APPENDIX B

Literature review on licensure test predictive validity

Debates around licensure tests frequently raise a central question: Does a teacher's score on (or passing versus not passing) an elementary content licensing test predict her effectiveness in the classroom, as measured by student outcomes? NCTQ endeavored to identify and review all the existing literature on this topic to attempt to answer this question.

No licensure test measures everything that constitutes an effective teacher. This dataset focuses on measures of the knowledge that teachers will be expected to teach. Teachers' content knowledge is a crucial part of effective teaching, but it does not comprise all that teachers need to know or be able to do. Because content knowledge is not the only factor that matters, it would be surprising to find a 1:1 correlation between performance on licensure tests and teachers' effectiveness in the classroom, or even something near that level of correlation. On the other hand, a negative correlation (or no correlation) should rightfully raise questions about the utility of licensure tests, suggesting that they are more an obstacle than a guardrail.

Measuring the predictiveness of licensing tests is made harder by the fact that most prospective teachers who fail the test generally do not qualify for a teaching license. As of 2020, only eight states allow elementary teachers to become fully certified without passing a content licensure test.¹ Because relatively few people become the teacher of record without having passed licensure tests, it is difficult to evaluate the effectiveness of teachers who never passed their licensure tests. Some research tries to navigate this limitation by considering whether teachers passed their test on the first attempt, or by taking advantage of state policy changes (e.g., adding a test requirement or changing the minimum passing score) to conduct a natural experiment.

Some states, however, give teachers one to three years to pass their tests, making these states prime candidates for such a study. Currently, little research exists that evaluates the effectiveness of teachers who could not pass their licensure tests after multiple attempts, because most of these aspiring teachers simply do not reach the classroom. Those who do are exceptions to the rule and are not likely representative of the broader pool of teachers who failed.

Process

A search for key terms² on Google Scholar, as well as on ETS's Research page, identified several thousand potentially relevant studies conducted between 1990 and 2020. A scan of the abstracts revealed considerable duplication, concluding that 269 studies were related closely enough to the research question to warrant a full review. Of those, only 15 were directly relevant to the question of whether licensure tests predict teacher effectiveness as measured by student outcomes.

Some of the eliminated studies either examined other teacher characteristics while mentioning licensing test scores' predictive validity in a literature review, or were simply reviews of existing research on teacher characteristics related to teacher effectiveness, including teacher licensing test scores. This review is limited to studies that directly examine the variable of teacher licensing test scores, independently or in conjunction with other measures of teacher characteristics or as part of a broader index of teachers' academic ability.³ We also included Hanushek (1997), a meta-analysis of various school-level inputs and their impact on student performance, which includes a statistical aggregation of 41 studies that examined teacher test scores (not necessarily specific to tests of content knowledge) or IQ scores on various measures of students' academic achievement.

Results

Of the 15 studies reviewed (see table below), 11 found a statistically significant positive relationship between teachers' test scores (or their ability to pass the content licensure test) and future effectiveness in the classroom, generally measured by students' gains on standardized tests in English language arts and math.⁴

Most of these 11 studies showed greater predictive validity for math performance than for English language arts. In general, findings were modest but statistically significant. For example, Chingos and Peterson (2011) found that failing Florida's licensure test at least once was associated with significantly lower student test scores (.036 standard deviations lower in fourth through fifth grade student math scores, .028 standard deviations lower in sixth through eighth grade math scores, and .03 standard deviations lower in sixth through eighth grade reading). These negative correlations were slightly higher than the positive correlations for teachers having National Board certification or for having attended a more selective college.

Of the research finding an equivalent or stronger correlation between teachers' licensure test scores and students' reading or ELA achievement, one study (Araujo et al., 2020) was based on data from Ecuador, which employs a very different teacher screening system and evaluates teachers passing the test in conjunction with their tenure status. More locally, Ferguson (1991) found Texas teachers' scores on a relicensing test covering basic literacy skills to be the most important school-level predictor of math and reading standardized test outcomes for their students.

Most recently, a study of the MTEL licensure tests in Massachusetts found that the tests were predictive of teachers' value-added scores as well as their evaluation ratings. Moreover, the MTEL scores were more strongly predictive of teacher performance ratings for teachers of color than they were for White teachers. The MTEL tests include both a basic skills test and a content knowledge test specific to the candidate's certification area.

