

Education Quality and
Accountability Office



Literature Review of the Empirical Evidence on the Connection Between Compulsory Teacher Competency Testing and Student Outcomes

August 2019

Contents

Executive Summary	3
An Overview of Historic and Current Teacher Competency Testing	5
Principle Aims of Teacher Competency Testing.....	5
Standardized Tests as a Method for Identifying Teachers of High Quality	6
Standardized Tests as a Method for Identifying Teachers of Low Quality.....	6
Subject-Specific Implementation of Standardized Testing	7
Timing of Teacher Testing.....	7
Measurement of Competency in Standardized Tests.....	7
Paths for Candidates Who Are Unsuccessful in Standardized Tests	8
Competency Test Design and Delivery Process	8
Current Test Formats and Included Items	8
Teacher Competency Testing and Student Outcomes.....	9
Challenges in Measuring the Relationship Between Teacher Competency and Student Outcomes	9
Evidence Linking Teacher Competency Testing and Student Outcomes.....	9
Associated Impacts of Teacher Testing	12
The Influence of Standardized Tests on Diversity and Inclusion	13
Teacher Characteristics and Evaluation Procedures That Predict Student Outcomes..	13
Conclusion	15
References.....	16

Executive Summary

This report provides a review of the current research and evidence available on the topic of the compulsory standardized testing of teachers and the relationship between these tests and student achievement. First, this report provides a brief overview of the history and current methodologies of teacher competency testing, highlighting key issues and relating these to policy concerns. Here the types of teacher competency tests are outlined, and a summary of the design and purpose of these tests is presented. This report then provides a summary of the empirical evidence concerning the relationship between standardized teacher competency testing and student achievement. This summary concentrates on key sources containing findings most relevant to public policymakers. Included in this section is an overview of some of the research about impacts on outcomes other than student achievement. Wherever possible, the report focuses on teacher testing in the area of mathematics. The report concludes with a brief overview of the research on other performance review processes and professional development programs that have shown an impact on achievement.

There is little question among researchers that better teachers lead to better student outcomes (Aaronson, Barrow, Sander, 2007; Coleman, 1968; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). However, as we attempt to establish a uniform strategy to measure teacher performance and correlate “successful” characteristics with student outcomes, the evidence becomes more mixed. In a review of the available evidence, a distinction must be made among teachers’ content knowledge, pedagogical content knowledge and performance on standardized tests to assess the value of each of these variables accurately. Additional questions are raised about the validity, generalizability and robustness of using a standardized test to measure either pure or pedagogical content knowledge.

Current available evidence shows that two fundamental types of a teacher’s subject matter knowledge directly influence student outcomes.¹ The first is good content knowledge (i.e., a foundation in and understanding of the subject) that would be familiar to anyone working in the subject area. The second is pedagogical content knowledge, which requires an understanding of how to teach the subject matter effectively.

For example, understanding how to multiply fractions is content knowledge, whereas understanding how to teach the multiplication of fractions is pedagogical content knowledge (Ball, Thames & Phelps, 2008; Rowland, Huckstep, & Thwaites, 2005). Research demonstrates that these types of content knowledge are related to student outcomes (Carlisle & Berebitsky, 2011; Mullens, Murnane, & Willett, 1996). However, pedagogical content knowledge is more strongly correlated with positive student outcomes (Hill, Kapitula & Umland, 2011; Kelcey and Carlise, 2013). Furthermore, pedagogical content knowledge has a mediating effect on pure content knowledge (Baumert, Kunter, Blum, Brunner, Voss, Jordan, Klusmann, Krauss, Neubrand & Tsai, 2010; Kelcey, Hill & Chin, 2019; Kersting, Givvin, Thompson, Santagata & Stigler, 2012). Content knowledge affects student outcomes differently depending on the level of pedagogical content knowledge.

¹ For the purposes of this review, student outcomes are measured by student performance on standardized achievement tests. These include both international achievement tests such as those of the Programme for International Student Assessment ([PISA](#)) and Trends in International Mathematics and Science Study ([TIMSS](#)) and jurisdiction-specific achievement tests such as the [National Assessment of Educational Progress \(NAEP\)](#) in the U.S. or the [IOWA](#) tests.

The conclusions to be drawn from research exploring the specific correlation between mathematical content knowledge and student outcome (as measured by standardized student tests) are not clear. The majority of studies demonstrate a positive relationship between content knowledge and student outcome in mathematics (Baumert et al., 2010; Campbell, Nishio, Smith, Clark, Conant, Rust, DePiper, Frank, Griffin & Choi, 2014; Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Hill, Rowan, & Ball, 2005; Rockoff, Jacob, Kane, & Staiger, 2011; Shechtman, Roschelle, Haertel, & Knudsen, 2010; Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J., 2010). However, other studies have found no significant connection between a teacher's mathematical knowledge and student outcomes (Kersting et al., 2012).

Conversely, the literature examining the connection between teachers' pedagogical mathematical knowledge and student outcomes is almost uniformly positive (Kelcey & Carlise, 2013; Hill et al., 2011; Kersting et al., 2012). Research using teacher observation base these findings on the teacher's abilities (1) to develop effective mathematics questions (Ball et al., 2008), (2) to present topics clearly and confidently (Hill et al., 2008) and (3) to provide constructive response to student questions and observable errors in classroom work (Ball et al., 2008).

There is much evidence to suggest that the relationship among content knowledge, pedagogical content knowledge and student achievement is further influenced by other factors. These factors can include curriculum materials (Shechtman et al., 2010), socioeconomic measures (Baumert et al., 2010), academic/non-academic tracks (Baumert et al., 2010) and the school or district environment (Corey, Phelps, Ball, Demonte, & Harrison, 2012; Petscher & Logan 2014; Tighe & Schatschneider, 2016).

The body of evidence examining the connection among standardized teacher competency testing on both pure content knowledge and pedagogical content knowledge and student outcomes is mixed. Many studies find that teacher competency test scores are positively related to student outcomes (Goldhaber, 2007; Kelcey et al., 2019; D'Agostino & Powers, 2009). However, almost all these analyses find this correlation to be weak and not universal. Moreover, research using multiple predictors of student outcomes observe that teacher competency test scores are the weakest predictor of student outcomes compared to (1) teacher experience, (2) teacher education college or university grades, (3) number of post-secondary math courses and (4) professional math certification courses for teachers (Clotfelter, Ladd, & Vigdor, 2010; D'Agostino & Powers, 2009; Ferguson & Womack, 1993; Shuls & Trivitt, 2015).

