# Licensure Tests and Teacher Supply

Alexis Orellana, PhD* Marcus A. Winters, PhD†

**Abstract**

Public school administrators struggle to staff classrooms with qualified teachers, especially within shortage areas such as STEM. Applying a regression discontinuity design to administrative data from Connecticut, we provide causal evidence that licensure test requirements reduce the supply of eligible teachers by deterring candidates who fail their first attempt. The effect is especially large for those seeking an endorsement to teach STEM subjects. The deterrent effect grows substantially larger for scores further away from the passing threshold. Consistent with findings from other states, licensure test scores are only modestly correlated with a teacher's later value-added contribution to student test scores.

Acknowledgements: We thank the Connecticut State Department of Education (CSDE) for providing the data necessary for this project, and we especially benefited from insights from Shuana Tucker, Ajit Gopalakrishnan, and David Alexandro. The content of this paper belongs to the authors alone and does not necessarily represent the views of the CSDE. Meagan Comb and Sidrah Baloch provided essential support for the research-practice partnership that makes this research possible. We also received valuable feedback from Dan Goldhaber, Roddy Theobald, James Cowan, Andrew Bacher-Hicks, Olivia Chi, Kirsten Slungaard-Mumma, William Delgado, and participants at the Boston University PREREQ seminar. All remaining errors are our own.

*Alexis Orellana, Ph.D. is a Postdoctoral Associate at the Wheelock Educational Policy Center housed within the Boston University Wheelock College of Education & Human Development. borellan@bu.edu

†Marcus A. Winters, Ph.D. is an associate professor and chair of the Educational Leadership & Policy Studies department at the Boston University Wheelock College of Education & Human Development. He is also faculty director of the Wheelock Educational Policy Center. marcusw@bu.edu

# 1 Introduction

Public schools rely heavily on occupational licensure requirements to ensure the competency of instructional staff. All states require prospective teachers to pass some form of licensure test to gain certification, and 15 states require candidates to pass basic skills tests to gain admission into a teacher preparation program (Putman and Walsh, 2021). Recent evidence that increased teacher attrition spurred by challenges associated with the Covid-19 pandemic (Bacher-Hicks et al., 2022; Zamarro et al., 2022; Camp et al., 2022; Goldhaber and Theobald, 2022) is exacerbating persistent teacher shortages, especially within difficult-to-staff areas such as special education and STEM subjects (Boe and Cook, 2006; Goldhaber et al., 2015; Cowan et al., 2016; Dee and Goldhaber, 2017; McVey and Trinidad, 2019), has renewed longstanding arguments in favor of reducing or removing licensure test requirements in order to increase the supply of eligible teachers (Ballou and Podgursky, 1998). However, the extent to which test requirements actually serve as a barrier to entering the classroom is not entirely clear, and there remains concern that reducing certification requirements would harm students by allowing ineffective teachers to enter the classroom (Putman and Walsh, 2021).

We link licensure test records to nearly two decades of administrative data from Connecticut in order to shed new light on the causal effects of licensure test requirements on the teacher workforce. Our primary strategy leverages exogenous variation in failing licensure tests occurring at the passing threshold of the test score distribution. We find that failing the first attempt at a licensure test significantly and substantially deters prospective teachers from eventually earning certification and teaching in a public school.

For those scoring at the passing threshold, failing a basic-skills test required for entry into a teacher preparation program reduces the probability a candidate eventually earns any teaching certification in the state by about 3.7 percentage points, reduces the likelihood they obtain a certification to teach within a shortage area by 1.6 percentage points, and reduces the likelihood they teach within a Connecticut public school by about 2.2 percentage points, or about 5.1% relative to the average first-time test-taker. Failing a subject-matter certification test typically administered near the completion of a preparation program reduces the probability of eventually earning any teaching certification within the state by 5.4 percentage points and the likelihood of eventually gaining an endorsement to teach within STEM by about 9.4 percentage points for those scoring at the passing threshold. Further, we show that nearly all of these estimates understate the overall treatment effect because the deterring effect of failing a licensure test grows larger for scores further away from the passing threshold. For instance, for those who score 0.5 standard deviations below the cutoff, failing the relevant subject-matter certification test reduces the likelihood of obtaining any teaching certification by 10 percentage points and reduces the probability of eventually obtaining an endorsement to teach in a STEM subject by 15.1 percentage points.

The extent to which the barrier to entry imposed by a licensure test requirement is harmful or beneficial largely depends on whether scores on the tests are predictive of a prospective teacher's future effectiveness. We show that for those who become teachers there is a statistically significant but arguably insubstantial relationship between scores on licensure tests and a teacher's measured value-added contribution to student test scores. Our results from Connecticut are consistent with recent studies that similarly find at best modest correlations between licensure scores and later teacher impacts (Clotfelter et al., 2006, 2007; Goldhaber, 2007; Buddin and Zamarro, 2009; Goldhaber and Hansen, 2010; Chingos and Peterson, 2011; Rockoff et al., 2011; Shuls and Trivitt, 2015; Goldhaber et al., 2017; Shuls, 2018; Cowan et al., 2020).

Our results add to a limited body of research evaluating the effect of licensure test requirements on teacher supply. Prior studies describe the relationship between observed characteristics and licensure scores but do not compare the later certification or employment of those who fail to a counterfactual group of candidates who passed (Goldhaber and Hansen, 2010; Goldhaber et al., 2017; Rucinski and Goodman, 2019; Cowan et al., 2020). Hanushek and Pace (1995) found that college students in states with more stringent licensure requirements are less likely to become a teacher, but their use of only cross-state variation at a point in time limits the ability to give their results a causal interpretation.[1]

[1]Both Angrist and Guryan (2004) and Larsen et al. (2022) exploit panel variation in the stringency of licensure requirements to investigate the effect of licensure requirements on the proportion of teachers who graduated from competitive colleges. These papers speak more to the effect of licensure requirements on the quality of the teacher labor force than the number of eligible teachers.

Our estimates for the deterrent impact of failing licensure tests by subject area are unique within the literature and highly relevant for policymakers' recent emphasis on addressing persistent teacher shortages. Despite the common perception of broad teacher shortages, in reality administrators consistently struggle to staff classrooms in a few key areas (Boe and Cook, 2006; Goldhaber et al., 2015; Cowan et al., 2016; Dee and Goldhaber, 2017; McVey and Trinidad, 2019). For example, during the 2011-12 school year 17% and 19% of public schools reported difficulty filling vacancies within special education and mathematics compared to only 2% and 4% reporting difficulty filling vacancies in general elementary and social studies, respectively (McVey and Trinidad, 2019).

Our results are also relevant to research investigating the factors that lead individuals to pursue employment as a public school teacher. Prior authors have identified that individuals are motivated to become a teacher in part by inherent factors such as job security, love of a specific subject, and an altruistic desire to work with children (Bastick, 2000; Rinke, 2008; Roness and Smith, 2010; Fokkens-Bruinsma and Canrinus, 2014). Our findings add to evidence suggesting that outside factors that can be adjusted through policy such as local teacher salaries (Figlio, 1997) and the quality of alternative labor market opportunities (Bacolod, 2007; Falch et al., 2009; Nagler et al., 2020) also contribute to one's decision to pursue a teaching career.

Finally, we contribute to the broader literature on the impacts of occupational licensing on labor supply. About 30% of U.S. workers are employed in an occupation that requires a government license (Kleiner and Krueger, 2013). Recent studies have found evidence that increases in licensure restrictions reduce labor supply in a host of occupations including cosmetology (Adams et al., 2002), physical and occupational therapy (Cai and Kleiner, 2020), and certified public accountants (Jacob and Murray, 2006). In contrast, some studies have failed to find significant labor supply responses to changes in licensure requirements for nurses (DePasquale and Stange, 2016; Law and Marks, 2017).

The remainder of the paper proceeds as follows. Section 2 describes the data and licensure testing requirements in Connecticut. Section 3 describes our methods and identifying assumptions. Section 4 presents results. Finally, Section 5 summarizes and concludes.

# 2 Data

## 2.1 Licensure Tests

The typical certification process in Connecticut requires an applicant to complete a state-approved educator preparation program and pass the subject-specific tests required to obtain an endorsement in their area of specialization. During our sample period, the state employed tests related to both of these certification requirements, all of which were created and administered by Educational Testing Service (ETS). Minimum passing scores for each test are determined by the Connecticut State Department of Education (CSDE).

We observe records for all licensure tests submitted to CSDE each year from 1995 to 2021. ETS routinely submits to CSDE all scores from test-takers who list Connecticut as their state of residence, take the test in Connecticut, or specify a preference for their scores to be submitted there. Each record contains an individual identifier, test-type identifier, score, and date. This information allows us to observe and distinguish each administration and test taken by each candidate during the sample period. Unfortunately, we do not observe demographic characteristics, such as gender or race, for all test-takers because ETS does not report such information to CSDE as part of the score transfer.

Schools do not typically observe an applicant's licensure test score(s) as part of the hiring process. Schools observe a candidate's certification and endorsement status, and thus can infer that a candidate has passed the necessary licensure tests.

### 2.1.1 Screen for Entering an Educator Preparation Program: Praxis I

Praxis I, also known as the Praxis Core, measures the reading, writing, and mathematics skills and content knowledge of candidates entering teacher preparation programs. CSDE required individuals to pass Praxis I to gain admission into a state-recognized educator preparation program until 2016. Since then the state no longer permits educator preparation programs to use Praxis I to screen candidates for entry, but programs can use scores on the test to determine whether the candidate needs additional support in particular areas.[2]

Figure A.1 in the Online Appendix describes the number of total and first-time-taker scores on Praxis I reported to CSDE each year. The number of tests administered peaked in 2002 and declined during the first decade of the 2000's until dropping sharply in 2017 following the policy change. The specific sub-tests that make up the Praxis I assessment have also changed over time, as illustrated in Figure A.2 in the Online Appendix.

### 2.1.2 Subject-Matter Certification Test: Praxis II

The second relevant licensure test in Connecticut is the various forms of Praxis II, also known as Praxis Subject, which assesses knowledge of specific subjects, as well as general and subject-specific teaching skills and knowledge. Candidates typically take these tests during the final year of their preparation program as part of applying to obtain a teaching certification or endorsement to teach a particular subject.

Each of the several subject-matter tests is linked to a particular teacher endorsement. Table A.1 in the Online Appendix shows the link between some of the endorsement codes offered in Connecticut and the Praxis II tests required. Some endorsement codes involve passing more than one test (for example, *Elementary Grades, K-6*). In these cases, we group all sub-tests and employ the minimum score as the forcing variable in the analysis described in Section 3.[3]

Figure A.3 in the Online Appendix reports the number of Praxis II test-takers overall and representing first administrations. Test administrations peaked in the mid-2000's and have gradually declined, consistent with declines in teaching candidates in the state over time.