Several of the 11 studies offer positive findings but warn of potential problems when the cut score is set too high, excluding higher numbers of effective would-be teachers. Goldhaber's 2007 study in North Carolina found that the state's "Curriculum" test (actually a test of content) better predicted teacher effectiveness than did the "Content test" (actually a test of

test takers' responses to teaching situations) with teachers in the top quintile of Curriculum test scores achieving between 2.2–3.5% of a standard deviation more in student reading and math achievement.8 But Goldhaber identified limitations to the policy significance of his study's findings, noting that increasing the test score cutoff in North Carolina to the higher standard used by Connecticut "would lead to the exclusion of less than 0.5 percent of the teacher work force estimated to be very ineffective teachers, but would also result in the exclusion of 7 percent of the teacher workforce estimated to be effective teachers" (p. 765). Similarly, based on teacher licensure test data from Arkansas, Shuls (2018) found a positive relationship for math but none for English language arts. Shuls also found that a higher cut score (by 0.25 standard deviations) would be a more effective screen for teacher quality in both math and ELA but would also have some "false negatives," meaning that it would remove some effective would-be teachers as well.9

Among the four remaining studies of the 15, Rockoff et al. reported finding no statistically significant relationship between passing the test on the first attempt and future classroom effectiveness, although only 8% of the teachers in the study reported failing the licensure test on their first attempt. ¹⁰ Buddin & Zamarro (2009) reported a null relationship between teachers' test scores on most of California's licensure tests and their students' test scores in Los Angeles Unified School District (LAUSD), but they did find a small negative association between teachers' scores on the CSET (a California test of content knowledge) with student reading achievement. They also found no relationship between whether teachers passed or failed tests on their first attempt and student outcomes, However, this study only evaluated pass licensure test data for teachers from the California State University system (representing about 17% of teachers in the school district analyzed) and only included teachers who passed their exams. ¹¹

The remaining two studies (of the four finding mixed or null results) illustrate the complexity of research on licensure tests.

Hanushek's (1997) meta-analysis of research on school resources included 41 studies that addressed teacher tests. He found that a third of studies (37%) found a positive relationship between teacher test scores (not always licensure tests) and student outcomes (largely but not exclusively referring to standardized test scores). While 10% of the studies found a negative relationship, Hanushek's conclusion was that "of all the explicit measures that lend themselves to tabulation, stronger teacher test scores are most consistently related to higher student achievement." ¹²

Goldhaber and Hansen's 2010 study found that the licensure tests had predictive validity but did not hold constant across all groups of teachers. Using Praxis II data from North Carolina, Goldhaber and Hansen found that the Praxis II tests appear to predict teachers' effectiveness in math but not reading. Digging deeper, they found that teacher licensing tests function differently as a "signaling tool" of teacher effectiveness, depending on a teacher's race, with the "Curriculum, Instruction, and Assessment test" (actually a test of content knowledge) being a better predictor of effectiveness for White teachers, and the "Content Area Exercises test" (actually a free–response essay exam focused on responding to teaching situations) being a better predictor for Black teachers. Moreover, the benefits of Black students learning from Black teachers (regardless of their licensure test performance) was equivalent to those

students having a White teacher who did well on the licensure test: "[W]hen teaching Black students, Black teachers in the lower end of the teacher test distribution are estimated to have impacts that are approximately the same as White teachers at the upper end of the distribution."

Finally, a study not included in the 15 described above explored a slightly different question with regard to licensure tests. Bastian et al. (2020) found that when cooperating teachers, the mentors who host teacher candidates for their student teaching placement, earned higher scores on their licensure tests, their student teachers went on to be higher rated early career teachers. ¹³

Relevant studies and findings

Citation

Araujo P. M. D., Heineck, G., & Cruz Aguayo, Y. (2020). Does test-based teacher recruitment work in the developing world? Experimental evidence from Ecuador (No. 165), BERG Working Paper Series

Findings

Significant, positive relationship between test-screened, tenured teachers and their students' performance on language and math, even after controlling for teacher education, experience, cognitive ability, personality traits and classroom practices. Outcomes were greater for language than math, and even more pronounced for students from socioeconomically disadvantaged backgrounds.

Notes

Based on data and a very different teacher screening system utilized in a developing country (Ecuador). Study also found that the characteristic of being tenured worked in combination with the test screen.

Test(s) reviewed and in what state(s)

"National skill and content knowledge tests" administered to tenured and non-tenured teachers in Ecuador

Outcomes measured

Students' performance in math and language

Citation

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 (No. w14021). National Bureau of Economic Research.