Teacher testing impacts more than just student outcomes and can influence the teacher recruitment pipeline and the curriculum taught to trainee teachers. Changes in these factors can cause a decrease in the enrolment of pre-service teachers from minority ethnic groups (Cobb, Shaw, Millard, & Bomotti, 1999; Graham, 2013; Nettles, Scatton, Steinberg, & Tyler, 2011; Petchauer, 2012; Shuls & Trivitt 2015).

An Overview of Historic and Current Teacher Competency Testing

In order to understand why policymakers and educators continue to support and undertake standardized teaching tests, it is essential to examine the history of this process. The use of standardized competency testing is not a new concept (Angrist & Guryan, 2008). Since the 19th century, teacher testing has been used primarily in the United States, yet it has also existed in several other countries around the world (Angrist and Guryan, 2008). The use of competency testing has fluctuated throughout the past century due to such factors as supply and demand for teachers, local political will, and public opinion on the issue (De Forest, 2006; Haney, Madaus, & Kreitzer, 1987; Santibanez, 2006).

Currently, standardized teacher competency testing is used in most states in the United States (Goldhaber, 2007), across [Australia](#) and in a few countries in Europe and Asia, such as the [United Kingdom](#) and [Saudi Arabia](#), although the last two countries are considering removing this policy (Gibb, 2019). Ontario had compulsory teacher testing for the period 2002–2004 but does not currently use such tests (Glassford, 2005; Glassford, 2007).

In the context of this literature review, a careful distinction must be made between teacher testing and teacher certification/training. The standardized competency test can exist in either paper or digital form and is usually multiple-choice. The principle aim of the test is to evaluate the participant's basic competency in the subject matter taught in public schools. These tests are therefore designed to measure an aspiring teacher's knowledge and skills and are used by educational institutions to confirm that educators are qualified to teach. These tests are separate from the exams or assessments within the college or university teacher education programs (Goodman, Arbona, & Dominguez de Rameriz, 2008).

Principle Aims of Teacher Competency Testing

Although the U.S. government now uses of standardized tests for trainee teachers, the use of such tests originates from the higher education community itself in an attempt for the community to better evaluate performance on teacher training programs (Pullin, 2017). In Canada, recent mandatory teacher testing arose as part of an effort to increase transparent accountability to the public and to ensure the workforce was competent in the teaching of the most up-to-date content possible (Glassford, 2007).

Scholars posit that one of the fundamental purposes of the mandatory teacher competency tests is to provide a valid indicator of teacher preparation. Inherent to this purpose is the evaluation of the effectiveness of teacher training programs. An educational system that has confidence in the functioning of a teacher training network, and as an extension, the trained teachers, might not require such standardized teacher testing (Goodman et al., 2008).

The primary explicit functions of standardized competency testing are to identify quality teachers, weed out unqualified teachers, and to establish an objective measure with which to identify these individuals (Goldhaber & Hansen, 2010; National Commission on Teaching and America's Future, 1996).

However, within the literature, there is a substantial discussion surrounding the implicit political motivations for these tests: for example, "[t]he relevance-utility of the PLACE tests has less to do with the rational screening-out function it purports to have and much more to do with the love

affair and comfort the public seems to have with licensing tests, no matter how functional they are” (Cobb et al. 1999, p. 172).

Standardized Tests as a Method for Identifying Teachers of High Quality

The evidence for the effective use of standardized tests in order to identify teachers of high quality is mixed. There is some evidence that standardized tests can accurately measure content knowledge for some candidates (Hill, Sleep, Lewis & Ball, 2007). However, there is much less evidence that achievement on these tests directly correlates to the same individual becoming a quality teacher. Darling-Hammond (2000) sums the situation up by writing, “Tests measure the competence of teacher candidates by assessing their knowledge of best teaching practices and their knowledge of the skills necessary for being an effective teacher. Unfortunately, the tests are unable to evaluate the extent to which teacher candidates exhibit best practices in applied contexts” (as cited in Goodman, Arbona, & Dominguez de Rameriz, 2008, p. 24).

Overall, there is little support for a direct relationship between teachers’ performance on standardized tests and the eventual quality of the teaching instruction of the individual (Hill, Umland, Litke, & Kapitula, 2012). Most of the concerns about licensing examinations are related to methodological issues of construct definition and predictive value (Libman, 2009). Currently there is no record of the Praxis tests, the standardized test used in most U.S. states, being validated against real-life teacher performance (Hill et al., 2012).

There appears to be a common theme in much of the research that one of the limitations of standardized teacher competency testing is the challenge of a multiple-choice test capturing multidimensional and often hard-to-measure skills and attributes of skilled teachers. “Quality teachers simply cannot be identified exclusively by their licensure route or their exam scores. It seems that there is something unobservable about an individual that makes him or her an effective teacher” (Shuls & Trivitt, 2015, p. 665).

Standardized Tests as a Method for Identifying Teachers of Low Quality

At their most basic, standardized competency tests are designed to act as a minimum quality guard (Shuls and Trivitt, 2015). The evidence for whether these tests can be used as a tool for removing the least qualified teachers is not clear. A study on the ability of standardized teacher tests to identify capable teachers and identify individuals unsuited to teaching found that the tests “failed to identify all teachers with such significantly weak practice” (Hill et al., 2012, p. 514).

Scholars examining the results of standardized testing in the U.S. write: “How many incompetent potential teachers who would have received their teaching licenses because of inappropriate or shoddy grading practices in their higher education training programs are appropriately prevented from receiving a teaching license because of failing the PLACE tests? Although we have no direct data to answer that question, our hunch is ‘not many’” (Cobb et al., 1999, p. 172).

Subject-Specific Implementation of Standardized Testing

Currently, there are two leading types of standardized teacher competency tests in use in various jurisdictions: basic-skills tests and specialized-skills tests. The former type measures the fundamental knowledge that all teachers are expected to have in areas such as math, reading and writing. The specialized-skills tests are intended to measure specific content knowledge in areas such as mathematics, history and science (Cobb et al., 1999; Angrist & Guryan, 2008).