## 2.2 Certification Data

We link applicants' scores on licensure tests to Connecticut's certification data between 2002 and 2021. These records contain the certificate type (e.g., initial educator, provisional educator, profes-

[2] See the Public Act No. 16-41, *An Act Concerning the Recommendations of the Minority Teacher Recruitment Task Force*. `https://www.cga.ct.gov/2016/act/pa/pdf/2016PA-00041-R00SB-00379-PA.pdf`

[3] Table A.1 in the Online Appendix shows a few endorsements require an additional test, *Foundations of Reading*, which is not administered by ETS. We exclude this test from our analyses.

sional educator), the date when the certification was issued, and the endorsement code indicating the subject in which the license grants the teacher permission to instruct.

In order to gain an Initial Educator Certification in the state, in addition to passing the relevant Praxis II test, an individual must hold a bachelor's degree, complete required coursework in professional education, general education, in some cases complete a subject-area major, and provide a recommendation for certification from a state-approved program. Once they believe they have fulfilled the requirements, individuals apply for certification by creating an account on the Connecticut Educator Certification System and paying a nominal fee. Thus, obtaining a certification requires an individual to actively apply and demonstrate that they have completed necessary benchmarks implies that those who hold a certification have some interest in obtaining a teaching position beyond what is evidenced by simply passing the licensure test.

## 2.3 Employment Records

We observe staff assignment data in all Connecticut public schools between 2002 and 2020. These records contain a unique Educator Identification Number (EIN), school code, position, and, in the case of teachers, the subject taught. We use the EIN identifiers to match teachers' information across datasets. Additionally, we employ these records to estimate the effect on the likelihood of observing an applicant serving as a teacher for at least five years.

## 2.4 Additional Teacher and Student Administrative Data

Our analysis describing the relationship between scores on licensure tests and a teacher's later impacts on students requires data matching students to teachers within the state over time. Student-level data contains test scores, demographic characteristics, and participation in programs such as special education and English language supplemental services. We use course offerings and student-course-grade information to construct a classroom identifier and link students to their teachers.

When estimating teacher value-added we restrict the analysis to the set of classrooms assigned to educators with a valid identifier. In addition, we only consider classrooms linked to one teacher during the corresponding school year. This restriction is necessary to correctly identify each teacher's contribution in our analysis.

We link teachers to students with valid test scores in Language or Math in grades 3 through 8 for each year from 2014-15 through 2020-21, except for 2019-20, when students did not take the test due to the Covid-19 pandemic. We successfully matched 95% of students to a single classroom teacher.

# 3 Empirical Strategy

## 3.1 Estimating the Causal Effect of Failing a Licensure Test on Progressing Toward Becoming a Teacher

The central challenge with estimating the effect of an individual failing their first attempt on a licensure test on their pathway to becoming a teacher is that candidates who fail the tests might differ from candidates who pass in unobserved ways that are also correlated with their trajectory towards becoming a public school teacher. We overcome this challenge by exploiting the sharp discontinuity in passing a given test that occurs at the designated cutoff.

Let $i$ denote an applicant taking test $j$ for the first time. Each test $j$ has a minimum passing score $\bar{x}_j$. We center scores around the corresponding cutoff and standardize them using the within-sample standard deviation.[4] We denote this variable $x_{ij}$. When a test $j$ considers more than one subtest, we define $x_{ij}$ as the minimum value across all subtests. Our main analyses are based on a sharp regression discontinuity design using the following specification:

$$y_{ij} = \alpha + f(x_{ij}) + \beta \mathbb{1}(x_{ij} < 0) + \epsilon_{ij} \tag{1}$$

The term $f(x_{ij})$ is a parametric function of the (normalized) score obtained by applicant $i$. In

[4]We employ standardized scores instead of raw scores because sometimes tests differ in their scale. For example, each applicant must approve two exams to earn an endorsement in Chemistry. The first one, *Chemistry: Content Knowledge*, is scored using 1-point intervals while *Chemistry: Content Essays* uses 5-point intervals.

our main analysis, we consider local linear regressions where $f$ is a linear function allowing for changes in the slope at the cutoff value:

$$f(x_{ij}) = \gamma_0 x_{ij} + \gamma_1 x_{ij} \times \mathbb{1}(x_{ij} < 0) \tag{2}$$

Finally, since tests have changed over time and they differ in terms of specific-subject knowledge, we deconstruct the unobserved component $\epsilon_{ij}$ in equation (1) into a test-specific component, a year-specific term, and a random term:

$$\epsilon_{ij} = \phi_j + \phi_t + \nu_{ij} \tag{3}$$

The first component, $\phi_j$, captures time-invariant factors common to all applicants taking a particular test (e.g., type of skills measured, test stringency). Including this set of fixed effects implies that identification of $\beta$ comes from comparing the outcomes of applicants who took the same test but differ only in that some of them passed on their first attempt while the other group failed. The second component, $\phi_t$, captures year-specific variation common to all test-takers in a given year. Finally, the residual term $\nu_{ij}$ represents idiosyncratic applicant-test level variation not captured by the explanatory variables.

In our analysis of Praxis I, we adjust equation (1) to separately measure the impact of failing the test during and following years when the state no longer allowed the test to be used as a screen for admission into a teacher preparation program. We do this by adding an interaction between the indicator for failing the test and an indicator for whether the observation is in 2016 or later.

Our key identifying assumption is that the relationship between a candidate's score on the test and the outcome would be smooth at the passing threshold if not for the fact that scoring above the line satisfied the passing requirement. In addition, we assume that individuals should not be able to manipulate their scores around the cutoff. The institutional features of the certification process in Connecticut make the violation of these assumptions very unlikely.[5]

[5] Figure A.4 in the Online Appendix shows the distribution of Praxis I test scores between the years 1995 and 2021.

We estimate equation (1) using local-linear regressions and a bandwidth of 1 standard deviation around the cutoff. In the Online Appendix we also report estimates from models that use optimal bandwidths calculated using the methodology of Calonico et al. (2014) (hereafter, CCT) and that estimate equation (1) using a uniform kernel. The results from alternative specifications are qualitatively similar, and in several cases larger than our primary estimates.

### 3.1.1 Interpreting Effects Away from the Passing Threshold

The estimate for $\beta$ represents the causal effect of failing the licensure test on the respective outcome locally for those with scores at the passing cutoff. This estimate will understate the average treatment effect if the impact of failing a licensure test increases for scores away from the threshold.

Dong and Lewbel (2015) show that under weak conditions within a sharp RD design the derivative of the treatment effect with respect to the forcing variable, which they label the Treatment Effect Derivative (TED), can be applied to estimate causal effects for observations away from the threshold.[6] Specifically, we can interpret $\gamma_1$ in equation (2) as the additional effect of the treatment for individuals at different points on the distribution for the forcing variable.[7]

Dong and Lewbel (2015) further show that within a sharp RD design the TED can also be interpreted as the change in the treatment effect that would result from a marginal change at the threshold under the assumption that an infinitesimal change in the assignment mechanism does not impact outcomes separate from the mechanism to assign treatments. However, we do not offer such an interpretation here because we suspect that this local policy invariance assumption likely does not hold in our case. It is at least plausible that differences in the effect of failing a licensure test on a candidate's probability of moving forward to become a teacher could be driven by the distance between the individual's score from the passing threshold rather than some underlying

We present densities separately because not all Praxis I tests have the same scale. These histograms do not show any evidence of manipulation around the cutoff values.

[6]The key assumption is that the forcing variable is differentiable within the neighborhood of the threshold, which Dong and Lewbel (2015) argue holds in any known application of sharp RD.

[7]Other authors (e.g., Bruce and Carruthers (2014), Bucarey et al. (2020), Eren et al. (2022), Aguirre (2022)) have utilized the covariate-based matching strategy for inferring treatment effects away from the threshold within an RD design described by Angrist and Rokkanen (2015). This approach is not available in our case, however, because our data do not satisfy its required conditional independence assumption.

difference in ability reflected by the score itself. That is, the effect of scoring any given number of points below the threshold may be similar regardless of where on the test score distribution the threshold is set.

## 3.2 Estimating the Relationship Between Licensure Scores and Later Teacher Value-Added

Similar to prior studies, we apply a two-stage approach to estimate the relationship between Praxis scores and later teacher impacts. The first stage uses a conventional value-added approach to estimate for each teacher the difference in the average test scores of students they instruct and the score that these students would be predicted to achieve based on their prior year test scores and other observed characteristics. The general model takes the form:

$$y_{ijst} = X'_{ijst}\beta + f(y_{ijst-1})\lambda + \phi_j + \xi_{ijst} \tag{4}$$

Where $y_{ijst}$ is the test score for student $i$ instructed by teacher $j$ within school $s$ during year $t$; $X$ is a vector of student and classroom characteristics and grade fixed effects; $f(y_{ijst-1})$ is a cubic function of the student's test score at the end of the previous year in math and language; $\phi_j$ is a teacher fixed effect; $\epsilon_{ijst}$ is a stochastic term, and $\beta$ and $\lambda$ are parameters to be estimated.

The objective of this step is to isolate $\hat{\phi}_j$, which is our estimate of teacher $j$'s contribution to student test scores conditional on the other covariates. Following the teacher value-added literature, we shrink our raw estimates to produce empirical Bayes estimates of teacher effects. Figure A.5 shows the distribution of the raw teacher fixed-effects $\hat{\phi}_j$ and the empirical Bayes estimates. Online Appendix A.2 provides further details.

We employ a cubic function for lagged test scores in order to allow for differences in expected growth for students at different points on the distribution of prior test scores. Prior research demonstrates that value-added models that account for prior test scores appear to be forecast unbiased when applied within large-scale administrative data (Kane et al., 2008; Chetty et al., 2014a; Koedel et al., 2015; Bacher-Hicks et al., 2019b).

For the second step in the analysis, we aggregate the data to the teacher level and estimate a regression where the dependent variable is the shrunken teacher's estimated value-added from the first stage, $\hat{\phi}_j$, and the independent variable is the teacher's score on the licensure test in question ($P_j$). Formally:

$$\hat{\phi}_j = \delta_0 + \delta_1 P_j + \eta_j \tag{5}$$

The estimate for $\delta_1$ represents the relationship between the teacher's score on the licensure test and their estimated value-added contribution to student test scores. We use this approach to separately investigate the predictive validity of the Praxis I and Praxis II tests on estimated test score value-added in ELA and math.

# 4 Results

## 4.1 Effect of Failing First Attempt at Praxis I Tests on Trajectory to Becoming a Teacher

We start our analysis by illustrating in Figure 1 nonparametric estimates for the relationship between the score obtained on Praxis I tests and the probability of completing additional steps towards becoming a public school teacher. The illustrated analyses are restricted to include only those who took Praxis I prior to 2016, when passing the test was used as a screen for entering an educator preparation program.