Findings

Estimated the effect of various teacher attributes using a value-added model and found a significant, positive relationship between teachers' math SAT scores and 4th and 5th grade student achievement in math in New York City schools This study found positive but not statistically significant results for passing the LAST on the first attempt and for LAST scores but notes that LAST scores are highly correlated with SAT scores, and so a model that considers both may undervalue the contribution of each.

Notes

Attributes much of the achievement-gap reduction in NYC between high- and low-poverty schools from 2000 to 2005 to the addition of new teachers with higher academic qualifications (including test scores).

Test(s) reviewed and in what state(s)

A composite of New York City teacher characteristics, including SAT scores and certification exam scores of individual teachers on the LAST (Liberal Arts and Sciences Test) and whether they passed on their first attempt

Outcomes measured

Students' performance in math and language

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103-115.

Findings

Found no relationship in most cases between teacher subject knowledge (as measured by their licensure test score) and student achievement and found a small negative relationship in the case of CSET scores and student reading achievement. Also found a small positive relationship between the RICA and student achievement in math in one model. Study also found no relationship between student outcomes and teachers failing one or more licensure tests on their first attempt.

Notes

Pass rate data is limited to six cohorts of graduates from the CSU system, representing about 17% of teachers in LAUSD. Study only includes teachers who passed their exams, and it presumably is based on best-attempt passing scores, rather than first-attempt passing scores (which some studies use instead), although the study also evaluates the relationship between student outcomes and teachers failing on their first attempt.

Test(s) reviewed and in what state(s)

CSET (subject-matter knowledge), CBEST (basic skills test), and RICA (reading pedagogy) tests in Los Angeles, California

Outcomes measured

Grades 2-5 student test scores on the California Achievement Test in reading and math

Citation

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(3), 449-465.

Findings

The study found a moderately strong positive relationship between certification exam performance and teachers' classroom effectiveness. Failing the certification exam one or more times was associated with significantly lower student test scores (.036 standard deviations lower in math for grades 4-5 math, .028 standard deviations lower in grades 6-8 math, and .03 standard deviations lower in grades 6-8 reading). There was no relationship with grades 4-5 reading scores.

Notes

Only 2% of teachers in the sample failed their certification score at least once. For 54% to 57% of teachers, certification exam scores were unknown.

Test(s) reviewed and in what state(s)

Florida's state certification exam

Outcomes measured

Value-added models that control for student characteristics, as well as school and teacher fixed effects, for grades 4-8 students

Citation

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *The Journal of Human Resources*, 41(4), 778-820.

Findings

Found a modest, positive relationship between teachers' licensure test scores and their students' performance on standardized assessments. A one standard deviation increase in teacher test scores was associated with up to a 0.017 standard deviation increase in average student scores in math and a smaller increase in reading.

Notes

Also finds that teachers with higher licensure test scores are more likely to work in schools with a higher proportion of White students, fewer students qualifying for free and reduced-price lunches, more college-educated parents, and higher prior-year test scores.

Test(s) reviewed and in what state(s)

Elementary Education and the Early Childhood Education licensing tests (1960s-1990s) and Elementary Education Curriculum and Elementary Education Content (mid-1990s to 2001) in North Carolina

Outcomes measured

Student test scores

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673-682.

Findings

Teachers' licensing test scores have a positive impact on student achievement, with larger effect sizes for math than reading.

Having a teacher at one of the extremes of the distribution of licensure test scores revealed a 0.130 standard deviation difference in students' math scores (between students whose teachers' test scores were 2 or more standard deviations above average versus 2 or more standard deviations below average.

Notes

Test(s) reviewed and in what state(s)

Elementary Education and the Early Childhood Education licensing tests (1960s-1990s) and Elementary Education Curriculum and Elementary Education Content (mid-1990s to 2001) in North Carolina

Outcomes measured

Student test scores and test score gains in reading and math for grades 3-5

Citation

Cowan, J., Goldhaber, D., Jin, Z., & Theobald, R. (2020). Teacher licensure tests: Barrier or predictive tool? CALDER Working Paper No. 245-1020.

Findings

MTEL scores are positive and statistically significant predictors of teachers' effectiveness (as measured by in-service performance ratings and value added to student test scores). A one standard deviation in teachers' MTEL performance is associated with an improvement in student test scores of about 0.024 when adjusting only for observable student characteristics but is weaker for language arts scores as compared to math.

