Licensure Tests and Teacher Supply

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

Abstract

We apply a sharp regression discontinuity design to administrative data from Connecticut to investigate the impact of failing the first attempt at a licensure test on teacher supply. We find deterrent effects from failing both a basic skills test required to enter an educator preparation program (Praxis I) and a subject-matter test used for ultimate certification (Praxis II). Failing Praxis II especially deters those seeking endorsement to teach within the shortage areas of STEM and special education. Failing Praxis II disproportionately pushes out relatively effective potential teachers and those who would have taught in schools with relatively smaller proportions of students who are Black or Hispanic, learning English, and eligible for subsidized lunch.

Acknowledgements: We thank the Connecticut State Department of Education (CSDE), especially 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. We received valuable feedback from Kevin Lang, Dan Goldhaber, Roddy Theobald, James Cowan, Andrew Bacher-Hicks, Olivia Chi, Kirsten Slungaard-Mumma, William Delgado, and participants at the BU-Wheelock PREREQ seminar, BU Economics Applied Microeconomics Seminar, the annual conferences for the Association of Education Finance and Policy, and the Society of Labor Economists. The data can be obtained by filing a request directly with the Connecticut State Department of Education. The authors are willing to assist. Neither author received financial support or has a conflicting interest.

*Alexis Orellana, Ph.D. is a Lead Economist at the School of Education and Social Policy at Northwestern University. boris.orellana@northwestern.edu

†Marcus A. Winters, Ph.D. is a Professor of Educational Leadership & Policy Studies at the Boston University Wheelock College of Education & Human Development and faculty director of the Wheelock Educational Policy Center. marcusw@bu.edu

doi:10.3368/jhr.0324-13465R2

1 Introduction

Recruiting effective teachers is a key element to producing high-quality public schools. In order to ensure that teachers meet at least a minimum competency standard, states typically require candidates to pass one or more tests as part of the certification process. But the extent to which licensure test requirements impact the size and quality of the eligible teacher workforce is unclear. Stringent licensure requirements could simultaneously push-out ineffective teachers by posing a disproportionately large barrier for them to obtain certification, and pull-out high-quality candidates by increasing the cost to becoming a teacher relative to other professions (Larsen et al., 2022). Further, evidence that a teacher's licensure test scores are at best modestly correlated with their later effectiveness (Clotfelter et al., 2006, 2007, 2010; Goldhaber et al., 2017) calls into question their value as a screen relative to the challenges they pose for filling persistent staffing shortages (Boe and Cook, 2006; Goldhaber et al., 2015; Cowan et al., 2016; Dee and Goldhaber, 2017; McVey and Trinidad, 2019).

One way that licensure tests could impact the size and quality of the teacher workforce is by deterring those who fail their first attempt from continuing to pursue teaching as a profession. We apply a sharp regression discontinuity design to more than two decades of administrative data from Connecticut to investigate the potential that failing the first attempt at a licensure test deters candidates from the pathway to teaching. We separately consider the deterrent effect from failing a basic skills test used as a requirement for entering an endorsed educator preparation program (Praxis I), and a subject-matter test typically administered near the completion of an educator preparation program and used as a screen for certification or endorsement to teach within a particular area (Praxis II).

We first demonstrate that failing the initial attempt at a licensure test significantly alters a candidate's pathway to teaching. Relative to the average test-taker, failing the first attempt at Praxis I reduces the likelihood of eventually obtaining any teaching certification by about 11.2% relative to the average licensure-test-taker and reduces the probability they eventually teach within a Connecticut public school by 8.3%. The deterrent effect from failing Praxis II differs substan-

tially by the subject area in which the candidate is seeking endorsement. Failing the first attempt at a licensure test necessary to obtain an endorsement to teach within the key shortage areas of STEM and special education reduces the likelihood of teaching within that area by 17% and 19%, respectively, while failing the first attempt at a test necessary to acquire a license to become a general elementary grade teacher does not impact the likelihood of obtaining a credential or becoming a teacher.

We further identify interesting differences in how failing Praxis II in different subject areas alters a potential teacher's pathway to the profession. Those who were deterred by failing a test to obtain endorsement to teach a STEM subject tend to not acquire a license to teach in another area and leave the profession entirely. In contrast, those deterred by failing a test for endorsement to teach special education typically obtain a license and are employed as a general elementary teacher.