On each figure there is a distinct discontinuity in the value for the respective outcome at the threshold. With the exception of obtaining a certification in a hard-to-staff subject, we also observe a clear change in the slope of the relationship between the forcing variable and the outcomes occurring at the cutoff. The relationship between Praxis I score and each outcome is essentially flat among passing scores but distinctly positive to the left of the cutoff. These patterns suggest that failing the first administration of the Praxis I test reduces the likelihood of moving forward toward becoming a teacher, and the magnitude of the effect grows larger for scores that are further from the passing threshold.

Table 1 shows estimates of equation (1) for each of these outcomes. We first consider our results for the effect of failing the test estimated at the passing threshold for years that it was used as a screen, found in the coefficient on $Failed$. Each regression finds a significant negative impact of failing the test on the respective outcome. For those scoring at the threshold, the test during this period decreased the probability that a candidate eventually gained any certification in the state by about 3.7 percentage points. Column (4) shows the effect of not passing the threshold on the probability of obtaining a certification in a subset of subjects defined as "hard-to-staff", which include STEM subjects (e.g., math, biology, chemistry, physics), English for Speakers of Other Languages (ESOL), and special education. Failing the first administration of Praxis I with a score at the threshold decreases the probability of earning a license in these areas by 1.6 percentage points, or about $11\%$ relative to the baseline. Finally, we examine to what extent failing Praxis I changes the likelihood of becoming a teacher and staying in the profession. Column (5) shows that failing Praxis I decreases the probability of being observed as a public school teacher in 2002-2020 by about 2.2 percentage points, or about $5.1\%$ relative to all test-takers. Column (6) shows that the probability of remaining in the profession for at least five years decreases by a similar magnitude. In this case, we restrict our analysis to the sub-sample of test takers before 2016 and thus we do not include the interaction $Forcing \times Year > 2016$. The point estimate of 2.2 percentage points corresponds to a decrease of $7.3\%$ relative to the baseline value.

We next turn to consider differences in the treatment effect for scores further away from the threshold. Consistent with the patterns illustrated on Figure 1, for those on the passing side of the threshold the relationship between the forcing variable and the outcome, found on the coefficient for $Forcing$, is positive but only of a modest magnitude. The significant positive coefficient on $Forcing \times Failed$ found for all outcomes except certification in a hard-to-staff subject implies that those who score further below the threshold are substantially less likely to move forward toward becoming a teacher than those who score at the passing threshold. For instance, candidates who score 0.5 standard deviations below the threshold on Praxis I are 9.6 percentage points less likely to obtain teacher certification and 5.5 percentage points less likely to be observed as a public school teacher in 2002-2020.

Finally, we consider the impact of failing the Praxis I test after 2016 when passing was no longer required to enter an educator preparation program. The estimated treatment effect at the cutoff during this period is found by the sum of $Failed$ and the interaction term $Failed \times Year > 2016$. During this period, failing Praxis I had no significant effect on the probability that a candidate eventually took Praxis II, passed Praxis II, obtained a teaching certification in any area, or was observed as a Connecticut public school teacher. However, we did not observe a significant difference across the pre- and post-requirement periods in the impact of failing the test on the likelihood the candidate obtained a certification in a hard-to-staff subject.

In the Online Appendix, we show the robustness of our main estimates to different choices of the bandwidth and kernels used to estimate (1). Panel (a) in Figure A.6 presents our estimates for different outcomes using various bandwidth choices. Overall, our estimates are similar, although there is a loss of precision in some cases. In addition, panel (a) in Figure A.7 shows how our reported estimates using the methodology of Calonico et al. (2014) vary when we use a triangular or Epanechnikov kernel. Overall, we do not find substantial changes.

## 4.2 Effect of Failing First Attempt at Praxis II Tests on Trajectory to Becoming a Teacher

Figure 2 illustrates non-parametric estimates for the relationship between the standardized score obtained on Praxis II tests and later outcomes. In each case there is again a distinct jump in the outcome occurring at the passing threshold and an upward trend in the relationship between scores and the respective outcomes to the left of the cutoff that flattens on the passing side of the threshold.

The results reported in Table 2 are consistent with these patterns. At the cutoff, failing the first administration of a subject-matter test decreased the likelihood of obtaining any teaching certification in the state by 5.4 percentage points. For scores at the threshold, failing the respective Praxis II test reduces the likelihood that a candidate eventually obtains an endorsement to teach a STEM subject by about 9.4 percentage points. The estimates for obtaining an endorsement in special education are negative but not statistically significant. We also find that failing Praxis

II reduces the probability of being observed as a teacher in a Connecticut public school by 1.6 percentage points. This estimate is statistically significant at the 10% level and is equivalent to a 3% decrease relative to the average level. In terms of retention, failing the first administration of a Praxis II test reduces the likelihood of an applicant teaching for at least five years by 1.5 percentage points, a decline of 3.8%.

The slope for the relationship between scores and the outcomes differs significantly on either side of the threshold. The estimates for $Forcing$ are near zero and for all but one outcome statistically insignificant, indicating a flat relationship between scores and the outcome on the passing side of the threshold. However, except for special education, the coefficient on $Failed \times Forcing$ is significant and positive, indicating that the treatment effect grows for failing scores further from the cutoff. For example, for a candidate who scored 0.5 standard deviations below the threshold failing the test reduced the probability of obtaining an endorsement to teach STEM by about 15.1 percentage points.

In the Online Appendix we show the robustness of our main estimates to different choices of the bandwidth and kernel. Panel (b) in Figure A.6 and Figure A.7 show the robustness of our results to changes in each of these, respectively.

### 4.3 Relationship Between Licensure Scores and Later Teacher Effectiveness

Figure 3 shows the association between licensure score and our empirical Bayes estimates of test score value-added by subject and test-type. We find no significant relationship between score on the Praxis I test and a teacher's impact on student ELA test scores, and indeed the estimate is negative. In math, we find a significant positive association to Praxis I scores, though the magnitude is quite small. A one standard deviation increase in Praxis I score is associated with a 0.0056$\sigma$ increase in value-added in math. In practice, our estimates suggest that the average difference between the math value-added of a teacher who received the lowest observed score on Praxis I and the highest observed score is less than 0.02$\sigma$.

The bottom panels illustrate the relationship between Praxis II scores and estimated value-added for observed teachers. For both subjects, we find a small, positive relationship. For

ELA teachers, a standard deviation increase in Praxis II score associates with a gain of $0.0052\sigma$ in test scores. For math, a standard deviation increase in Praxis II scores associates with a gain of $0.0081\sigma$.

Our estimates from Connecticut are comparable to previous research assessing whether basic skills and subject area knowledge licensure tests are predictive of future teacher effectiveness. Our results are similar in terms of the magnitude of these associations and the differences previously found between Math and Language teachers.[8]

# 5 Conclusion

We provide evidence that licensure tests serve as a barrier to becoming a public school teacher and that scores on the tests have little relationship with the impact a teacher has on student learning. Our results make several unique contributions to a prior research on the impacts of occupational licensure requirements in education and have important implications for active policy conversations.

To our knowledge, we provide the first causal estimates for the extent to which failing a licensure test alters a potential teacher's trajectory towards entering the classroom. We find impacts at two distinct points of the teacher pipeline: when candidates take a screening test to enter an educator preparation program and when they take subject-matter certification tests as they near the end of their pre-service training. We find especially large impacts for prospective teachers pursuing endorsement to teach STEM subjects, a key shortage area.

The magnitude of the effect of failing licensure tests grows larger for scores further away from the passing cutoff. One potential interpretation of this pattern is that the licensure score is measuring some underlying attribute that also correlates with the treatment effect. In our view, a more plausible interpretation is that those who fail the tests consider the distance between their score and the passing threshold, and this would hold regardless of where on the distribution the

[8] Clotfelter et al. (2006) document that an increase of 1 s.d. in licensure tests associates with increases of $0.02\sigma$ and $0.01\sigma$ in Math and Reading test scores, respectively. Clotfelter et al. (2007) find increases of $0.015\sigma$ and $0.014\sigma$ in Math and Reading test scores, respectively. Clotfelter et al. (2010) find an association of $0.047\sigma$ in Algebra/Geometry and $-0.02\sigma$ in English. Finally, Goldhaber et al. (2017) find that an increase of 1 s.d. in the WEST-B basic skills tests associates with an increase of 0.01-$0.03\sigma$ in middle school math.

cutoff was placed.

That licensure test requirements would deter some individuals from entering a given profession is anticipated by theory (Kleiner, 2000) and has long been expected to be the case in public education (Ballou and Podgursky, 1998). In addition to confirming this suspicion, our results are important because they assign a meaningful magnitude to the effect. That failing a licensure test would substantially deter prospective teachers is not obvious given the combination of factors that many prospective teachers are driven by strong internal desire to work with children and that those who fail have unlimited opportunities to retake licensure tests that are viewed to be not especially rigorous (Putman and Walsh, 2021).

Our findings suggest that reducing or removing licensure test requirements would likey increase the supply of eligible teachers. However, policymakers considering such a change in the education workforce must weigh the need to increase teacher supply against the potential that doing so would dilute teacher quality.

Consistent with prior studies, we show in the Connecticut context that licensure scores are at best modestly correlated with a teacher's later effectiveness. Our most optimistic estimate suggests that a one standard deviation increase in a teacher's Praxis II score is associated with only a $0.0081\sigma$ increase in math value-added. In comparison, prior research finds that the difference in being assigned to a first-year teacher and a teacher with two years of experience is about $0.08$-$0.12\sigma$ depending on the specification used (Wiswall, 2013; Kraft and Papay, 2014)

Data limitations unfortunately prohibited us from investigating heterogeneity in the causal effect of failing licensure exams on the pathway to becoming a teacher by a candidate's race/ethnicity. Descriptive evidence that teachers who identify as members of historically underrepresented subgroups are less likely to pass licensure tests on their first attempt (Goldhaber, 2007; Rucinski and Goodman, 2019; Cowan et al., 2020) and are less likely to retake and eventually pass the test if they fail on their first attempt (Cowan et al., 2020) suggests that such analysis is highly necessary. An additional limitation is that we are only able to measure the relationship between licensure scores and value-added for those teaching within the tested grades, 3 through 8. We might expect that the subject-matter knowledge assessed by Praxis II could be especially important for high school teachers.