There are several different methods of incorporating subject matter into standardized teacher testing. Nearly all jurisdictions requiring competency testing include some components of a basic-skills assessment. Such skills include basic competence in mathematics, writing and reading. Virtually every jurisdiction that requires teacher competency testing also mandates a basic-skills test. Additionally, standardized tests in specific subject matter are a requirement in many regions. These are becoming increasingly common for teachers practicing at the later (e.g., secondary school) compared to earlier (e.g., elementary or middle school) years. Finally, some regions also require a standardized test of pedagogy knowledge, although this is the least common type of mandatory standardized testing (D'Agostino & Powers, 2009; Wayne & Youngs 2003; Shuls & Trivitt, 2015).

Timing of Teacher Testing

Most commonly, teachers are tested for these skills during their teacher training courses. Most individuals take these standardized competency tests before they become certified professional teachers (D'Agostino & Powers, 2009). However, many regions require a standardized competency test even before admission into teacher professional education programs (Libman, 2009; Wayne & Youngs, 2003).

Measurement of Competency in Standardized Tests

Very few jurisdictions that use compulsory teacher competency testing publicize the first-time pass rates. The few places that do have such a high level of variation in these data that it is challenging to create a clear picture of a valid pass rate. For example, state-level pass rates for the Praxis II math exam can range from 52% to 93%, with only seven states reporting (Putman & Walsh, 2019). The Australian government reports that pass rates for their standardized teacher tests are [above 90%](#) for first-time test takers. As a result, no precedent or indicators exist within current testing practices that would lead to a pass rate applicable to various locations.

In general, jurisdictions use these standardized competency tests to determine a pass/fail result and do not specify individual scores. The pass score varies even on the most widely used standardized tests in the U.S. (e.g., the Praxis tests) and is determined at the state level (Shuls, & Trivitt, 2015). Recommendations on these scores most commonly originate from a panel of experts, which combine information about item and examinee characteristics (Hill et al., 2012).

Researchers have studied the potential impact of policies that attempt to improve teacher quality by raising the pass score. The evidence shows that an increase in the score required to pass could increase the quality of teachers as measured by student outcomes on standardized tests. However, the change would almost certainly result in a reduction in the number of minority group teachers and lead to adverse outcomes in disadvantaged schools (Shuls & Trivitt, 2018).

The evidence also shows that increasing the score needed to pass the standardized teacher tests would remove 0.2% of the most ineffective teachers while excluding 7% of the effective teachers (Goldhaber, 2007).

Paths for Candidates Who Are Unsuccessful in Standardized Tests

In most jurisdictions in which there is mandatory testing, there is no regulatory limit on the number of times one person can try to take the test and pass (Libman, 2009).

Many regions that facilitate mandatory teacher testing, such as North Carolina, Florida and the United Kingdom, allow graduates of a licensed teacher training program who fail the exams to work for several years with a temporary teaching license (Goldhaber, 2007; Libman, 2009).

Competency Test Design and Delivery Process

The literature shows that while it is the government that implements the need for compulsory teacher competency testing, the design and administration of these tests originates from non-governmental bodies (Glassford, 2005; Libman, 2009; Wang et al., 2003). In the United States, most testing is delivered by the National Evaluation System, a branch of the for-profit corporation Pearson Education or by the Educational Testing Service, a private non-profit organization (Donovan & Cannon, 2018; Libman, 2009). In Australia, the [Australian Council for Educational Research](#) is an independent organization devoted to educational research that is responsible for facilitating a standardized test for pre-service teachers in testing centers via an online platform.

Current Test Formats and Included Items

Modern standardized teacher competency tests are primarily multiple-choice but also include a few short-answer modules. They are delivered either on paper or digitally. Tests are delivered in this format in order to achieve efficiencies in time and cost; and the format enhances the opportunity to assess specific knowledge across a broad domain (D'Agostino & Powers, 2009).

Common basic-skills tests contain about 50 to 55 questions in each area, usually math, reading and writing. Most tests also contain some short-answer questions. Several tests (e.g., the California Basic Educational Skills Test) contain a small number of essay questions in the reading and writing modules.

The multiple-choice format of almost all the standardized mathematical competency tests is a point of contention for many researchers. Heather C. Hill, a leading scholar in developing and understanding teacher mathematical competency measures, writes, "Our results also suggest that although teaching proficiency is multidimensional, that multidimensionality is unlikely to be captured by a single multiple-choice certification assessment" (Hill et al., 2012, p. 515).

Teacher Competency Testing and Student Outcomes

There is a large body of empirical evidence that demonstrates a positive relationship between teacher quality and student outcomes (Aaronson et al., 2007; Coleman, 1968; Rivkin et al., 2005; Rockoff, 2004). However, the evidence concerning how to measure teacher quality and how to identify any causal relationships between teacher quality and student achievement is much less conclusive. For instance, several studies have found that teacher credentials (e.g., degree level, licensure status) do not predict student achievement in statistically significant ways (e.g., Aaronson et al., 2007; Goldhaber & Brewer, 2000; Harris & Sass, 2011). Scholars have also not identified a relationship between degree type (e.g., Aaronson et al., 2007; Monk & King, 1994) or college entrance exams (e.g., Kane & Staiger, 2008) and student achievement in mathematics. The literature review finds mixed evidence concerning the association between standardized teacher competency testing and student achievement.

Challenges in Measuring the Relationship Between Teacher Competency and Student Outcomes

Assessing the relationship between standardized teacher competency testing and student achievement is complex. One challenge is that there are very few points of comparison. Researchers cannot compare teachers who pass the competency tests with those who do not pass, since they do not remain in the education system (Shuls & Trivitt, 2015).

Additionally, there is no research that combines longitudinal data from before and after the implementation of mandatory teacher testing. Therefore, there are no tests of whether the existence of standardized competency tests impacts the quality of the workforce (Goldhaber, Gratz et al., 2017).

Also, there is a correlation between other teacher characteristics and teacher performance on standardized competency tests. Teacher characteristics such as grades in teacher post-secondary education, number of math classes taken and licensing pathway have all been shown to impact student achievement and are correlated with teacher performance on standardized competency tests (Seebuck, 2015). It is not impossible to disentangle these effects, but it requires nuanced methodology and high-quality data (D'Agostino & Powers 2009; Seebuck, 2015).