We then present new evidence on the extent to which failing the first attempt at a licensure test disproportionately deters candidates with higher or lower latent value-added, by which we mean the contribution to student test scores they would make if they were to become a teacher in the state.[1] Prior studies measuring the correlation between licensure score and value-added are limited by the fact that the researchers do not observe the performance of licensure-test-takers who never enter the classroom (Clotfelter et al., 2006, 2007, 2010; Goldhaber et al., 2017). We address this challenge by measuring the discontinuity at the passing threshold in the relationship between initial licensure score and the value-added scores for those we observe as teachers. The intuition underlying our approach is the same as a conventional balance test: For the population of those who take a licensure test, the distribution of latent value-added should be smooth at the passing threshold, and thus a discontinuity in *observed* value-added at the passing threshold reflects treatment-induced differences in the relationship between an individual's *latent* value-added and the likelihood that we observe their value-added score.

We find suggestive evidence that failing the first attempt at Praxis II disproportionately pushes out those who would have been relatively more effective teachers had they entered the classroom. In both math and English Language Arts (ELA) we find a discontinuous decline in

average observed value-added score of about 0.02σ on the failing side of the threshold, though the estimate is statistically significant only for ELA. Nonetheless, taken together the pattern of results across both subjects suggests that the deterrent effect from failing a licensure test does not differentially push out lower-performing teachers but may especially discourage higher-quality prospects. Applying this strategy, we also find evidence that failing the first attempt at Praxis II disproportionately deters those who otherwise would have taught in schools with relatively smaller percentages of students who are Black or Hispanic, learning English, and eligible for subsidized lunch.

Our results have broad policy relevance that extend beyond the context of Connecticut. All states require prospective teachers to pass some form of licensure test to gain certification or endorsement, and 15 states currently require candidates to pass basic skills tests to gain admission into a teacher preparation program (Putman and Walsh, 2021). Half of all states use the Praxis series of tests and apply the same or very similar cutoffs as Connecticut.

Although Connecticut is a relatively small New England state, its teacher workforce characteristics and the challenges in filling vacancies mirror those faced by school systems nationwide. Connecticut's teacher workforce aligns closely with the national average across several key indicators, including starting salary, teacher attrition, recent changes in student and teacher populations, and the proportion of schools reporting hard-to-fill vacancies.[2] Further, five out of the seven subjects that Connecticut identifies as shortage areas were also reported as such by more than half of all states.[3]

Our estimates by subject area are unique within the literature and highly relevant for policy. Prior studies rely on aggregated data that does not distinguish the subject area in which the candidate is seeking endorsement to investigate the relationship between licensure testing and the characteristics of the overall teacher workforce (Hanushek and Pace, 1995; Angrist and Guryan, 2004, 2008; Larsen et al., 2022). Our finding that the effect of failing Praxis II varies substantially by the subject in which the candidate is seeking endorsement suggests that analyses of the impact of teacher certification requirements in the aggregate could miss important impacts within key subject areas. From a policy perspective, separately considering impacts by subject area is highly

important because administrators consistently struggle to find teachers with the necessary credentials to staff classrooms only 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).

We add to a limited body of research measuring the relationship between the stringency of licensure requirements and teacher supply. 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 limits the ability to give their results a causal interpretation. Analyses leveraging within-state variation in the stringency of licensure requirements over time find no significant impact on the selectivity of undergraduate institutions from which the average teacher is drawn (Angrist and Guryan, 2004, 2008; Larsen et al., 2022). Larsen et al. (2022) further find that strengthening academic coursework requirements for teacher licensure was associated with a significant increase in the bottom decile and a statistically insignificant decrease in the top decile of the selectivity of undergraduate institution for entering teachers. We contribute new estimates for the causal effect of a binding licensure test requirement on the size and quality of the entering teacher workforce derived by applying a different compelling identification strategy to administrative data in which we can measure differential effects on a direct measure of teacher value-added contributions to student test scores.

Our findings are also relevant for understanding the relationship between licensure test scores and teacher quality. In particular, our results suggest that prior estimates for the correlation between a teacher's licensure test score and their value-added impact on student outcomes could suffer from selection bias due to differences in the relationship between latent value-added and the likelihood of observing value-added among those who passed or failed their initial attempt (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).

We also build upon prior 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 valuing 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). We 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). Our finding that failing the first attempt at a licensure test deters some prospective teachers is consistent with recent studies reporting 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). Prior studies tend to find little positive effect from licensure requirements on measures of average workforce quality (Carroll and Gaston, 1981; Kleiner and Kudrle, 2000; Kugler and Sauer, 2005; Hall et al., 2019; Kleiner and Soltas, 2019; Farronato et al., 2020), though a recent study by Anderson et al. (2020) finding positive impacts from licensing for midwives on maternal mortality in the early 1900's is an important recent exception. And there is some evidence that licensure requirements can raise the floor for workforce quality (Ramseyer and Rasmusen, 2015; Bhattacharya et al., 2019; Larsen et al., 2022). We build upon this work on the link between licensure and workforce quality with new evidence showing that failing the first attempt at a licensure test has a larger deterrent effect on more promising prospective teachers. Finally, our finding that failing a licensure test differentially pushes out prospective STEM teachers raises questions about how licensure may especially amplify shortages in high-skill areas with high-salary outside options that are worthy of future consideration.