Finally, we note that given the importance of teachers for determining short- and long-term student outcomes (Kane and Staiger, 2008; Chetty et al., 2014b; Jackson, 2018; Gershenson et al., 2022), concern about weakening or even removing a key screen for a prospective teacher's quality is real and, in our view, quite reasonable. Though some may apply our results to call for the elimination of licensure requirements, another reasonable interpretation is that there is a need to identify screens that are better aligned with a teacher's later impacts. Regardless, it is important that policymakers considering changes to teacher employment systems weigh our results for the impact that licensure test requirements have on reducing teacher supply, along with the mounting evidence that scores on the tests are not highly predictive of a candidate's later impacts on their students' learning.

# References


Adams, A. F., Jackson, J. D., and Ekelund, R. B. (2002). Occupational licensing in a "competitive" labor market: The case of cosmetology. *Journal of Labor Research*, 23(2):261–278.

Aguirre, J. (2022). How can progressive vouchers help the poor benefit from school choice? evidence from the chilean voucher system. *Journal of Human Resources*, 57(3):956–997.

Angrist, J. and Rokkanen, M. (2015). Wanna get away? regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association*, 110(512):1331–1344.

Angrist, J. D. and Guryan, J. (2004). Teacher Testing, Teacher Education, and Teacher Characteristics. *American Economic Review*, 94(2):241–246.

Bacher-Hicks, A., Billings, S. B., and Deming, D. J. (2019a). "The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime". Working Paper 26257, National Bureau of Economic Research.

Bacher-Hicks, A., Chi, O. L., and Orellana, A. (2022). Two Years Later: How COVID-19 has Shaped the Teacher Workforce. EdWorking Paper 22-572, Annenberg Institute at Brown University.

Bacher-Hicks, A., Chin, M. J., Kane, T. J., and Staiger, D. O. (2019b). An experimental Evaluation of Three Teacher Quality Measures: Value-Added, Classroom Observations, and Student Surveys. *Economics of Education Review*, 73(C).

Bacolod, M. P. (2007). Do alternative opportunities matter? the role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89(4):737–751.

Ballou, D. and Podgursky, M. (1998). Teacher recruitment and retention in public and private schools. *Journal of Policy Analysis and Management*, 17(3):393–417.

Bastick, T. (2000). Why teacher trainees choose the teaching profession: Comparing trainees in metropolitan and developing countries. *International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft/Revue Internationale de l'Education*, 46(3/4):343–349.

Boe, E. E. and Cook, L. H. (2006). The Chronic and Increasing Shortage of Fully Certified Teachers in Special and General Education. *Exceptional Children*, 72(4):443–460.

Bruce, D. J. and Carruthers, C. K. (2014). Jackpot? the impact of lottery scholarships on enrollment in tennessee. *Journal of Urban Economics*, 81:30–44.

Bucarey, A., Contreras, D., and Muñoz, P. (2020). Labor Market Returns to Student Loans for University: Evidence from Chile. *Journal of Labor Economics*, 38(4):959–1007.

Buddin, R. and Zamarro, G. (2009). Teacher Qualifications and Student Achievement in Urban Elementary Schools. *Journal of Urban Economics*, 66(2):103–115.

Cai, J. and Kleiner, M. M. (2020). The labor market consequences of regulating similar occupations: The licensing of occupational and physical therapists. *Journal of Labor Research*, 41:352–381.

Calonico, S., Cattaneo, M., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82:2295–2326.

Camp, A., Zamarro, G., and McGee, J. B. (2022). Changes in teachers' mobility and attrition in arkansas during the first two years of the covid-19 pandemic. EdWorking Paper 22-589, Annenberg Institute at Brown University.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9):2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9):2633–79.

Chingos, M. M. and 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., and Vigdor, J. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3).

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, 41(4):778–820.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007). Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects. *Economics of Education Review*, 26(6):673–682.

Cowan, J., Goldhaber, D., Hayes, K., and Theobald, R. (2016). Missing Elements in the Discussion of Teacher Shortages. *Educational Researcher*, 45(8):460–462.

Cowan, J., Goldhaber, D., Jin, Z., and Theobald, R. (2020). Teacher Licensure Tests: Barrier or Predictive Tool. Technical report, CALDER Working Paper.

Dee, T. S. and Goldhaber, D. (2017). Understanding and Addressing Teacher Shortages in the United States. *The Hamilton Project*, 5:1–28.

DePasquale, C. and Stange, K. (2016). Labor supply effects of occupational regulation: Evidence from the nurse licensure compact. NBER Working Papers 22344.

Dong, Y. and Lewbel, A. (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics*, 97(5):1081–1092.

Eren, O., Lovenheim, M., and Mocan, H. N. (2022). The effect of grade retention on adult crime: Evidence from a test-based promotion policy. *Journal of Labor Economics*, 40(2):361–395.

Falch, T., Johansen, K., and Strøm, B. (2009). Teacher shortages and the business cycle. *Labour Economics*, 16(6):648–658.

Figlio, D. N. (1997). Teacher salaries and teacher quality. *Economics Letters*, 55(2):267–271.

Fokkens-Bruinsma, M. and Canrinus, E. T. (2014). Motivation for becoming a teacher and engagement with the profession: Evidence from different contexts. *International Journal of Educational Research*, 65:65–74.

Gershenson, S., Hart, C. M. D., Hyman, J., Lindsay, C. A., and Papageorge, N. W. (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, 14(4):300–342.

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., Cowan, J., and Theobald, R. (2017). Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. *Journal of Teacher Education*, 68(4):377–393.

Goldhaber, D. and Hansen, M. (2010). Race, Gender, and Teacher Testing: How Informative a Tool is Teacher Licensure Testing? *American Educational Research Journal*, 47(1):218–251.

Goldhaber, D., Krieg, J., Theobald, R., and Brown, N. (2015). Refueling the STEM and Special Education Teacher Pipelines. *Phi Delta Kappa*, 97(4):56–62.

Goldhaber, D. and Theobald, R. (2022). Teacher attrition and mobility in the pandemic. *Center for Analysis of Longitudinal Data in Education Research*.

Hanushek, E. A. and Pace, R. R. (1995). Who Chooses to Teach (and Why)? *Economics of Education Review*, 14(2):101–117.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, 126(5):2072–2107.

Jacob, J. and Murray, D. (2006). Supply-side effects of the 150-hour educational requirement for cpa licensure. *Journal of Regulatory Economics*, 30(2):159–178.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6):615–631.

Kane, T. J. and Staiger, D. O. (2008). "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation". Working Paper 14607, National Bureau of Economic Research.

Kleiner, M. M. (2000). Occupational licensing. *Journal of Economic Perspectives*, 14(4):189–202.

Kleiner, M. M. and Krueger, A. B. (2013). Analyzing the extent and influence of occupational licensing on the labor market. *Journal of Labor Economics*, 31(S1):S173–S202.

Koedel, C., Mihaly, K., and Rockoff, J. E. (2015). Value-Added Modeling: A Review. *Economics of Education Review*, 47:180–195.

Kraft, M. A. and Papay, J. P. (2014). Can professional environments in schools promote teacher development? explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4):476–500.

Larsen, B., Ju, Z., Kapor, A., and Yu, C. (2022). The Effect of Occupational Licensing Stringency on the Market for Public School Teachers. Technical report.

Law, M. T. and Marks, M. S. (2017). The labor-market effects of occupational licensing laws in nursing. *Industrial Relations: A Journal of Economy and Society*, 56(4):640–661.

McVey, K. P. and Trinidad, J. (2019). Nuance in the Noise: The Complex Reality of Teacher Shortages. *Bellwether Education Partners*.

Nagler, M., Piopiunik, M., and West, M. R. (2020). Weak markets, strong teachers: Recession at career start and teacher effectiveness. *Journal of Labor Economics*, 38(2):453–500.

Putman, H. and Walsh, K. (2021). State of the States 2021: Teacher Preparation Policy. *National Council on Teacher Quality*.

Rinke, C. R. (2008). Understanding teachers' careers: Linking professional life to professional path. *Educational Research Review*, 3(1):1–13.

Rockoff, J. E., Jacob, B. A., Kane, T. J., and Staiger, D. O. (2011). Can you Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, 6(1):43–74.

Roness, D. and Smith, K. (2010). Stability in motivation during teacher education. *Journal of Education for Teaching*, 36(2):169–185.

Rucinski, M. and Goodman, J. (2019). Racial Diversity in the Teacher Pipeline: Evidence from Massachusetts. *Policy Brief. Harvard Kennedy School: Rappaport Institute for Greater Boston. Cambridge, MA.*

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.

Shuls, J. V. and Trivitt, J. R. (2015). Teacher Effectiveness: An Analysis of Licensure Screens. *Educational Policy*, 29(4):645–675.

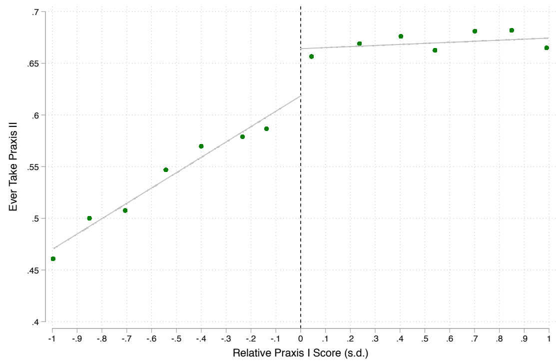
Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, 100(C):61–78.

Zamarro, G., Camp, A., Fuchsman, D., and McGee, J. B. (2022). Understanding how Covid-19 has changed teachers' chances of remaining in the classroom. EdWorking Paper 22-542, Annenberg Institute at Brown University.