Another challenge arises from the fact that students are not enrolled in classes randomly assigned to teachers and vice versa. This reality makes it likely that a relationship already exists between teachers who perform well on competency tests and the types of students they tend to teach and the types of classrooms they tend to teach in (Chetty, Friedman, & Rockoff, 2014; Kane & Staiger, 2008; Jackson, Rockoff, & Staiger, 2014).

Evidence Linking Teacher Competency Testing and Student Outcomes

When studying evidence linking teacher competency testing to student outcomes, a careful distinction must be made among teacher certification, teachers' pure content knowledge, teachers' pedagogical content knowledge and teacher testing. As discussed above, compulsory standardized teacher testing is one form of monitoring teacher content knowledge. There is not a perfect relationship between a teacher's actual content knowledge and their results on a standardized competency test (Goodman, Arbona & Dominguez de Ramirez, 2008). Further,

there is even less of a relationship between a teacher's pedagogical content knowledge and that teacher's results on a standardized competency test (Hill, 2005; Hill, 2007. Hill et al., 2008; Kelcey et al., 2019).

While a substantial body of evidence finds that teachers' pure content knowledge has a favourable relationship with student outcomes (Campbell et al., 2014; Carlisle & Berebitsky, 2011) and that this relationship is mediated by context (Kelcey, Hill, & Chin, 2019); the evidence around the association between standardized competency testing and student outcomes is not conclusive (Podgursky, 2005).

Such mixed conclusions from empirical research, which illustrate that standardized teacher competency testing can be statistically significant or insignificant in determining student outcomes is not surprising for several reasons: first, due the methodological challenges explained above; and, second, that such variation in the significance of this relationship can be explained by the individual use of these tests as pass/fail for completion of a course. Further, these tests were not intended to act as nuanced predictors of performance in the classroom but rather as "a minimum quality guard" (Shuls & Trivitt, 2015, p. 664). While several studies find no relationship between teacher performance on standardized competency tests and student achievement (Buddin & Zamarro, 2009; Quirk, Witten, & Weinberg, 1973; Wiseman & Al-bakr, 2013), the majority of empirical evidence suggests that standardized teacher competency tests provide some signal of teacher quality, yet this association is weak, with small effect sizes, and is also inconsistent across contexts (D'Agostino & Powers, 2009; Goldhaber, 2007; Kelcey et al., 2019).

Praxis II, the most commonly used U.S test for teacher competence, has been studied for its predictive relationship to student outcomes in several settings. Shuls and Trivitt (2015) used the Praxis data from the Arkansas Department of Education and student outcomes on test scores from pupils in Grades 3–8. This data set did not allow them to match students to teachers, so the results are based instead on school–grade level correlations. Their analysis identifies a statistically significant correlation between teacher and student test scores. They also report a significant correlation between teacher certification pathways and experience on the one hand and student test scores on the other. The teacher certification effect size is approximately double the other two effect sizes (Shuls & Trivitt, 2015).

Researchers in Colorado focused on the basic-skills test that is required as part of the state's licensing process and found that it "seemed overwhelmingly burdensome and lacked relevance and utility, except in a few cases" (Cobb et al., 1999, p. 173).

A large study in Texas evaluated the relationship between teachers' mathematics knowledge, classroom decision making and student achievement and found that the mathematical knowledge of teachers was only very weakly related to student achievement. Further, it found that the relationship, where it did exist, was nonlinear and strongly influenced by other instructional factors such as the curriculum, the technology available and the school environment (Shechtman et al., 2010). The standardized test used to evaluate teachers in the Texas study was the MKT instrument designed by Hill and team (Hill et al., 2005; Hill et al., 2012).

A review of the international evidence regarding the association of teacher certification and student achievement found that requiring standardized testing for certification did not predict student achievement, and by association, teacher quality (Wiseman & Al-bakr, 2013).

Additional international evidence originates from a study spanning 17 countries by Ingvarson and Rowley (2017). This research explores the relationship between national-level teacher quality assurance processes and student outcomes on international mathematics exams (PISA and TIMSS). Although the study did not consider the effect of standardized testing in isolation, it did include teacher competency testing as part of the assessment. The researchers rated each country on the strength of their teacher recruitment policies, teacher accreditation and education policies, and thirdly, teacher certification policies. Standardized teacher competency testing falls into the third category. Student results on both the mathematics exams were most strongly predicted by the second category a country's strength in maintaining quality in teacher education and accreditation (.75, $p < .01$). Student results were also significantly related to the teacher certification policies for the PISA tests only (0.52, $p < .05$) (Ingvarson & Rowley, 2017). The difference between the PISA and TIMSS test scores is summarized by Wu (2010): "a country with a high score in PISA shows that the students are good at 'everyday mathematics,' while a high score in TIMSS shows that the students are good at 'school mathematics'" (p. 96).

D'Agostino and Powers (2009) conducted an extensive meta-analysis examining the ability of teacher test scores to predict teacher performance and student outcomes and found that "test scores have been less related to teaching performance than students' success level in the preservice programs the tests were designed to hold accountable" (p. 165). In their large cross-jurisdictional analyses, the authors compared the impact on student performance of teacher performance in teacher training college and performance on teacher competency tests. D'Agostino and Powers (2009) found that teacher performance in teaching training or college was more strongly correlated with students' outcomes than the teacher's performance on standardized teacher competency tests taken upon graduation. However, the standardized teacher competency test results did have a weak positive correlation with student outcomes. The effect size of standardized tests was about half the effect size of teacher performance in college courses (D'Agostino & Powers, 2009).

Studies that have compared the relative impact of teacher characteristics, including results on standardized testing, sometimes show a positive correlation with student achievement. However, these studies are never the most predictive. For example, Begle (1979) reported that credit accumulation in mathematics methods courses was correlated with student performance. Similarly, Ferguson and Womack (1993) suggested that the amount of coursework completed by teachers accounted for more than four times the variance in teacher performance (16.5%) than any measures of content knowledge (less than 4%).

A study using North Carolina's data shows that standardized teacher testing is predictive of teacher effectiveness measured by student performance on standardized testing. These researchers conclude that teachers passing the Praxis II test increases student achievement by 3–6% of a standard deviation higher in math compared to teachers not passing the test (Goldhaber, 2007).