2 Data

2.1 Licensure and Licensure Tests

Similar to other states, in Connecticut the traditional certification process requires an applicant to complete a state-approved educator preparation program and pass subject-specific tests associated with the area of specialization in which they seek an endorsement. During our sample period, the state employed tests related to both entry into an educator preparation program and subject-matter endorsement, 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.

During 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. However, schools do not typically observe an applicant's specific licensure test score(s) or information about the number of attempts the candidate required to pass.

The state allows public schools to hire on a one-year permit teachers who have not fulfilled the proper licensure requirements in order to fill an urgent staffing need. However, Connecticut's requirements for hiring teachers under such a temporary emergency license are more stringent than is typical in other states. In addition to holding a bachelor's degree, the individual must have completed at least 12 credit hours of coursework in the relevant subject area and either be currently enrolled or demonstrate they plan to enroll in an educator preparation program. Per-

haps as a consequence of these more stringent standards, though the proportion of schools in the state reporting unfilled or hard-to-fill teaching vacancies is similar to the national average, only 1.3% of Connecticut public school teachers are not properly certified for their position compared to 3.7% nationwide.[4]

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.[5] All of our analyses for Praxis I include only tests administered prior to 2016.

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 subtests that make up the Praxis I assessment have also changed over time, as illustrated in Figure A.2 in the online Appendix.

Praxis requirements used in Connecticut are similar to those used in other states. Table A.3 in the online Appendix shows Praxis I passing scores used by Connecticut before 2016 and summary statistics of current passing scores in other states.[6] Currently, 25 states use the three Praxis I subtests and the large majority of them use the same passing score. In 2019, the Praxis I subtests were replaced by new versions (subtests 5713, 5723, and 5733). For this reason, we also look at differences using the subtests valid in 2016. All states, including Connecticut, used the same passing scores as shown in Table A.3, with the exception of Washington and North Dakota.[7]

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. 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 endorsement. Table A.4 in the online Appendix shows the link between some of the endorsement codes offered in Connecticut and the associated Praxis II tests. Some endorsement codes involve passing more than one test (for example, *Elementary Grades, K-6*). In these cases, we group all subtests and employ the minimum score as the forcing variable in the analysis described in Section 3.[8]

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.

Score requirements used in Connecticut are similar to those in other states. Table A.5 in the online Appendix shows passing scores for all tests used in Connecticut alongside summary statistics of passing scores of the same tests in other states. Taken together, Tables A.3 and A.5 show that states do not vary substantially in terms of their test score requirements. In addition, these cutoffs typically remain fixed over time. In the case of Praxis II, in 20 out of the 26 tests, the passing score in Connecticut is the same as the modal passing score across other states.

2.2 Certification Data

We link applicants' scores on licensure tests to Connecticut's certification data between 2002 and 2021. For each person who applied to the state for certification and/or endorsement, these records contain the certificate type, 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 addition, these data also include basic demographic information for those applying for certification, includ-

ing the candidate's race/ethnicity and gender.

For our analyses, we define a certified teacher as one who has obtained a renewable Initial or Provisional Educator Certificate.[9] 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. Since 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, we consider it to be a reasonable proxy for seeking a teaching position.

We separately distinguish those who teach on a nonrenewable Interim Educator Certificate or permit to teach within a shortage area.[10] Though all teaching within a Connecticut public school should have one of these certification types, we observe a small number of teachers with valid initial licensure scores who we do not match to a license.

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.

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 ELA 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 in this sub-sample to a single classroom teacher.

2.5 Summarizing the Pathway to Becoming a Teacher

Table 1 describes differences in the pathway toward becoming a teacher related to initial licensure test score and certification status. The top panel reports results for Praxis I and the bottom panel reports results for Praxis II.

Those who fail their first attempt at Praxis I are about 19 percentage points less likely to eventually pass Praxis II, and 11.5 percentage points less likely to teach within a Connecticut public school than those who passed. Among those who eventually gain certification, there is a statistically significant but insubstantial difference in the likelihood of teaching between those who passed or failed their first attempt at the basic skills test.