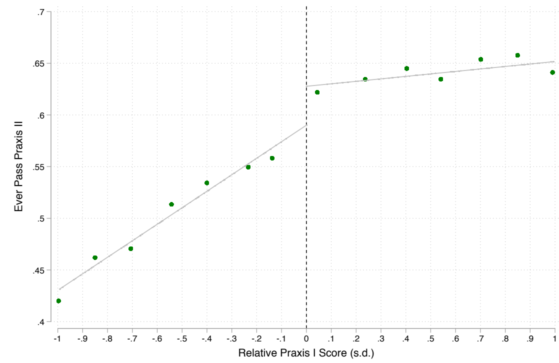
# 6 Figures and Tables

Figure 1: Nonparametric Estimates of the Effect of Failing Praxis I Test

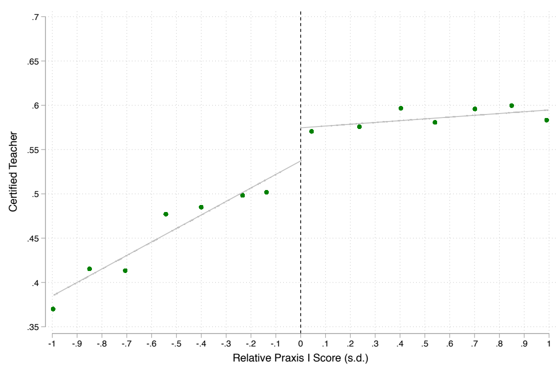
(a) Take Praxis II Test



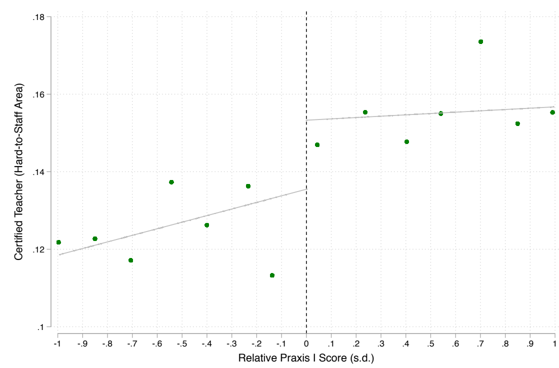

(b) Pass Praxis II Test



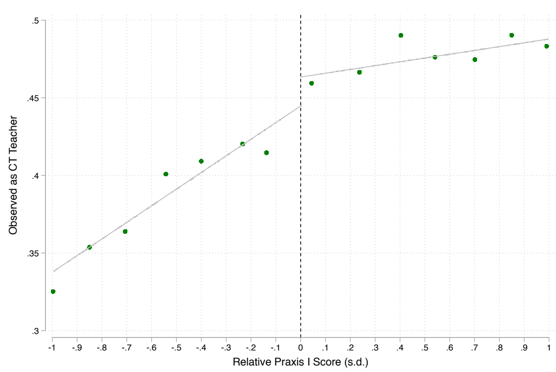

(c) Obtain Teacher Certification



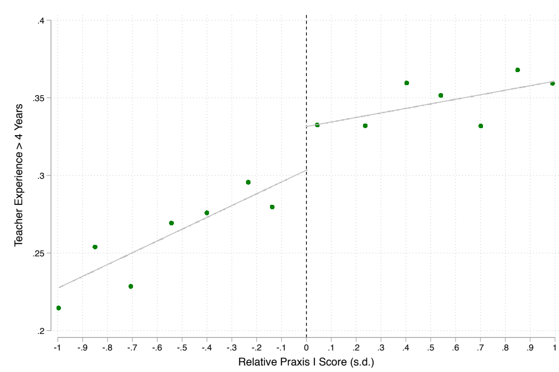

(d) Obtain Teacher Certification in Hard-to-Staff Area




(e) Observed in a Connecticut Public School



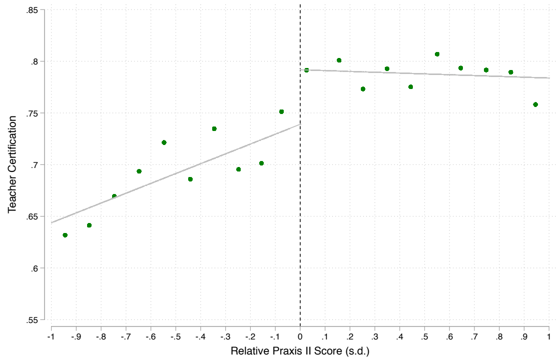

(f) Experience of 5 Years or More



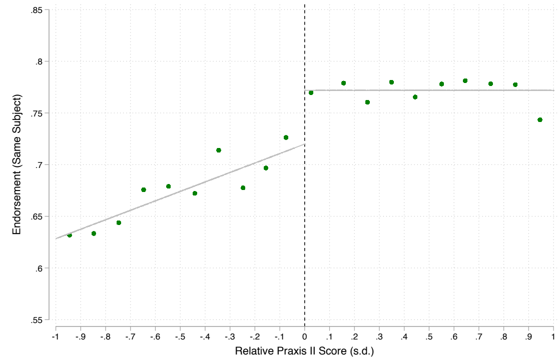

Notes: This figure illustrates the probability of observing different outcomes conditional on each applicant's Praxis I relative score. Panel (a) and (b) show the probability of taking and passing any Praxis II exam, respectively. Panel (c) shows the probability of earning teacher certification in any subject area. Panel (d) shows the probability of obtaining certification in STEM, Special Education, or English as a Second Language. Panel (e) shows the probability of observing the applicant in a Connecticut public school between 2002 and 2020. Panel (f) shows the probability of observing the applicant teaching in a Connecticut public school for five years or more. Green dots correspond to sample means of the dependent variable using 0.15 standard deviation bins. Grey lines correspond to the linear fit for points above and below the cutoff.

Figure 2: Nonparametric Estimates of the Effect of Failing Praxis II Test

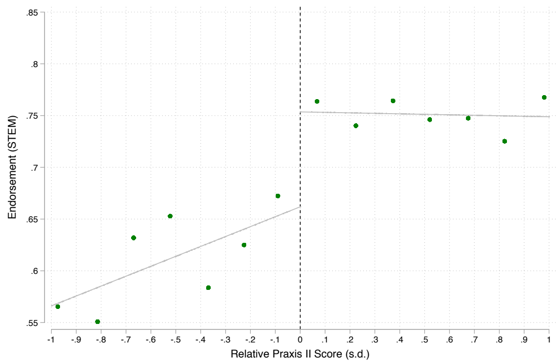
(a) Obtain Teacher Certification



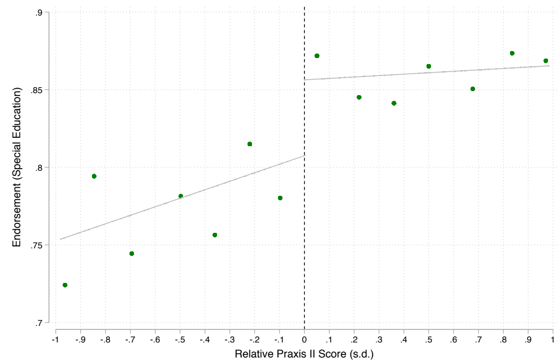

(b) Obtain Same-Subject Certification



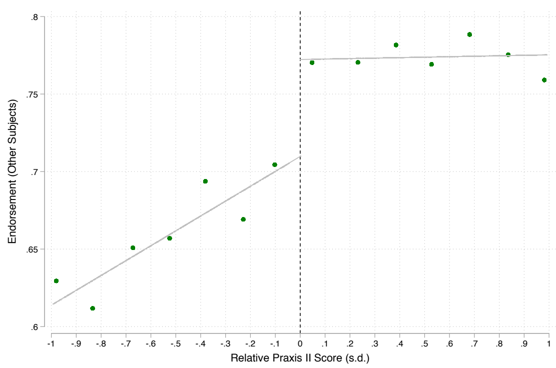

(c) STEM



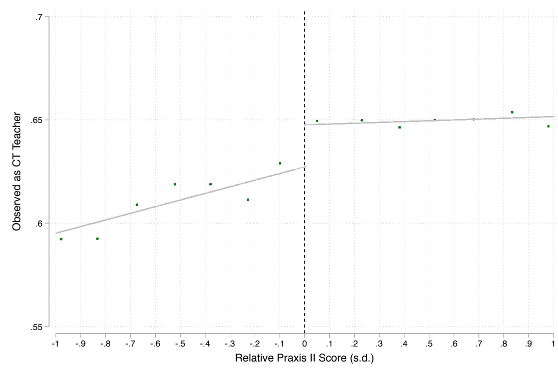

(d) Special Education




(e) Other Subjects



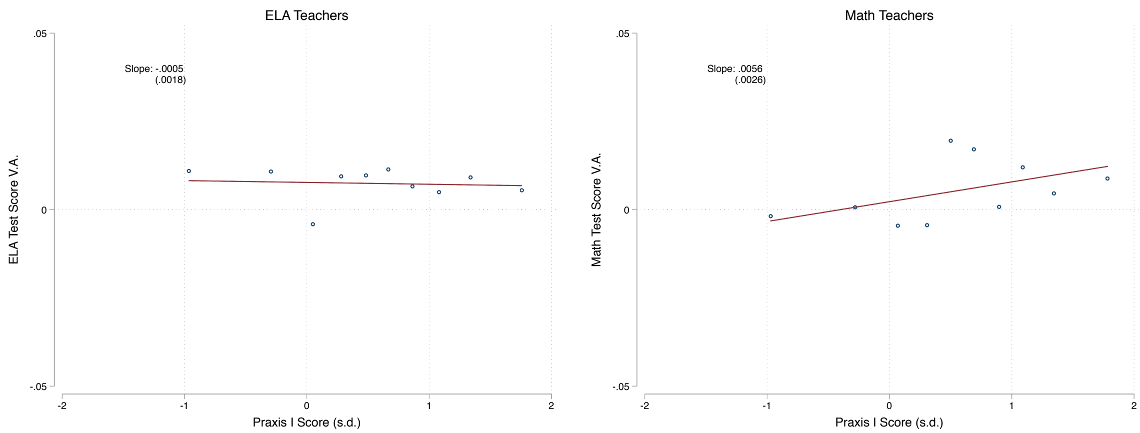

(f) Observed in a Connecticut Public School



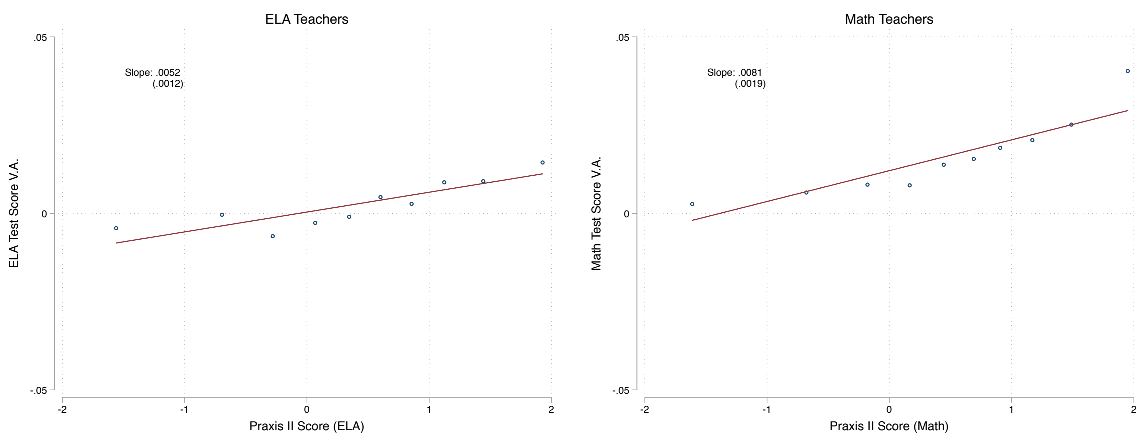

Notes: This figure illustrates the probabilities of observing different outcomes conditional on each applicant's Praxis II relative score. Panels (a) and (b) show the probability of earning teacher certification in any subject area and in the same Praxis II subject, respectively. Panels (c), (d), and (e) show the probability of earning teacher certification in STEM, Special Education, or English as a Second Language, for applicants who took the Praxis II tests in each subject area. Panel (f) shows the probability of observing the applicant teaching in a Connecticut public school between 2002 and 2020. Green dots correspond to sample means of the dependent variable using 0.1 standard deviation bins in panels (a) and (b), and 0.15 standard deviation bins in panels (c) - (f). Grey lines correspond to the linear fit for points above and below the cutoff.