Clotfelter, Ladd and Vigdor (2010) also used the North Carolina Praxis data to evaluate drivers of student achievement and identify a positive relationship between teacher competency test results and students' achievement on standardized math tests. The Clotfelter (2010) study explored multiple teacher characteristics and found that teacher competency test scores were the least predictive of student outcomes, with an effect size of 0.01. The teachers' years of experience had an effect size of 0.05; the license type had an effect size of 0.06; and the teacher possessing the certification to teach math had an effect size of 0.12.

There is some evidence that the relative importance of teachers' content knowledge and the associated standardized teacher testing may increase as the complexity of the subject matter increases in higher grades (Appleton, 2013). A study using data from Florida's ABCTE testing finds that teachers who pass both the professional knowledge test and the subject area test are associated with higher student outcomes and that the relationship is strongest when looking only at student achievement in Grades 6–10 (Goldhaber, Gratz et al., 2017; Sass, Semykina, & Harris, 2014).

Researchers used the Washington State data and looked at the relationship between teacher content knowledge test scores and middle school student math achievement and Grade 9 student algebra achievement. They found that there is a moderately positive relationship between teacher math competency test scores and student performance in middle school math, and no relationship for Grade 9 algebra (Goldhaber et al., 2017).

This all said, the empirical evidence does suggest that multiple-choice standardized teacher competency tests can be weakly to moderately predictive of classroom practice and student performance, particularly in middle or high school (Hill et al., 2012). However, standardized test scores are much less related to student outcomes than are teacher certification (both general and subject-specific), teacher experience and other contributors to teacher effectiveness within the classroom and at district levels.

Associated Impacts of Teacher Testing

Many of the studies testing the association between teacher testing and student outcomes discuss the impacts of teacher testing outside student achievement on standardized tests. Cobb et al. (1999) point out that since such a small proportion of potential teachers fail the basic-skills test, “an enormously large proportion of students must take the tests with little discriminatory power associated with that time and effort” (p. 173). Cobb et al. also suggest that test content is, for the most part, equivalent to the tests that the students have already undertaken in order to be admitted to higher education.

There is also evidence that mandatory teacher competency testing impacts the teacher workforce pipeline, workforce, teacher training and the diversity of the teacher population. “The transaction costs may prevent competent individuals from ever considering a career in teaching” (Shuls & Trivitt, 2015, p. 647).

Results from the Ingersoll (2007) study suggested that single filters, such as standardized tests, could be perceived as barriers to competence and professionalism unless accompanied by consideration of other factors such as remuneration and working conditions.

Furthermore, evidence suggests that some of the standardized teacher competency tests may be dissuading individuals with potential, rather than objectively unqualified teachers, from entering the pipeline. Tests that claim superiority in selecting teachers who have graduated from certification programs are estimated to have as much as “an 80% false-rejection rate and a mere 20% false-positive rate” (D’Agostino & Powers, 2009, p. 151). Empirical evidence shows that state regulators may not be in the best position to assess quality teaching. Podgursky (2005) suggests that teaching quality may best be considered at the local level by school administrators, reporting that school administrators can identify teachers whose methods support gains in student outcomes.

The Influence of Standardized Tests on Diversity and Inclusion

Related to the restricted pipeline resulting from standardized teacher testing is the serious impact that these assessments have on racial diversity within the teacher pool. There is mounting evidence that mandatory teacher competency testing has a limiting effect on diversity and inclusion in the teacher workforce. Angrist and Guryan (2008) illustrate that an analysis of the intersection between race and standardized teaching tests shows that the mandated assessments prove to take the role of a biased barrier rather than a screen for quality. Further research shows that a “high stakes” system of standardized examinations prejudices weaker social groups or minorities (Libman, 2009, p. 11). Multiple studies have found that the pass rate on common standardized teacher competency tests are significantly lower for people of colour than for white applicants (Cobb et al. 1999; Graham, 2013; Nettles et al., 2011); Petchauer, 2012). Additionally, there is evidence that such standardized tests are biased against almost all vulnerable classes of potential teachers other than women (Angrist & Guryan, 2008). Empirical evidence demonstrates that black teacher candidates (Petchauer, 2012) and Hispanic teacher candidates (Cobb et al., 1999) are much more likely to fail the standardized teacher tests than their white peers.

Researchers offer concepts such as social capital theory, stereotype threat or assessment bias as potential reasons racialized populations are disproportionately failing standardized teaching tests, such as Praxis (Graham, 2013). While no definitive conclusions exist concerning which factor is the most influential, research continually indicates that these tests negatively impact prospective pre-service teachers. Graham (2013) highlights research from Dr. Dara Wakefield, whose work has led to the conclusion the Praxis test is “more than any other criterion is single-handedly reducing the number of minorities who enter the teaching profession” (see Wakefield, D. (2003). Screening teacher candidates: Problems with high-stakes testing. *The Educational Forum*, 67, 380-388).

According to several studies, in addition to reducing the pool of diverse applicants, standardized competency testing has the potential to impact the distribution of qualified teachers (Podgursky, 2005; Seebruck, 2015). The unbalanced distribution of quality teachers has the unintended consequence of lowering average teacher quality in vulnerable, high-poverty or rural districts (Podgursky, 2005; Seebruck, 2015).

Teacher Characteristics and Evaluation Procedures That Predict Student Outcomes

The literature on the relationship between standardized teacher testing and student outcomes also includes some findings on the characteristics of teachers and teacher evaluation procedures that influence student outcomes. No single characteristic is consistently reported as the “best” quality of a teacher. Commonly, teacher experience and high-quality teacher education were identified as positive predictors of student success (Harris & Sass, 2011). Yet, even here, the results are mixed (Steinberg & Sartain, 2015).

Research from the province of Quebec, where standardized student math test scores are high relative to the rest of Canada, attributes such student achievement to post-secondary mathematics education programs offered in the province. This research finds a uniquely strong emphasis on requiring trainee teachers to undertake more courses in both mathematics methodology and mathematics content (Shulman 1987; Vashchyshyn & Chernoff, 2016).

Currently, most regions, particularly in North America, do not require elementary teachers to complete a large number of hours dedicated to mathematics while in pre-service education, nor do these jurisdictions have rigorous question sets in their licensure exams to determine mathematics competency (Georges, Borman, & Lee, 2010). Both factors have been strongly linked to increases in student performance in Quebec and elsewhere (Begle, 1979; Borman & Lee, 2010).