The pattern of descriptive results among Praxis II test takers is similar to the results from Praxis I, though the differences between those who passed or failed their initial attempt are smaller. Though those who pass Praxis II on their first attempt are more likely to obtain certification and teach within the state, most who fail the first administration continue on the pathway to teaching. However, re-take rates on Praxis II differ notably by subject area. Among those who fail their first attempt, 66% and 72% retake a test if their initial test was in a STEM subject or special education, respectively, compared to 81% in other subjects. Relative to those who pass Praxis II on their first

attempt, those who pass on a retest have similar trajectories but are 2 percentage points more likely to be observed as a teacher. Those who fail the first administration of Praxis II are slightly more likely to teach on an interim or emergency certification than those who passed their first attempt, but very few candidates enter the classroom in this way. Finally, among test-takers who eventually gain certification, those who failed the first administration of Praxis II are about 3 percentage points more likely to be observed as a public school teacher.

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

3.1 Empirical Strategy

In this section, our goal is to estimate the causal effect of an individual failing their first attempt on a licensure test on their pathway to becoming a teacher. A naive comparison is likely biased by unobserved differences related to the likelihood of failing and one's trajectory towards becoming a public school teacher. We overcome this challenge by leveraging the sharp discontinuity in passing 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.[11] 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:

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

The term $f(x_{ij})$ is a parametric function of the (normalized) score obtained by applicant i, which our primary model employs as linear and allows for changes in the slope at the cutoff value.[12] The indicator $\mathbb{1}(x_{ij} < 0)$ is our variable of interest and represents the discontinuity at the threshold value. We account for changes in the tests over time and differences across subject-area tests by

including fixed effects for year (ϕ_t) and specific test subject administered (ϕ_j) to the prospective teacher. We estimate local linear regressions to observations that fall within optimal bandwidths from the cutoff as calculated using the methodology of Calonico et al. (2014) (hereafter, CCT). Our primary results are from models that employ a triangular kernel, and we report results that use a uniform kernel in the online Appendix. The sample includes an individual's first observed score on the relevant licensure test, excluding first-time test-takers who we observe teaching within a Connecticut public school in a prior year. This latter exclusion should account for current teachers whose first attempt took place in a year prior to our data beginning.[13]

The key identifying assumption for β is that the relationship between a candidate's score and the outcome would be smooth at the passing threshold if not for the fact that scoring above the line satisfied the passing requirement. There are two particular threats to this assumption. The first is the potential for individuals to manipulate their scores around the cutoff. The institutional features of the certification process in Connecticut make violating this assumption unlikely. Figure A.4 in the online Appendix shows the distribution of Praxis I test scores between 1995 and 2021. We present densities separately because not all Praxis I tests have the same scale. These histograms show no indication of manipulation around the cutoff values. We formally test the existence of discontinuities around the cutoff by implementing the test proposed by Cattaneo et al. (2018). Figure 1 shows no statistical evidence to reject the null hypothesis of continuity around the passing threshold. The p-value of the discontinuity test for Praxis I and Praxis II is 0.54 and 0.35, respectively.

The second threat to identification is the potential for discontinuities in the value for confounders around the threshold. To investigate the potential for this threat, authors typically look for balance in the value of observed baseline characteristics on either side of the threshold. Unfortunately, conventional balance tests are not available to us because we observe demographic information only for individuals who apply for a certification or endorsement. Nonetheless, given the nature of the tests we argue that it is highly unlikely for the characteristics of test-takers would differ systematically at the passing threshold.[14]

3.2 Results

Figures 2 and 3 illustrate non-parametric estimates for the relationship between the score obtained on Praxis I or Praxis II and select outcomes. For both tests we illustrate the likelihood of obtaining any certification and of teaching within a Connecticut public school. In Figure 2, for Praxis I we also illustrate the relationship for the likelihood of eventually obtaining a certification in a hard-to-staff subject. For Praxis II we illustrate the relationship between initial score and obtaining an endorsement to teach within STEM, special education, and other subjects, separately. These analyses by subject-area are restricted to the first administration of the test required to obtain the associated endorsement.

In all but one figure we observe a distinct discontinuity in the value for the respective outcome at the threshold, which is indicative of a causal treatment effect. However, consistent with the summary statistics reported on Table 1, the figures also make apparent that many who fail the first administration of the licensure tests nonetheless persist on the pathway to becoming a teacher.

[FIGURE 2 ABOUT HERE]

[FIGURE 3 ABOUT HERE]

Table 2 reports regression discontinuity estimates for the effect of failing the first administration of a licensure test on certification and the likelihood of teaching within a Connecticut public school. The results reported in the top panel indicate that failing the first administration of the Praxis I basic-skills test required for entry into an educator preparation program reduced the likelihood the test-taker later obtained any teaching certification within the state by about 6.7 percentage points (ppts), or 11.2% relative to the average test-taker within the sample. Failing the first attempt at Praxis I reduced the likelihood of eventually becoming a Connecticut public school teacher by 3.8 ppts (8.3%).[15] The estimates reported in Columns (1) and (2) suggest that failing Praxis I reduces the likelihood that a test-taker continues on the pathway towards becoming a teacher to the point that they take the Praxis II test.