Figure 3: Association between Praxis scores and Teacher Value-Added

(a) Praxis I




(b) Praxis II




Notes: This figure illustrates the relationship between each applicant's Praxis I and Praxis II score and measures of Language and Math test score value-added. Panel (a) shows the association between Praxis I standardized scores and test score value-added. Panel (b) shows the association using same-subject Praxis II standardized scores. Praxis scores correspond to the standardized running variable described in section 3. Test score value-added is estimated using a sample of math and language teachers linked to 3rd-8th grade students. We use specification (4) to obtain teacher fixed effects and then compute empirical Bayes estimates following Kane et al. (2008)

Table 1: RD Estimates of the Effect of Failing Praxis I

| | Take Praxis II | Pass Praxis II | Certification Any Area | Certification Hard-to-Staff | Observed Teacher | Experience $\geq$5 Years |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Failed | -0.045*** | -0.035*** | -0.037*** | -0.016** | -0.022** | -0.022** |
| | (0.010) | (0.010) | (0.010) | (0.007) | (0.010) | (0.009) |
| Forcing | 0.016* | 0.030*** | 0.020** | 0.013** | 0.016* | 0.008 |
| | (0.009) | (0.009) | (0.009) | (0.006) | (0.009) | (0.008) |
| Failed $\times$ Forcing | 0.110*** | 0.114*** | 0.117*** | -0.001 | 0.075*** | 0.053*** |
| | (0.018) | (0.018) | (0.018) | (0.013) | (0.018) | (0.017) |
| Failed $\times$ Year>2016 | 0.051** | 0.044* | 0.055** | -0.027 | 0.024 | |
| | (0.024) | (0.024) | (0.024) | (0.018) | (0.024) | |
| Average Outcome | 0.62 | 0.59 | 0.52 | 0.15 | 0.43 | 0.30 |
| N | 46749 | 46749 | 46749 | 46749 | 46749 | 44436 |

Notes: This table presents estimates of the effects of failing Praxis I scores on different outcomes. Each column shows estimates of equation (1) using a bandwidth of 1 standard deviation around the cutoff. Sample includes an individual's initial Praxis I test administration observed in years 1995 - 2021. Dependent variables in Columns (1) through (4) indicate if the individual ever took Praxis II, passed Praxis II, or obtained a certification at any point between initial Praxis I administration and 2021. "Hard-to-staff" areas include STEM subjects (e.g., math, biology, chemistry, physics), English for Speakers of Other Languages (ESOL), and special education. The dependent variable in Column (5) is an indicator for whether the individual was observed as a Connecticut public school teacher between 2002 and 2020, years for which we observe employment data. Column (6) is an indicator for whether the applicant is observed for at least five years as a Connecticut public school teacher using the sub-sample of applicants between 1995 and 2016. Each regression includes year and test fixed effects. Standard errors are presented in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2: RD Estimates of the Effect of Failing Praxis II

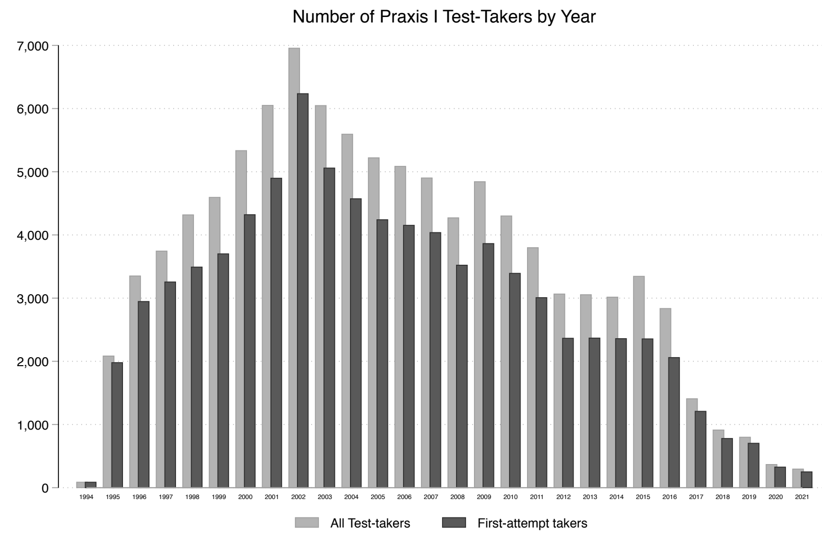
| | Certification Any Area | Certification Same Subject | STEM | Special Ed | Other Subject | Observed Teacher | Experience $\geq$5 Years |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Failed | -0.054*** | -0.051*** | -0.094*** | -0.033 | -0.056*** | -0.016* | -0.015* |
| | (0.008) | (0.008) | (0.021) | (0.025) | (0.009) | (0.009) | (0.009) |
| Forcing | -0.015* | 0.001 | -0.019 | 0.019 | -0.003 | 0.006 | -0.017* |
| | (0.008) | (0.008) | (0.022) | (0.020) | (0.009) | (0.009) | (0.009) |
| Failed x Forcing | 0.091*** | 0.081*** | 0.114*** | 0.069 | 0.081*** | 0.029* | 0.037** |
| | (0.015) | (0.015) | (0.038) | (0.046) | (0.016) | (0.017) | (0.016) |
| Average Outcome | 0.76 | 0.71 | 0.74 | 0.18 | 0.71 | 0.60 | 0.40 |
| N | 45833 | 45833 | 7606 | 4649 | 39370 | 45833 | 45833 |

Notes: This table presents estimates of the effects of failing Praxis II scores on different outcomes. The dependent variable in columns (1) - (5) is an indicator for whether the individual is ever observed to have obtained certification/endorsement in the area described at the top of the respective column. Column (6) is an indicator for whether the applicant is observed as a Connecticut public school teacher between 2002 and 2020, years for which we observe employment data. Column (7) is an indicator for whether the applicant is observed for at least five years as a Connecticut public school teacher using the sub-sample of applicants between 1995 and 2016. Each regression includes year and test fixed effects. Sample includes first-time Praxis II takers from 1995 - 2021 who scored within 1 standard deviation on either side of the passing threshold for the respective test. $Forcing$ is the centered score on the relevant Praxis II test for the respective outcome. Standard errors are presented in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

# A Online Appendix

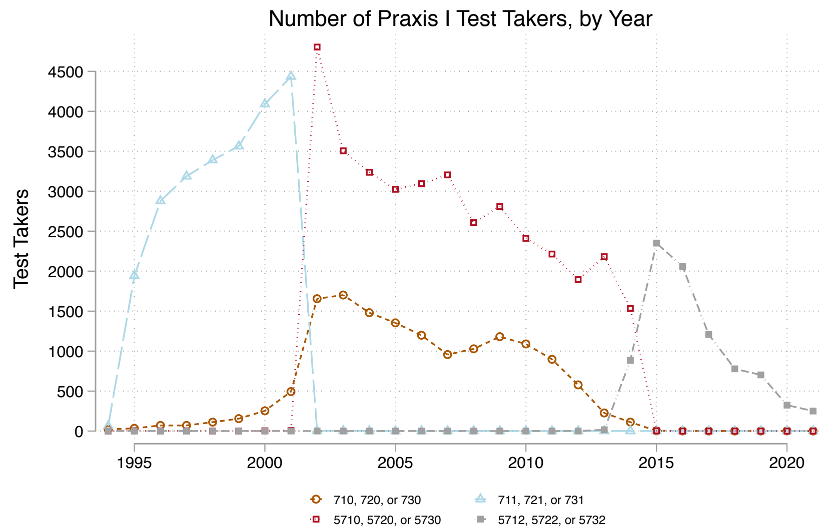
## A.1 Additional Figures and Tables

Figure A.1: Praxis I Test-takers



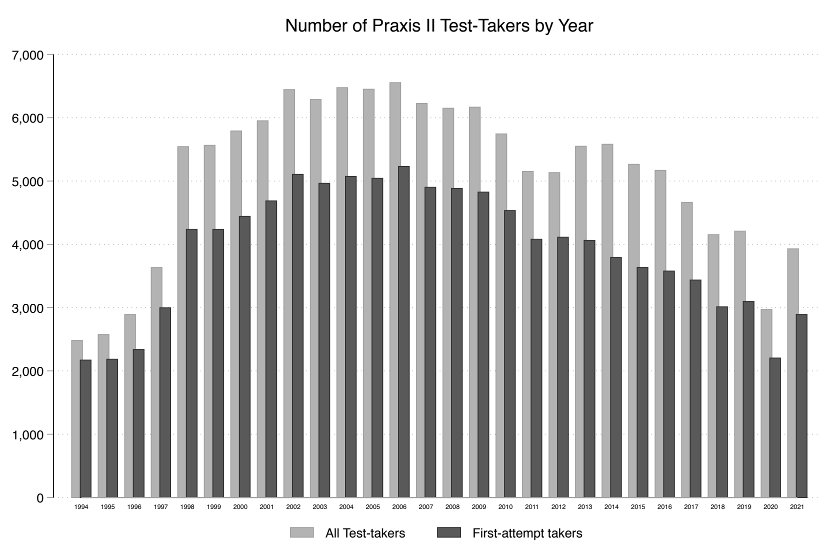

Notes: This plot shows the number of applicants who took a Praxis I test between 1995 and 2021. Gray and black bars represent the total number of test-takers and the total number of first-time applicants in each year, respectively.

Figure A.2: Different Praxis I tests over time




Notes: This plot displays the changes in Praxis I examinations between 1995 and 2021. Each examination consists of three subtests: reading, writing, and math. The lines show the number of applicants who took the corresponding set of subtests in each year.