Several of the studies evaluating teacher effectiveness report that teacher experience is an important predictor of student outcome (Goldhaber, 2007; Huang & Moon, 2009). The Huang study, a multilevel analysis of teacher characteristics and student achievement, found that “students taught by seasoned grade level teachers for 4 years in a row may wind up scoring more than one standard deviation higher or approximately one grade level more than students taught by beginning teachers” (Huang & Moon, 2009, p. 226).

Other researchers found strong predictive power in the potential teachers’ performance in their college or university teacher training, especially during student teaching (D’Agostino & Powers, 2009; Marder, 2016). One study focusing specifically on mathematics found that increasing the quality and quantity of required mathematics courses at the pre-service level was one of the most helpful steps toward improving student outcomes (Georges, Borman, & Lee, 2010).

The body of research concerned with evaluating the method of teacher evaluation most predictive of positive student outcomes is consistent in its findings. This evidence recommends focusing on process-oriented and pedagogy-oriented evaluation measures rather than on semi-immutable characteristics such as prior experience, education or cognitive ability (Goe, Bell, & Little, 2008). As written by Clotfelter et al. (2010), “[t]he discrepancy between the overall variation in teacher quality and that predicted by credentials alone implies that it would be a mistake for policymakers to put so much weight on measurable credentials in determining teacher quality that they ignore other contributors to teacher effectiveness, many of which only can be determined by observation at the school or classroom level” (p. 676).

The Briole and Maurin (2019) study found that teacher evaluation makes a significant difference to student outcomes and that the most effective is an in-person evaluation conducted by a senior professional consisting of one class observation and a debriefing session. This study’s findings indicate a significant and large (of about 4.5% of a standard deviation) increase in student math scores. “This effect is seen not only for students taught by the teacher the year of the evaluation but also for students taught by the same teacher the subsequent years, suggesting that evaluations improve teachers’ core pedagogical skills” (Briole & Maurin, 2019, p. 15).

A similar study estimating the impact of a newly launched teacher evaluation system in Cincinnati in 2000 found that brief in-person evaluations of teachers that included a combination of content and pedagogical feedback significantly increased student performance on tests in a lasting manner (Taylor & Tyler, 2012).

Conclusion

Teacher quality is identified as influential to student achievement (Aronson et al., 2007; Coleman, 1968; Rivkin et al., 2005; Rockoff, 2004). Yet individual variables (e.g., degree level or licensure status) appear to detect and predict significant changes in student outcomes (Aronson et al., 2007; Goldhaber & Brewer, 2000; Harris & Sass, 2011).

Overall, the standardized teacher competency tests, which have been propagated and instituted across the United States and in several locations around the world, are intrinsically linked to public perception of teaching robustness and quality. These standardized tests originated during a movement that called for increased accountability in the quality and effectiveness of local teachers and teacher training programs. These practices are linked with some research that creates a pathway enabling potential improvements in the quality and performance of teachers within the classroom. However, while these tests may receive support from policymakers, educators and the public, the fundamental goal of these tests—to improve student learning—is often not met. Current research demonstrates that standardized teacher tests is not linked with a level of performance consistency that justifies their widespread implementation at this time. The use of caution with these tests is advised by many researchers on the basis that these tests are not consistently associated with the positive benefits that are often claimed. Furthermore, the potential negative impacts of these programs, including bias against marginalized groups and the decrease in the availability of qualified teachers, are more consistent impacts of these tests.

Without identifying the best methods to increase student success, creating a restrictive pathway to becoming a teacher may cause more harm than good. Many researchers in this field, along with the American Psychological Association, advocate against the use of standardized tests as the single metrics for aptitude or achievement (Goodman et al., 2008). As outlined by this review, the positive qualities and merit of teachers go beyond that which can be measured solely by a standardized multiple-choice test. The research shows that creating a system devoted solely to standardized tests, in the absence of other assessment, may lead to adverse impacts on both the teacher pool and educational outcomes for students. Until standardized teacher competency tests become associated with consistent, measurable benefits, they should be used with caution and with consideration for the potential impact on local communities.

The “discrepancy between the variation in teacher quality and [the teacher quality] predicted by credentials alone implies that it would be a mistake for policymakers to put so much weight on measurable credentials in determining teacher quality” (Clotfelter, Ladd, & Vigdor, 2010, p. 676).

Furthermore, other crucial factors essential for teacher effectiveness are qualitative in nature and cannot be measured by standardized testing methods; instead they require observation at the local level (Clotfelter, Ladd, & Vigdor, 2010).

References

- 2018 Teacher Prep Review. (n.d.). Retrieved July 28, 2019, from National Council on Teacher Quality (NCTQ) website: <https://www.nctq.org/publications/2018-Teacher-Prep-Review>
- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135. <https://doi.org/10.1086/508733>
- Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, 27(5), 483–503. <https://doi.org/10.1016/j.econedurev.2007.03.002>
- Appleton, K. (2013). *Elementary Science Teacher Education: International Perspectives on Contemporary Issues and Practice*. <https://doi.org/10.4324/9781315045443>
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special. *Journal of Teacher Education*, 59(5), 389–407.
- Ball, Hill, H. C., & Bass. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, 29(1), 14–46.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. & Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. <https://doi.org/10.3102/0002831209345157>
- Begle, E. G. (1979). *Critical Variables in Mathematics Education: Findings from a Survey of the Empirical Literature*.
- Blömeke, S., & Delaney, S. (2014). Assessment of teacher knowledge across countries: A review of the state of research. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & W. H. Schmidt (Eds.), *International Perspectives on Teacher Knowledge, Beliefs and Opportunities to Learn: TEDS-M Results* (pp. 541–585). https://doi.org/10.1007/978-94-007-6437-8_25
- Briole, S., & Maurin, E. (2019). *Does Evaluating Teachers Make a Difference?* (SSRN Scholarly Paper No. ID 3390297). Retrieved from Social Science Research Network website: <https://papers.ssrn.com/abstract=3390297>
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115. <https://doi.org/10.1016/j.jue.2009.05.001>
- Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). *Teaching for Excellence and Equity: Analyzing Teacher Characteristics, Behaviors and Student Outcomes with TIMSS*. <https://doi.org/10.1007/978-3-030-16151-4>
- Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., Neumayer DePiper, J., Frank, T. J., Griffin, M. J & Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, 45(4), 419–459.
- Carlisle, J. F., & Berebitsky, D. (2011). Literacy coaching as a component of professional development. *Reading and Writing*, 24(7), 773–800. <https://doi.org/10.1007/s11145-009-9224-4>