[TABLE 2 ABOUT HERE]

The bottom panel reports regression discontinuity estimates for the effect of failing the first attempt at Praxis II within the full sample that includes all test subjects. Failing the first attempt at Praxis II reduced the likelihood that the test-taker eventually obtained some type of teaching certification by about 6.6 ppts (8.4%) and the likelihood of teaching in a Connecticut public school by about 2.1 ppts (3.5%), but the estimate is not statistically significant.[16]

We extend our analyses to investigate the effect of failing the first attempt at a licensure test on the subject area in which a candidate obtains a license and ultimately teaches. Table 3 presents our results of failing the first administration of a Praxis I test on the likelihood of obtaining certification and teaching in each subject area.[17] To make our interpretation comparable, for each outcome we report all estimates using the same bandwidth. Failing the first administration of Praxis I decreases the probability of obtaining a certification in STEM by 3.3 ppts and of teaching within a STEM subject by about 3.0 ppts, but does not impact the probability of teaching in areas such as special education, ESOL, English, or History. By contrast, failing Praxis I made a candidate about 2.1 ppts more likely to teach in Art or Music.

[TABLE 3 ABOUT HERE]

Tables 4, 5, and 6 report the results from analyses measuring the deterrent effect of failing the first administration of the Praxis II test associated with endorsement to teach within a STEM subject, special education, and elementary education, respectively. These analyses provide further context for how failing the test impacts the career pathways of prospective teachers by measuring whether and the extent to which those who fail a test for endorsement in one area instead obtain an endorsement and/or teach within a different subject area or leave the profession entirely. Our results suggest that the deterrent effect of failing Praxis II varies considerably by the subject area in which the candidate is seeking endorsement.

We first consider the results for the effect of failing the initial licensure test to teach within the shortage areas of STEM and special education. Failing the first attempt at a STEM-subject licensure test reduced the likelihood of obtaining a license to teach within STEM by about 10.7 ppts

(16%), and failing the first attempt at a special education licensure test reduced the likelihood of obtaining a special education license by about 10.6 ppts (13%). However, the effect of failing the test within these respective areas had different impacts on the candidate's later professional trajectory. Those who failed the test for endorsement in STEM typically did not obtain an endorsement in another subject, but rather tended to leave teaching entirely. Failing the first attempt at a STEM licensure test increased the likelihood that the candidate was not observed as a teacher within the state by about 8.7 ppts (24%).

[TABLE 4 ABOUT HERE]

The results reported in Table 5 suggest that those who were deterred from obtaining a special education endorsement and teach within special education because they failed the first attempt at the associated licensure test were more likely to take a test within a different subject area and obtain employment as a general elementary teacher. Failing the first attempt at a special education licensure test reduced the likelihood of teaching within special education by about 11.1 ppts (19%) and increased the likelihood of teaching within general elementary by about 8.5 ppts (70%). We also find some suggestive evidence that failing a special education licensure test increases the likelihood that the prospect is not later observed as a teacher, but this result is estimated imprecisely and is not statistically significant at any conventional level.

[TABLE 5 ABOUT HERE]

In contrast, the results reported in Table 6 find that failing a test for a license to become a general elementary teacher reduces the likelihood of obtaining that endorsement by only about 3.5 ppts (4%) and has no significant effect on the likelihood that they become a teacher or more specifically of becoming a general elementary teacher in the state.

[TABLE 6 ABOUT HERE]

We present several robustness checks and additional results to our main analysis in the online Appendix. Figures A.6 and A.7 show that our estimates are robust to the choice of different

bandwidths, while Figure A.8 and Table A.8 present our estimates using a uniform kernel instead of a triangular kernel.

Table A.9 presents additional analyses on longevity and the likelihood that they teach in a tested grade and subject, allowing us to observe their value-added score, among licensure-test-takers who we observe as a teacher in the data. We find suggestive evidence that among prospects who become teachers, those who failed their first attempt at Praxis I are more likely to attrit early in their career and are less likely to have an observed value-added score. For Praxis II, our estimates for longevity of those who become teachers are smaller and, with the exception of one case, not statistically significant. However, among those who become teachers, we find no difference in the likelihood we observe value-added for those who passed or failed their first attempt at Praxis II.

Finally, Tables A.11 and A.12 present estimates of the deterrent effect of failing the first administration of the Praxis II test in the English and Art-Music subjects, respectively.

