In D. W. Putwain, & K. Smart (Eds.). *British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education* (pp. 5–22). Oxford: Wiley.
Klassen, R. M., & Kim, L. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.
Klassen, R. M., Perry, N. E., & Frenzel, A. C. (2012). Teachers' relatedness with students: An underemphasized component of teachers' basic psychological needs. *Journal of Educational Psychology, 104*, 150–156.
Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59–76.
Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., et al. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education, 64*, 90–106.
Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence, 36*, 153–160.
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.
Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after spearman's (1904) "'General intelligence', objectively determined and measured". *Journal of Personality and Social Psychology, 86*, 96–111.
Mårdberg, B., & Carlstedt, B. (1998). Swedish Enlistment Battery (SEB). Construct validity and latent variable estimation and profile prediction of cognitive abilities by the CAT-SEB. *International Journal of Selection and Assessment, 6*, 107–114.
*Memory, D. M., Antes, R. L., Corey, N. R., & Chaney, D. E. (2001). Should Tougher Basic Skills requirements be viewed as a means of strengthening the teaching force? *Journal of Personnel Evaluation in Education, 15*, 181–191.
*Memory, D. M., Coleman, C. L., & Watkins, S. D. (2003). Possible tradeoffs in raising basic skills cutoff scores for teacher licensure: A study with implications for participation of African Americans in teaching. *Journal of Teacher Education, 54*, 217–227.
Miller, R., & Hollist, C. (2007). Attrition bias. In N. Salkind (Ed.). *Encyclopedia of measurement and statistics* (pp. 57–60). Thousand Oaks, CA: SAGE Publications.
Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation, 12*, 53–74.
Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*, 231–256.
Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., et al. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*, 130–159.
Ones, D. S., Dilchert, S., Viswesvaran, C., & Salgado, J. F. (2012). Cognitive ability. In N. Schmitt (Ed.). *The Oxford handbook of Personnel assessment and selection* (pp. 179–224). New York, NY: Oxford University Press.
Pfaffel, A., Kollmayer, M., Schober, B., & Spiel, C. (2016). A missing data approach to correct for direct and indirect range restrictions with a dichotomous criterion: A simulation study. *PloS One, 11*(3), e0152330.
*Preston, C. (2014). *Relationships between teacher preparation and beginning teacher quality in North Carolina public universities*(Doctoral Dissertation). Available from: Vanderbilt University Electronic Theses and Dissertations 07182014-093908.
Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1–48.
Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science, 29*, 1358–1369.
*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.
Rostker, L. E. (1945). The measurement of teaching ability. Study number one. *The Journal of Experimental Education, 14*, 6–51.
Rotherham, A. J., & Mead, S. (2003). Teacher quality: Beyond no child left behind. A response to Kaplan and Owings (2002). *NASSP Bulletin, 87*, 65–76.
Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics, 125*, 175–214.
Rowan, B. (1994). Comparing teachers' work with work in other occupations: Notes on the professional status of teaching. *Educational Researcher, 23*, 4–21.
Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15*, 187–210.
Schmidt, F. L. (2015). Select on intelligence. In E. A. Locke (Ed.). *Handbook of Principles of organizational behavior* (pp. 3–14). Hoboken, NJ: John Wiley & Sons.
Schmidt, F. L., Oh, I. S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology, 59*, 281–305.
Schrijvers, M., Janssen, T., Fialho, O., & Rijlaarsdam, G. (2018). Gaining insight into human nature: A review of literature classroom intervention studies. *Review of Educational Research, 89*, 3–45.
Schweder, S. (2019). Mastery goals, positive emotions and learning behavior in self-directed vs. teacher-directed learning. *European Journal of Psychology of Education,* 1–19.
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.
Shulman, L. S. (2004). Professional development: Leaning from experience. In S. Wilson (Ed.). *The wisdom of practice: Essays on teaching, learning, and learning to teach* (pp. 503–522). San Francisco, CA: Jossey-Bass.
*Smith, K., Wageman, J. J., Anderson, S. K., Duffield, S., & Nyachwaya, J. (2019). The relationship of academic indicators and professional disposition to teaching skills: A secondary data analysis. *Administrative Issues Journal: Connecting Education, Practice, and Research, 8*, 106–118.
Spearman, C. (1904). "General intelligence", objectively determined and measured. *American Journal of Psychology, 15*, 201–292.
Sternberg, R. J. (2012). Intelligence. *Dialogues in Clinical Neuroscience, 14*, 19–27.
Sternberg, R. J. (2015). Successful intelligence: A model for testing intelligence beyond IQ tests. *European Journal of Education and Psychology, 8*, 76–84.
Sternberg, R. J., & Grigorenko, E. L. (2002). *The general factor of intelligence: How general is it?* Mahwah, NJ: Lawrence Erlbaum.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques.* New York, NY: Wiley.
Van de Cavey, J., & Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition, 146*, 172–184.
*Walter, O., & Marcel, A. (2013). Emotional intelligence and teaching self-efficacy as predictors of teaching ability. *Journal of Studies in Education, 3*, 57–74.
Wayne, A. M., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*, 89–122.
Wonderlic, E. F. (1983). *Wonderlic Personnel test manual.* Northfield, IL: E. F. Wonderlic & Associates.