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499–531.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679. <https://doi.org/10.1257/aer.104.9.2633>

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(3), 655–681. <https://doi.org/10.3368/jhr.45.3.655>

Cobb, R. B., Shaw, R., Millard, M., & Bomotti, S. (1999). An examination of Colorado's teacher licensure testing. *The Journal of Educational Research*, 92(3), 161–175. <https://doi.org/10.1080/00220679909597592>

Coleman, J. S. (1968). Equality of educational opportunity. *Equity & Excellence in Education*, 6(5), 19–28. <https://doi.org/10.1080/0020486680060504>

Corey, D. L., Phelps, G., Ball, D. L., Demonte, J., & Harrison, D. (2012). Explaining variation in instructional time: An application of quantile regression. *Educational Evaluation and Policy Analysis*, 34(2), 146–163. <https://doi.org/10.3102/0162373711431098>

Council, N. R., Education, D. of B. and S. S. and, Education, C. for, & States, C. on the S. of T. P. P. in the U. (2010). *Preparing Teachers: Building Evidence for Sound Policy*. National Academies Press.

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(1), 146–182. <https://doi.org/10.3102/0002831208323280>

Darling-Hammond, L. (2010). *Evaluating Teacher Effectiveness: How Teacher Performance Assessments Can Measure and Improve Teaching*. Retrieved from <https://eric.ed.gov/?id=ED535859>

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57–77. <https://doi.org/10.3102/01623737023001057>

DeForest, J. (2006). New York City's failed teacher selection project: Political reality trumps educational research, 1947-1953. *Teachers College Record*, 108(4), 726–747. <https://doi.org/10.1111/j.1467-9620.2006.00666.x>

Donovan, M. K., & Cannon, S. O. (2018). The university supervisor, edTPA, and the new making of the teacher. *Education Policy Analysis Archives*, 26, 28. <https://doi.org/10.14507/epaa.26.2849>

Ferguson, P., & T. Womack, S. (1993). The impact of subject matter and on teaching performance. *Journal of Teacher Education*, 44(1), 55–63. <https://doi.org/10.1177/0022487193044001008>

Georges, A., Borman, K. M., & Lee, R. S. (2010). Mathematics reform and teacher quality in elementary grades: Assessments, teacher licensure, and certification. *Education Policy Analysis Archives*, 18, 13. <https://doi.org/10.14507/epaa.v18n13.2010>

- Gibb, Nick Hn. (2019) Changes to the professional skills test for teachers, *UK Government, Department of Education*. <https://www.gov.uk/government/news/changes-to-the-professional-skills-test-for-teachers>
- Glassford, L. A. (2005). A triumph of politics over pedagogy? The case of the Ontario Teacher Qualifying Test, 2000-2005. *Canadian Journal of Educational Administration and Policy*.
- Glassford, L. A. (2007). Quality assurance or election fodder? The politics of teacher testing in Ontario, 1999 - 2003. *Historical Studies in Education / Revue d'histoire de l'éducation*, 25–52.
- Goe, L. (2007). *The Link between Teacher Quality and Student Outcomes: A Research Synthesis*. Retrieved from <https://eric.ed.gov/?id=ED521219>
- Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. *National Comprehensive Center for Teacher Quality*.
- Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4), 765–794.
- Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1), 134–150.
- Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education*, 68(4), 377–393. <https://doi.org/10.1177/0022487117702582>
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–145. <https://doi.org/10.3102/01623737022002129>
- 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. <https://doi.org/10.1016/j.econedurev.2017.09.002>
- Goldhaber, D., & Hansen, M. (2010). Assessing the potential of using value-added estimates of teacher job performance for making tenure decisions. Working Paper 31. *National Center for Analysis of Longitudinal Data in Education Research*.
- Goodman, G., Arbona, C., & Dominguez de Rameriz, R. (2008). High-stakes, minimum-competency exams: How competent are they for evaluating teacher competence? *Journal of Teacher Education*, 59(1), 24–39. <https://doi.org/10.1177/0022487107309972>
- Graham, A. (2013). Black teacher education candidates' performance on PRAXIS I: What test results do not tell us. *Negro Educational Review*, 64(1/4), 9.
- Haney, W., Madaus, G., & Kreitzer, A. (1987). Chapter 5: Charms talismanic: Testing teachers for the improvement of American education. *Review of Research in Education*, 14(1), 169–238.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7), 798–812. <https://doi.org/10.1016/j.jpubeco.2010.11.009>
- Hill, H., & Ball, D. L. (2009). The curious—and crucial—case of mathematical knowledge for teaching. *Phi Delta Kappan*, 91(2), 68–71. <https://doi.org/10.1177/003172170909100215>

- Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, 29(2), 95–114. <https://doi.org/10.3102/0162373707301711>
- Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, 41(5), 513–545.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, 26(4), 430–511. <https://doi.org/10.1080/07370000802177235>
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. <https://doi.org/10.3102/0002831210387916>
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406. <https://doi.org/10.3102/00028312042002371>
- Hill, H. C., Sleep, L., Lewis J. M., & Ball, D. L., (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts? In Lester, F. K., (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (p. 111-155). Charlotte, N.C: Information Age Pub. Retrieved from http://hub.mspnet.org/media/data/Hill_etal_final.pdf?media_000000005636.pdf
- Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education*, 118(4), 489–519. <https://doi.org/10.1086/666380>
- Huang, F. L., & Moon, T. R. (2009). Is experience the best teacher? A multilevel analysis of teacher characteristics and student achievement in low performing schools. *Educational Assessment, Evaluation and Accountability*, 21(3), 209–234. <https://doi.org/10.1007/s11092-009-9074-2>
- Ingersoll, R. M. (2007). A comparative study of teacher preparation and qualifications in six nations. CPRE Research Report Series. *Consortium for Policy Research in Education*.
- Ingvanson, L., & Rowley, G. (2017). Quality assurance in teacher education and outcomes: A study of 17 countries. *Educational Researcher*, 46(4), 177–193. <https://doi.org/10.3102/0013189X17711900>
- Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, 6(1), 801–825. <https://doi.org/10.1146/annurev-economics-080213-040845>
- Kane, T. J., & Staiger, D. O. (2008). *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation* (Working Paper No. 14607). <https://doi.org/10.3386/w14607>
- Kelcey, B., & Carlisle, J. F. (2013). Learning about teachers' literacy instruction from classroom observations. *Reading Research Quarterly*, 48(3), 301–317. <https://doi.org/10.3102/0002831213492844>
- Kelcey, B., Hill, H. C., & Chin, M. J. (2019). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel quantile mediation analysis. *School Effectiveness and School Improvement*, 0(0), 1–34. <https://doi.org/10.1080/09243453.2019.1570944>