4 Investigating Differential Deterrent Effects by Latent Quality and First School

In this section our goal is to investigate the extent to which failing the first administration of a licensure test differentially pushes out more or less effective potential teachers and teachers who would have taught in schools with particular characteristics. Our primary challenge is that we observe value-added and school assignments only for those who are employed as a public school teacher. Our strategy to address this challenge is to use differences in observed value-added on either side of the passing threshold to infer differences in the latent value-added of those who were or were not deterred by failing their initial attempt at the licensure test. To ease the discussion, in what follows we focus on describing our analysis and interpretation within the context of differences in the individual's latent value-added. We then apply the same logic to investigate differences in the characteristics of the school in which the individual would have taught had they not failed their first attempt the licensure test.

We define a prospect's latent value-added, θ_i^*, as the independent contribution that the

teacher would have on student test scores if they were to become a teacher. For those whose performance we observe, we assume that calculated value-added, θ_i, is an unbiased estimate for θ_i^*.[18] Further, we assume that θ_i^* is fixed at the time of their first attempt at a licensure test. This last assumption would be violated, for example, if possession of a license itself improved a prospective teacher's ability to affect student outcomes, or if preparing to pass a retake systematically improved their inherent potential. Though we cannot directly test its validity, we argue that this assumption is justified by the nature of the tests, the availability of multiple retakes, and that licensure scores correlate weakly with the observed value-added of those who eventually become teachers.[19]

In essence, we treat θ_i^* as a fixed baseline characteristic possessed by all licensure-test-takers. The key identifying assumption of a regression discontinuity design thus implies that the value for θ_i^* should be smooth at the passing threshold. That is, if we were able to estimate a model taking the form:

$$\theta_i^* = \tau + g(x_{ij}) + \lambda \mathbb{1}(x_{ij} < 0) + \psi_j + \psi_t + \eta_{ij} \tag{2}$$

we would expect to find $\lambda = 0$, implying that θ_i^* is balanced at the threshold.

However, we do not directly observe θ_i^*. Thus, we must instead estimate a model that uses θ_i as the outcome.

$$\theta_i = \tau + g(x_{ij}) + \lambda \mathbb{1}(x_{ij} < 0) + \psi_j + \psi_t + \eta_{ij} \tag{3}$$

Notice that because it is an unbiased estimate for θ_i^*, if we were able to calculate value-added for all licensure test-takers we would similarly expect to find $\lambda = 0$. That is, absent differential selection in the quality of those whose value-added we observe, estimated value-added should also balance at the threshold.

In practice, however, we only observe θ_i for a subset of licensure test-takers, and thus the estimation sample is necessarily restricted to only those with an observed value-added score. Our above assumptions imply that failing the test does not directly impact θ_i. However, we would observe a discontinuity in the value of θ_i at the threshold if failing the initial licensure test reduced

the likelihood of observing the value-added for those test-takers with higher/lower θ_i^*. That is, as in a conventional balance test, a finding that $\lambda \neq 0$ would imply selectivity associated with latent value-added occurring on the failing side of the threshold. However, what is worrisome selection bias in a typical application of a regression discontinuity design in our case has a policy-relevant interpretation. Namely, the existence of a discontinuity in observed value-added at the threshold suggests that failing the test disproportionately pushed out more/less effective prospective teachers before they entered the classroom, making it less likely that their value-added was observed.

It is important to note that when considering differences in latent value-added the estimation sample for this exercise is not representative of the population of licensure test-takers. Considering the grades and subjects tested, these sub-samples contain a larger fraction of elementary, English, and mathematics teachers relative to our full sample.[20] Models that apply this strategy to analyze the impacts of failing Praxis I are underpowered and thus we focus on measuring the effect of failing the first attempt at Praxis II.

4.1 Results

Figure 4 illustrates the relationship between initial score on Praxis II and average observed value-added in ELA and mathematics. The results reported in the first row ("Failed") of Table 7 coincide with the respective figures. For both subjects we find a discontinuous drop of about -0.02σ in average observed value-added scores on the failing side of the threshold, implying that the treatment disproportionately reduced the likelihood we observe value-added scores for test-takers with higher latent value-added. The estimate is statistically significant in ELA, but not in math (p-value = 0.15).

[FIGURE 4 ABOUT HERE]

Figure 5 illustrates the relationship between initial Praxis II score and select demographic characteristics of the school in which the individual was first employed. The sample for these regressions is larger than the value-added analyses because they include all licensure test-takers we observe teaching within a Connecticut public school, not only those who also teach in a tested grade

and subject. The pattern of results suggest that on average those who are deterred from teaching by failing the first attempt at Praxis II would have taught in schools with smaller percentages of Black and Hispanic students, English learners, and students eligible for free or reduced priced lunch. That is, failing the first attempt at Praxis II differentially pushes out those who would have otherwise taught in schools with more advantaged student populations, on average.