Notes

Finds that licensure test scores are predictive of value added for both White teachers and teachers of color, and that licensure test scores are more predictive of performance ratings for teachers of color than for White teachers. Also finds that candidates of color are less likely to retake the licensure test if they fail on their first attempt, compared with White candidates.

Test(s) reviewed and in what state(s)

Massachusetts Tests for Educator Licensure (MTEL)

Outcomes measured

Teachers' value-added scores and performance ratings

Citation

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28, 465-498.

Findings

Relicensing test scores were a significant, positive predictor of student performance.

Notes

Finds that licensure test scores are predictive of value added for both White teachers and teachers of color, and that licensure test scores are more predictive of performance ratings for teachers of color than for White teachers. Also finds that candidates of color are less likely to retake the licensure test if they fail on their first attempt, compared with White candidates.

Test(s) reviewed and in what state(s)

Texas' relicensing test covering teachers' basic literacy skills

Outcomes measured

Their teachers' relicensing test performance was the most important school-level predictor of math and reading outcomes for students on standardized tests.

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.

Findings

Found that teacher licensure test scores were predictive of teacher effectiveness, particularly in math. Praxis II "Curriculum" test was a better predictor of student performance than the "Content" test, with students of teachers in the top quintile of the Curriculum test scoring 2.2% standard deviation higher in reading and 3.5% standard deviation higher in math than those whose teachers scored in the bottom quintile.

Notes

Identified policy limitations to findings, noting trade-offs of increasing the test score cutoff in NC to the higher standard used by CT, which would exclude an estimated 0.5% of ineffective teachers vs. 7% estimated to be effective.

Test(s) reviewed and in what state(s)

NTEs, Praxis I (basic skills), and Praxis II Multiple Subjects tests in North Carolina

Outcomes measured

Student standardized test scores in reading and math tied to North Carolina's Standard Course of Study (controlling for prior test scores and other student characteristics)

Citation

Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, 47(1), 218-251.

Findings

Positive correlation between passing the licensure test and teacher effectiveness in math but not in reading.

Notes

Warned that the two tests (CIA and CAE) function differently as "signals of quality," depending on a teacher's race, with CIA scores (which actually measure content knowledge) being a better predictor for White teachers and CAE (which actually measures responses to classroom situations) being a better predictor for Black teachers' effectiveness. Race/ethnicity matching (Black teachers with Black students) could be as significant as licensure test performance in predicting student outcomes.

Test(s) reviewed and in what state(s)

Praxis II Curriculum, Instruction, and Assessment test (CIA) and the Praxis II Content Area Exercises test (CAE) in North Carolina

Outcomes measured

Student achievement using end-of-grade reading and math tests for grades 4-6, controlled for prior-year test scores

Citation

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19(2), 141-164.

Findings

A meta-analysis of research exploring how school resources affect student performance asserted that about a third (37%) of the 41 studies that included teacher test scores found a positive relationship between teacher test scores (though not necessarily content licensure test scores) and student achievement. Ten percent of studies found a negative and statistically significant relationship. When limited to the studies that use value-added measures of student performance, 27% of the 11 studies examining teacher test scores found a positive and significant relationship, while 9% found a negative and significant relationship.

Notes

The authors conclude that "of all the explicit measures that lend themselves to tabulation, stronger teacher test scores are most consistently related to higher student achievement, even though only 37% provide positive and statistically significant effects."

Test(s) reviewed and in what state(s)

Varies by study

Outcomes measured

Three-quarter of the studies in the analysis measure student performance by standardized tests; other research includes measures such as dropout rates and labor market earnings.

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(1), 43–74.

Findings

Positive but not statistically significant relationship between passing licensure test and effectiveness and between passing and mentor evaluation

Notes

Only 8% of the sample failed the LAST on the first attempt.

Test(s) reviewed and in what state(s)

New York City teachers' Liberal Arts and Sciences Test (LAST) passage rates (on first attempt) and self-reported ACT/SAT scores

Outcomes measured

Student achievement data (math in grades 4-8) and subjective evaluations of teachers by their mentors

Citation

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.

Findings

Praxis licensure exams used are correlated to teacher effectiveness, but that correlation is relatively low. Praxis II is a better screening tool than Praxis I. Raising the state's cut score on Praxis II would increase the quality of the teacher workforce, as measured by value-added student achievement.