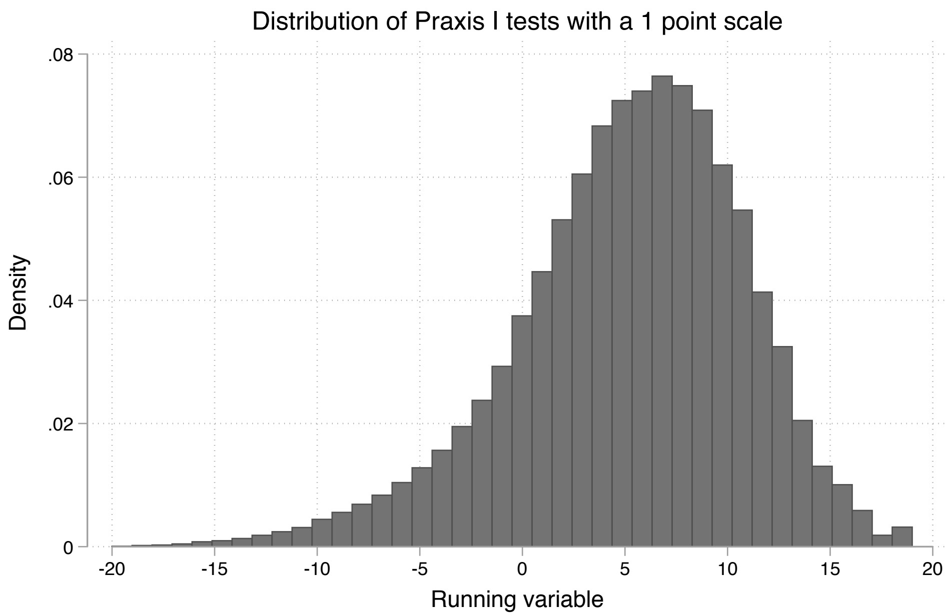
Figure A.3: Praxis II Test-takers



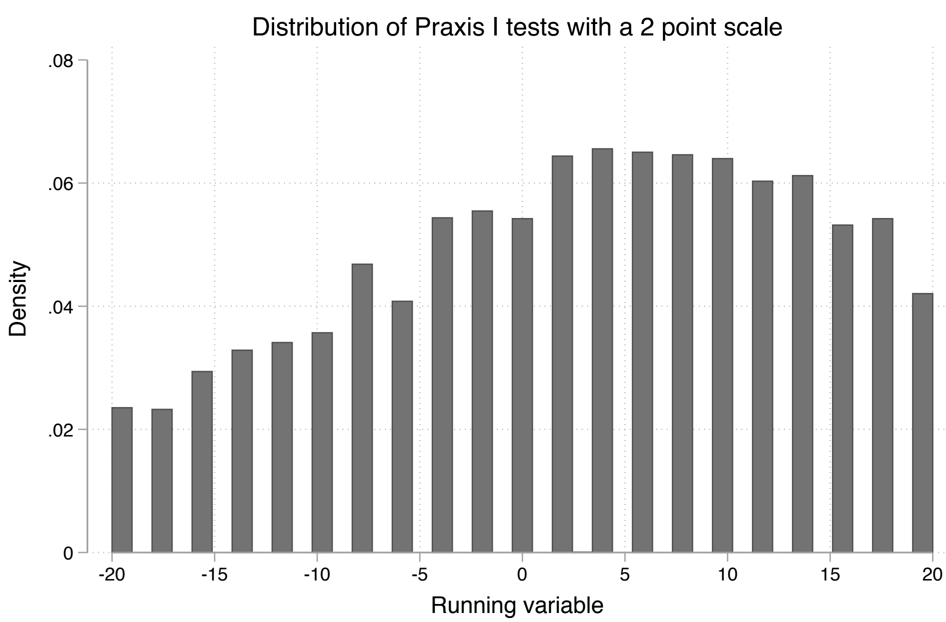

Notes: This plot shows the number of applicants who took a Praxis II test between 1995 and 2021. Gray and black bars represent the total number of test-takers and the total number of first-time applicants in each year, respectively.

Figure A.4: Distribution of Praxis I scores

(a) Tests using a 1-point scale




(b) Tests using a 2-point scale




Notes: This figure shows the distribution of the running variable for Praxis I tests between 1995 and 2021. Each panel shows the distribution for tests using a one-point and two-point scale, respectively.

Table A.1: Praxis II Tests and Teaching Endorsements in Connecticut

| Endorsement | Description | Praxis II Test | Additional Test |
|---|---|---|---|
| 13 | Elementary Grades K-6 | 5002 + 5003 + 5004 + 5005 | Foundations of Reading |
| 15 | English 7-12 | 44, 49 or 5039 | |
| 26 | History/Social Studies 7-12 | 81 or 5081 | |
| 29 | Mathematics 7-12 | 61 or 5161 | |
| 30 | Biology 7-12 | 235 or 5235 | |
| 31 | Chemistry 7-12 | 242 + 245 or 5245 | |
| 32 | Physics 7-12 | 262 + 265 or 5265 | |
| 33 | Earth Science 7-12 | 571 or 5571 | |
| 34 | General Science 7-12 | 433 + 435 or 5435 | |
| 47 | Technology Education PK-12 | 51 or 5051 | |
| 49 | Music PK-12 | 111+ 113 or 114 or 5114 | |
| 111 | TESOL PK-12 | 361 or 5362 | |
| 165 | Comprehensive Special Education K-12 | 543 or 5543 | Foundations of Reading |
| 215 | English Middle School 4-8 | 5047 | |
| 226 | History/Social Studies Middle School 4-8 | 89 or 5089 | |
| 229 | Mathematics Middle School 4-8 | 69 or 5169 | |
| 230, 231, 232, 233, 234, 235 | Middle Grades Science | 5540 | |
| 305 | Elementary Grades 1-6 | 5032 + 5033 + 5034 + 5035 | Foundations of Reading |

Notes: This table presents the Praxis II test requirements to earn a teaching certification in Connecticut. We employ this correspondence to identify whether applicants obtained a certification in the same Praxis II subject. The first and second columns display the code and subject-area description of each endorsement. The third column details which Praxis II tests are required in each case. The last column indicates whether an additional test (Foundations of Reading) is also required. This additional test is not used in our analyses since it is not administered by ETS.

Table A.2: RD Estimates of the Effect of Failing Praxis I Using Optimal Bandwidths

| | Take Praxis II | Pass Praxis II | Certification Any Area | Certification Hard-to-Staff | Observed Teacher | Experience $\geq$5 Years |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Failed | -0.055*** | -0.046** | -0.062*** | -0.038*** | -0.037* | -0.017 |
| | (0.019) | (0.018) | (0.023) | (0.014) | (0.021) | (0.015) |
| Average Outcome | 0.62 | 0.59 | 0.52 | 0.15 | 0.44 | 0.31 |
| Years | 1995-2016 | 1995-2016 | 1995-2016 | 1995-2016 | 1995-2016 | 1995-2016 |
| Fixed Effects | Year, Test | Year, Test | Year, Test | Year, Test | Year, Test | Year, Test |
| Bandwidth | (-0.56,1) | (-0.67,0.99) | (-0.46,0.98) | (-0.50,0.92) | (-0.48,1.03) | (-0.64,0.90) |
| N | 41020 | 41884 | 37748 | 36935 | 41843 | 39309 |

Notes: This table presents estimates of the effects of failing Praxis I scores on different outcomes. Each column presents estimates of equation (1) using the methodology proposed by Calonico et al. (2014) and a triangular kernel. Each regression includes year and test fixed effects. Standard errors are presented in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.3: RD Estimates of the Effect of Failing Praxis II Using Optimal Bandwidths

| | Certification Any Area (1) | Certification Same Subject (2) | STEM (3) | Special Ed (4) | Other Subject (5) | Observed Teacher (6) | Experience ≥5 Years (7) |
|---|---|---|---|---|---|---|---|
| Failed | -0.064*** | -0.052*** | -0.069** | -0.081** | -0.060*** | -0.021 | -0.028** |
| | (0.013) | (0.012) | (0.034) | (0.041) | (0.013) | (0.014) | (0.014) |
| Average Outcome | 0.74 | 0.72 | 0.62 | 0.76 | 0.73 | 0.60 | 0.40 |
| Years | 1995-2021 | 1995-2021 | 1995-2021 | 1995-2021 | 1995-2021 | 1995-2021 | 1995-2021 |
| Fixed Effects | Year, Test | Year, Test | Year, Test | Year, Test | Year, Test | Year, Test | Year, Test |
| Bandwidth | (-0.54,1.05) | (-0.73,1.12) | (-0.59,0.78) | (-0.72,0.98) | (-0.69,1.04) | (-0.71,1.18) | (-0.64,0.97) |
| N | 41638 | 46459 | 5618 | 4433 | 37350 | 47551 | 40549 |

Notes: This table presents estimates of the effects of failing Praxis II scores on different outcomes. Each column presents estimates of equation (1) using the methodology proposed by Calonico et al. (2014) and a triangular kernel. Each regression includes year and test fixed effects. Standard errors are presented in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

## A.2 Teacher Value-Added

As it is common practice in the value-added literature (Kane and Staiger, 2008; Chetty et al., 2014b; Jackson, 2018; Bacher-Hicks et al., 2019a), we generate empirical Bayes shrunken estimates of $\hat{\phi}_j$ to account for sampling error and minimize mean square prediction errors. We construct residuals $\hat{\xi}_{ijst}$ from equation (4) and assume these can be decomposed into a component attributable to teachers ($\phi_j$), classroom-level shocks ($\theta_{ct}$), and student-level idiosyncratic error ($\epsilon_{ijst}$). Using these variance components, we generate empirical Bayes shrunken estimates of teacher effects following Kane and Staiger (2008). Specifically, we multiply the weighted average of teacher-level residuals by an estimate of its reliability, which accounts for the number of observations in each classroom cell:

$$\hat{\phi}_j^{EB} = \bar{\xi}_j \times \frac{\hat{\sigma}_\phi^2}{\hat{\sigma}_\phi^2 + \left(\sum_{m_j} \hat{\sigma}_{jt}^2\right)^{-1}} \tag{6}$$

Where:

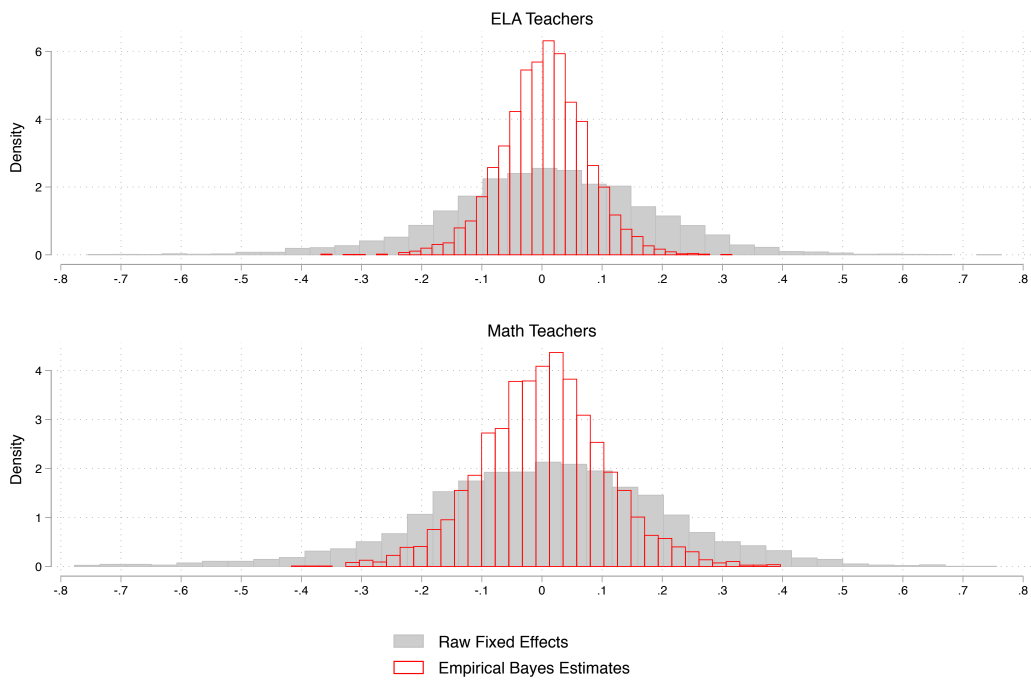
$$\bar{\xi}_j = \sum_t \bar{\xi}_{jt} \times \frac{\hat{\sigma}_{jt}^2}{\sum_l \hat{\sigma}_{jl}^2} \tag{7}$$

$$\hat{\sigma}_{jt}^2 = \left(\hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\xi^2}{N_{cj}}\right)^{-1} \tag{8}$$

In equations (6), (7), and (8), the teacher-level variance $\hat{\sigma}_\phi^2$ corresponds to the covariance in classroom-level average residuals for the same teacher over time $\hat{\sigma}_\phi^2 = cov(\bar{\xi}_{jct}, \bar{\xi}_{jc't'})$. We estimate the student-level idiosyncratic variance $\hat{\sigma}_\epsilon^2$ as the variance in within-classroom deviations in student outcomes. Finally, we estimate the variance of classroom-level shocks as the remainder of the total variation: $\hat{\sigma}_\theta^2 = Var(\xi_{ijst}) - \hat{\sigma}_\phi^2 - \hat{\sigma}_\epsilon^2$.

Figure A.5 shows the distribution of the raw fixed effects ($\hat{\phi}_j$) and the Empirical Bayes estimates ($\hat{\phi}_j^{EB}$) for Math and ELA teachers.

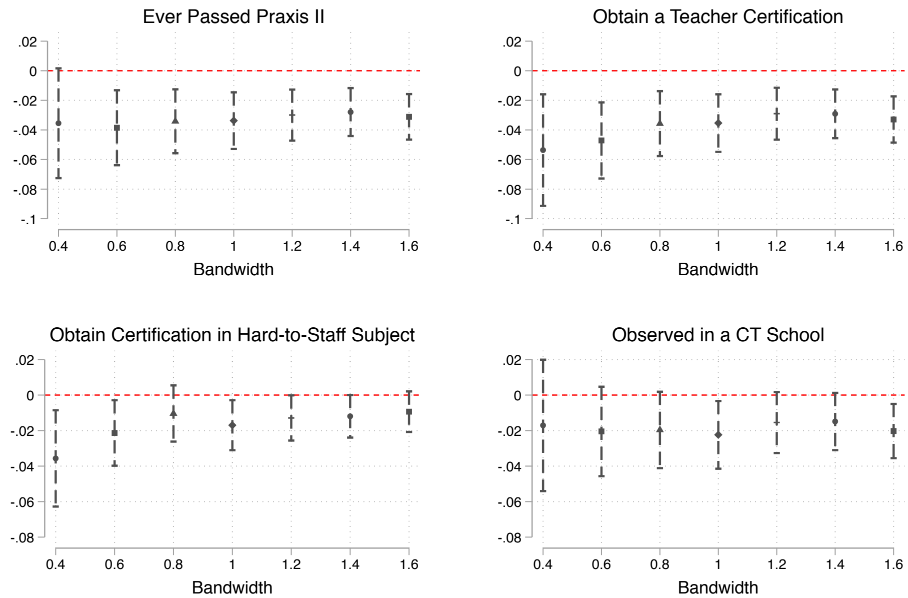
Figure A.5: Distribution of Empirical Bayes Estimates



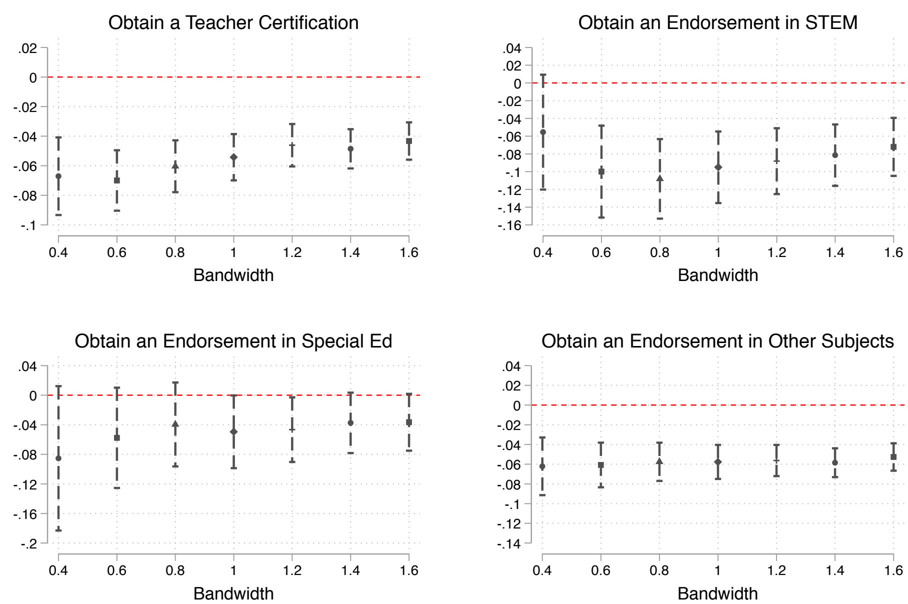

Notes: This figure shows the distribution of raw teacher fixed effects and shrunken empirical Bayes estimates obtained from equation (4). We construct empirical Bayes estimates following Kane and Staiger (2008). See section A.2 for details.

Figure A.6: Different Bandwidths

(a) Effect of Failing Praxis I



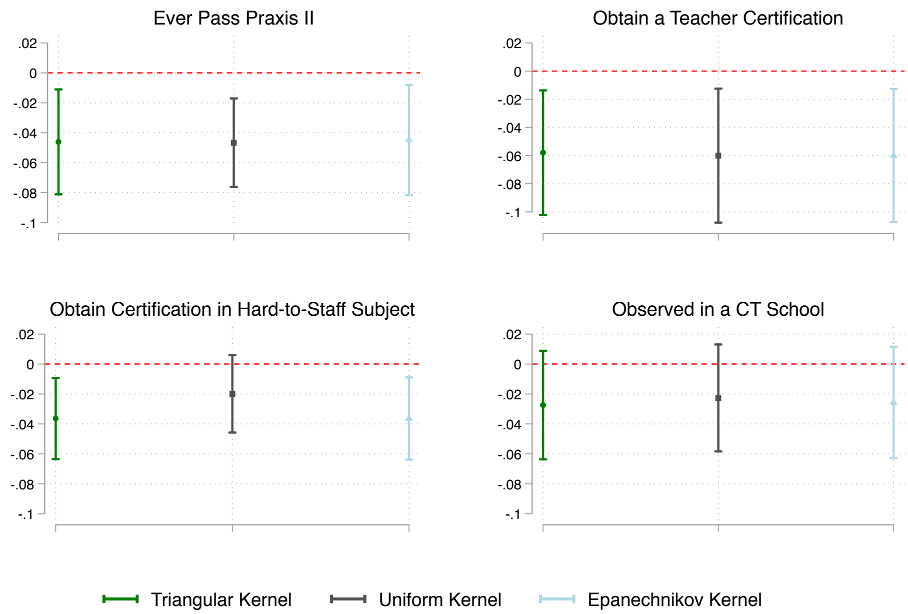

(b) Effect of Failing Praxis II



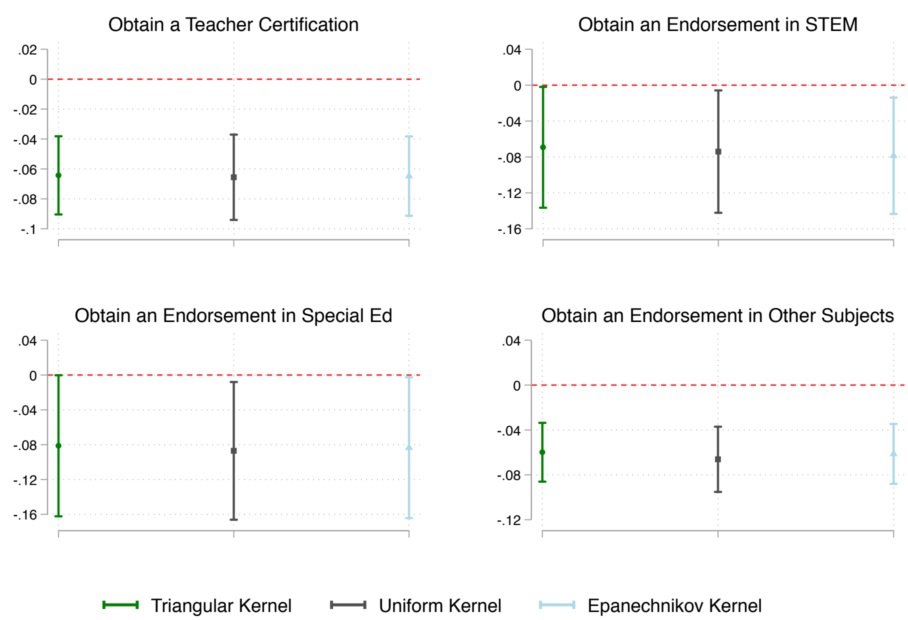

Notes: This figure illustrates how our estimates from Tables 1 and 2 vary with the choice of the bandwidth. The x-axis corresponds to different bandwidths used to build these figures. Each coefficient and their confidence intervals come from estimating equation (1) using the corresponding bandwidth around the cutoff.

Figure A.7: Different Kernels

(a) Effect of Failing Praxis I




(b) Effect of Failing Praxis II




Notes: This Figure illustrates the robustness of our results to the choice of the kernel used in the local linear regressions. Each plot shows the estimation of equation (1) using the methodology proposed by Calonico et al. (2014) under different choices of the kernel. Each point illustrates the estimated effect, and the lines denote the 95% confidence intervals. Panels (a) and (b) shows different outcomes associated with failing the Praxis I and Praxis II tests, respectively.