[FIGURE 5 ABOUT HERE]

[TABLE 7 ABOUT HERE]

There are two potential mechanisms by which increased attrition due to failing the first attempt at a licensure test could produce differential selection consistent with the patterns described on Figures 4 and 5.[21] First, there could be an interactive treatment effect such that test-takers with higher latent value-added or aiming to teach in a higher-SES school are more greatly deterred by failing the test. Second, we detail in online Appendix C that a uniform treatment effect could also produce differential selection by exacerbating a pre-existing relationship between the characteristic of interest (i.e., latent value-added or type of school in which they would be first employed), initial licensure score, and the likelihood that a test-taker would complete the pathway to becoming a teacher independent of whether they fail the test. In short, the potential for differential selection despite a homogeneous treatment effect can occur if there is residual variation in the candidate's characteristic of interest that is correlated with the likelihood that the candidate would become a teacher independent of the policy, what we might call "natural attrition." Though it is unfortunate that our analysis cannot disentangle the relative impacts of these mechanisms, it is the overall treatment effect that is policy relevant.

Our finding that failing the test differentially pushes out those who would have taught in higher-SES schools is perhaps unexpected in light of concern that licensure tests could serve as a disproportionately large barrier to entry for prospective teachers from lower-income families and who identify as part of a minority subgroup, and we expect that such candidates would be more likely to teach within lower-SES schools. One potential explanation is that this pattern could reflect a homogeneous treatment effect that exacerbates a pre-existing positive correlation between the

likelihood of teaching within a higher-SES school and both licensure score and "natural attrition". In addition, it is possible that the likelihood of teaching in a higher-SES school is correlated with latent value-added, and thus if the deterrent effect is primarily driven by differentially pushing out candidates who have more attractive alternative labor market opportunities, then we would also expect to find differential selection of those who would have taught in higher-SES schools.

Finally, we attempt to quantify the relative latent value-added and school characteristics for those who are deterred from teaching because they failed their initial attempt at a licensure test. The difference in average observed value-added and school characteristics at the threshold reported in the first row of Table 7 is a weighted average reflecting both the effect of failing on the likelihood of becoming a teacher and the difference in characteristics of those who are deterred from teaching and those who nonetheless enter the classroom. Thus, to quantify the differences on these characteristics between those who were or were not deterred by failing the test we must re-scale these differences in average observed values at the threshold by the proportion of failers who were deterred.

It is clarifying to consider this correction within the framework of a local instrumental variables approach where failing the first administration of a licensure test serves as an instrument for an aspect of the teaching pipeline and latent value-added is the outcome of interest. We can classify the pool of eligible test-taking teachers into one of three categories: Always-Teachers persist to teach even if they fail their first administration of the test, Never-Teachers do not teach regardless of whether they pass or fail, and Compliers become teachers if they pass but not if they fail. For those who scored above the passing threshold we observe the value-added for both Always-Teachers and Compliers, while among those who scored below the threshold we observe value-added only for Always-Teachers. In this framework, Equation (3) represents the reduced-form, measuring the difference in average observed teacher quality among those who passed or failed the test. Identifying the difference in the latent value-added of Always-Teachers and Compliers requires dividing this reduced-form estimate by the Complier share, which we acquire from our results for the effect of failing the first administration of a test on a given stage of the teacher pipeline as described in Equation (1).

As a practical matter, we cannot directly estimate the IV model because Compliers on the failing side of the threshold are missing the outcome and thus conventional software excludes them from the analysis. We generate appropriate standard errors by bootstrapping the procedure. Specifically, we estimate the reduced-form and first-stage within 1,000 randomly drawn (with replacement) samples from the full data. The average for the ratio of the reduced-form and first-stage across these iterations is the bootstrapped estimate for the magnitude and the standard deviation is the bootstrapped standard error.

The second row ("IV") of Table 7 reports our estimates quantifying the difference in latent value-added between Always-Teachers and Compliers. Adjusting the reduced-form by the relatively weak first-stage leads to very large but very imprecisely measured IV coefficients. Our results suggest that the small proportion of prospective teachers who are deterred by failing their first attempt at Praxis II have substantially higher (about a third of a standard deviation) latent value-added in both ELA and math and would have taught in schools with much higher percentages of English learners and those eligible for subsidized lunch than the group of Always-Teachers. However, we caution that the analysis cannot rule out that these differences between Always-Teachers and Compliers are much smaller than represented by the coefficients. We suggest that readers interpret these results as suggestive evidence.

5 Discussion and Conclusion

Applying a regression discontinuity strategy to administrative data from Connecticut, we find evidence that failing the first attempt at required licensure tests at different points on the teacher preparation pipeline serves as a barrier to entering the teaching profession that can have unintended consequences for the teacher workforce.