Notes

Study asserts that Important trade-offs have to be considered when using licensing test scores as a binary pass-fail screening tool, namely that potentially effective teachers, many of them minority candidates, may be removed from the pool.

Test(s) reviewed and in what state(s)

Praxis I & II in Arkansas

Outcomes measured

Grades 5-8 student VAM scores on standardized tests

Citation

Shuls, J. V., & Trivitt, J. R. (2012). What makes a teacher effective? An analysis of teacher credentials and student achievement. In Annual Conference Association for Education Finance and Policy, Boston, MA.

Findings

Significant, positive relationship between teacher test scores and student outcomes.

Praxis II scores are significantly correlated with increased student achievement in math in both elementary and middle school. The connection to language arts performance was marginally significant at the elementary level but not at the middle school level.

Notes

Test(s) reviewed and in what state(s)

Praxis I & II tests in Arkansas

Outcomes measured

Student test scores on state tests in English and math, grades 3-8

Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. *Educational Policy*, 29(4), 645-675.

Findings

Teachers' licensure test scores, particularly on Praxis II, were a significant, positive predictor of student achievement. Finds that test scores are more strongly predictive of students' math results than of ELA results. Teachers in the top 10% and top 25% of Praxis scores are significantly more effective at increasing student achievement in both subjects, with greater statistical significance in math.

Notes

Also finds that teachers' Praxis II scores are highly correlated with Praxis I scores.

Test(s) reviewed and in what state(s)

Praxis I & Praxis II (professional knowledge) tests in Arkansas

Outcomes measured

Student achievement data for grades 3-8

ENDNOTES

- Putman, H. & Walsh, K. (2021). State of the States 2021: Teacher Preparation Policy. Washington, D.C.: National Council on Teacher Quality. Retrieved from https://www.nctq.org/publications/State-of-the-States-2021:-Teacher-Preparation-Policy#content.
- Search Terms: ETS Research page ("Elementary content"); Google Scholar, between 1990 and 2020 ("elementary content" "licens* test" "certification test" "teacher" "student outcomes" "elementary" "core subjects" "core knowledge" "outcomes" "student learning" "learning gains" "Praxis" "teacher effective" "race" "Black" "Latino" "Hispanic" "Latinx" "predictive validity" "predict*" "licensing" "certification" "Pearson")
- 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 (No. w14021). National Bureau of Economic Research.
- Araujo P. M. D., Heineck, G., & Cruz Aguayo, Y. (2020). Does test-based teacher recruitment work in the developing world? Experimental evidence from Ecuador (No. 165). BERG Working Paper Series; Boyd et al. (2008); Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. Economics of Education Review, 30(3), 449-465; Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacherstudent matching and the assessment of teacher effectiveness. The Journal of Human Resources, 41(4), 778-820; Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). How and why do teacher credentials matter for student achievement? (No. w12828). National Bureau of Economic Research; Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. Harvard Journal on Legislation, 28, 465-498; 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; Cowan, J., Goldhaber, D., Jin, Z., & Theobald, R. (2020). Teacher licensure tests: Barrier or predictive tool? CALDER Working Paper No. 245-1020; Shuls, J. V., & Trivitt, J. R. (2012). What makes a teacher effective? An analysis of teacher credentials and student achievement. In Annual Conference Association for Education Finance and Policy, Boston; Shuls, J. V., & Trivitt, J. R. (2015). Teacher effectiveness: An analysis of licensure screens. Educational Policy, 29(4), 645-675; 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.
- Araujo P. M. D., Heineck, G., & Cruz Aguayo, Y. (2020). Does test-based teacher recruitment work in the developing world? Experimental evidence from Ecuador (No. 165). BERG Working Paper Series.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. Harvard Journal on Legislation, 28, 465–498.
- 7 Cowan, J., Goldhaber, D., Jin, Z., & Theobald, R. (2020). Teacher licensure tests: Barrier or predictive tool? CALDER Working Paper No. 245–1020.
- 6 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.
- 9 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.
- 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(1), 43–74.
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115.

- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19(2), 141–164.
- This study found that a standard deviation increase in cooperating teachers' licensure test scores was associated with a positive and moderately statistically significant increase in early career teachers' evaluation ratings and a positive but nonsignificant increase in value-added scores.

 Bastian, K. C., Patterson, K. M., & Carpenter, D. (2020). Placed for success: Which teachers benefit from high-quality student teaching placements? *Educational Policy*.