- Kersting, N., Givvin, K., Thompson, B., Santagata, R., & Stigler, J. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568–589. <https://doi.org/10.3102/0002831212437853>
- Libman, Z. (2009). Teacher licensing examinations—True progress or an illusion? *Studies in Educational Evaluation*, 35(1), 7–15. <https://doi.org/10.1016/j.stueduc.2009.01.003>
- McNeal, K., & Lawrence, S. (2009). Teachers from the “Neighborhood”: Standardized testing as a barrier to certification of minority candidates. *Online Yearbook of Urban Learning, Teaching, and Research*. Retrieved from <https://eric.ed.gov/?id=EJ859480>
- Meroni, E. C., Vera-Toscano, E., & Costa, P. (2015). Can low skill teachers make good students? Empirical evidence from PIAAC and PISA. *Journal of Policy Modeling*, 37(2), 308–323. <https://doi.org/10.1016/j.jpolmod.2015.02.006>
- Monk, D. H., & King, J. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. *Choices and Consequences: Contemporary Policy Issues in Education*, 29–58.
- Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review*, 40(2), 139–157. <https://doi.org/10.1086/447369>
- Nettles, M. T., Scatton, L. H., Steinberg, J. H., & Tyler, L. L. (2011). Performance and passing rate differences of African American and white prospective teachers on Praxis™ examinations: A joint project of the National Education Association (NEA) and Educational Testing Service (ETS). *ETS Research Report Series*, 2011(1), i–82.
- Petchauer, E. (2012). Teacher licensure exams and Black teacher candidates: Toward new theory and promising practice. *The Journal of Negro Education*, 81(3), 252–267. <https://doi.org/10.7709/jnegroeducation.81.3.0252>
- Petscher, Y., & Logan, J. A. R. (2014). Quantile regression in the study of developmental sciences. *Child Development*, 85(3), 861–881. <https://doi.org/10.1111/cdev.12190>
- Podgursky, M. (2005). Teacher licensing in U.S. public schools: The case for simplicity and flexibility. *Peabody Journal of Education*, 80(3), 15–43. https://doi.org/10.1207/s15327930pje8003_3
- Pullin, D. (2017). What counts? Who is counting? Teacher education improvement and accountability in a data-driven era. In J. Nuttall, A. Kostogriz, M. Jones, & J. Martin (Eds.), *Teacher Education Policy and Practice: Evidence of Impact, Impact of Evidence* (pp. 3–16). https://doi.org/10.1007/978-981-10-4133-4_1
- Putman, H., & Walsh, K. (2019). *A Fair Chance: Simple Steps to Strengthen and Diversify the Teacher Workforce*. Retrieved from <https://eric.ed.gov/?id=ED593388>
- Quirk, T. J., Witten, B. J., & Weinberg, S. F. (1973). Review of studies of the concurrent and predictive validity of the National Teacher Examinations. *Review of Educational Research*, 43(1), 89–113.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>

- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252.
<https://doi.org/10.1257/0002828041302244>
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2010). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1), 43–74.
https://doi.org/10.1162/EDFP_a_00022
- Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, 8(3), 255–281.
- Santibanez, L. (2006). Why we should care if teachers get A's: Teacher test scores and student achievement in Mexico. *Economics of Education Review*, 25(5), 510–520.
<https://doi.org/10.1016/j.econedurev.2005.08.001>
- Sass, T. R., Semykina, A., & Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38, 9–23.
- Seebruck. (n.d.). *Teacher Quality and Student Achievement: A Multilevel Analysis of Teacher Credentialization and Student Test Scores in California High Schools*. Retrieved from https://www.mcgill.ca/msr/files/msr/seebruck_2015.pdf
- Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge, to classroom practice, to student learning in the instructional system of the middle-school mathematics classroom. *Cognition and Instruction*, 28(3), 317–359.
<https://doi.org/10.1080/07370008.2010.487961>
- Shuls, J. V. (2018). Raising the bar on teacher quality: Assessing the impact of increasing licensure exam cut-scores. *Educational Policy*, 32(7), 969–992.
<https://doi.org/10.1177/0895904816682315>
- Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy*, 29(4), 645–675. <https://doi.org/10.1177/0895904813510777>
- Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy*, 10(4), 535–572. https://doi.org/10.1162/EDFP_a_00173
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628–3651. <https://doi.org/10.1257/aer.102.7.3628>
- Thibodeaux, A. K., Labat, M. B., Lee, D. E., & Labat, C. A. (2015). The effects of leadership and high-stakes testing on teacher retention. *Academy of Educational Leadership Journal; Arden*, 19(1), 227–249.
- Tighe, E. L., & Schatschneider, C. (2016). A quantile regression approach to understanding the relations among morphological awareness, vocabulary, and reading comprehension in adult basic education students. *Journal of Learning Disabilities*, 49(4), 424–436.
<https://doi.org/10.1177/0022219414556771>
- Vashchyshyn, I. I., & Chernoff, E. J. (2016). A formula for success? An examination of factors contributing to Quebec students' strong achievement in mathematics. *Canadian Journal of Education / Revue Canadienne de l'éducation*, 39(1), 1–26. Retrieved from JSTOR.
- Wakefield, D. (2003). Screening teacher candidates: Problems with high-stakes testing. *The Educational Forum*, 67(4).

Wang, J., & Hartley, K. (2003). Video technology as a support for teacher education reform. *Journal of Technology and Teacher Education*, 11(1), 105–138.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89–122.
<https://doi.org/10.3102/00346543073001089>

Wiseman, A. W., & Al-bakr, F. (2013). The elusiveness of teacher quality: A comparative analysis of teacher certification and student achievement in Gulf Cooperation Council (GCC) countries. *PROSPECTS*, 43(3), 289–309. <https://doi.org/10.1007/s11125-013-9272-z>

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach For America in high school. *Journal of Policy Analysis and Management*, 30(3), 447–469.
<https://doi.org/10.1002/pam.20585>