We present unique evidence for how the deterrent effect of failing a licensure standard impacts the pathway to becoming a teacher overall and by subject area. Failing a licensure test does not deter candidates who are pursuing a general teaching license, but presents a barrier to teaching within STEM and special education. Deterrence from failing subject-matter licensure tests thus

contributes to the persistent staffing shortages in these areas.

Our finding that failing the first attempt at the associated licensure test tends to push prospective STEM teachers out of the profession but instead pushes prospective special educators to teach in other areas is worthy of future research. One possible explanation for this result is that candidates who are interested in teaching STEM subjects may have relatively more attractive opportunities in the outside labor market that they are more willing to pursue if they face an obstacle to becoming a teacher, while those pursuing an endorsement to teach special education are less interested in career options other than teaching. In the case of special education, schools may consider adopting strategies to encourage and support general education teachers who previously signaled an interest in teaching students with disabilities by attempting to obtain the relevant license to retake the test.

We also contribute new evidence for the relative quality of prospective teachers who are deterred by failing their first attempt at a licensure test. We leverage the underlying assumptions of a regression discontinuity analysis to present a unique set of results that effectively measure the value-added impact that teachers who were deterred by failing the test would have achieved had they entered the classroom. We find evidence that failing the first attempt at the Praxis II subject-matter test required to obtain a certification tends to push out relatively higher-quality potential teachers. We emphasize that these analyses are estimated imprecisely, and thus should be considered suggestive. Nonetheless, the fact that our analyses can effectively rule out that failing a licensure test differentially deters the lower-quality prospective teachers that licensure policies intend to weed out from the system is highly relevant to policy discussions. Further, these findings suggest the need for future research to better understand the extent to which more promising teachers have more attractive career alternatives that can pull them from the profession.

From a policy perspective, our results suggest potential benefits from loosening the use of licensure tests as a restriction to teaching, especially within key shortage areas. For example, states could allow those seeking a position to teach in special education or a STEM subject to pass the respective licensure tests within two years of entering the classroom – effectively converting the test from a gatekeeper into a timed benchmark that candidates must meet once supported by

real-world practice. Increasing flexibility in key shortage areas could focus relief precisely where the deterrent effect is largest, and our results suggest would differentially retain the more promising prospective teachers.

Finally, it is important to keep in mind that the deterrent effect from failing the first attempt at a licensure test that is the focus of our analysis is just one component of modern teacher licensure requirements. For example, the very existence of a testing requirement may affect the pool of individuals pursuing a teaching career. Further, to obtain certification, candidates must also complete an approved educator preparation program, and Larsen et al. (2022) find evidence that stricter coursework requirements increase the lower tail of the teacher quality distribution. Thus, though our results raise important questions about the common use of licensure tests as a barrier to teaching, especially for teaching within key shortage areas, it remains possible that the licensure process overall could benefit the teaching profession.

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.

Anderson, D. M., Brown, R., Charles, K. K., and Rees, D. I. (2020). Occupational Licensing and Maternal Health: Evidence from Early Midwifery Laws. *Journal of Political Economy*, 128(11):4337–4383.

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

Angrist, J. D. and Guryan, J. (2008). Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements. *Economics of Education Review*, 27(5):483–503.

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., 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.

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.

Bhattacharya, V., Illanes, G., and Padi, M. (2019). Fiduciary Duty and the Market for Financial Advice. Working Paper 25861, National Bureau of Economic Research.

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.

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.

Carroll, S. L. and Gaston, R. J. (1981). Occupational Restrictions and the Quality of Service Received: Some Evidence. *Southern Economic Journal*, 47(4):959–976.

Cattaneo, M., Jansson, M., and Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *Stata Journal*, 18(1):234–261.

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 Paper 22344.

Falch, T., Johansen, K., and Strøm, B. (2009). Teacher Shortages and the Business Cycle. *Labour Economics*, 16(6):648–658.

Farronato, C., Fradkin, A., Larsen, B., and Brynjolfsson, E. (2020). Consumer Protection in an Online World: An Analysis of Occupational Licensing. Technical report, National Bureau of Economic Research.

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.

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.

Hall, J., Hicks, J., Kleiner, M. M., and Solomon, R. (2019). Occupational Licensing of Uber Drivers. *Stanford Institute for Theoretical Economics*.

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. 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.

Kleiner, M. M. and Kudrle, R. T. (2000). Does Regulation Affect Economic Outcomes? The Case of Dentistry. *The Journal of Law and Economics*, 43(2):547–582.

Kleiner, M. M. and Soltas, E. J. (2019). A Welfare Analysis of Occupational Licensing in US States. Technical report, National Bureau of Economic Research.

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

Kugler, A. D. and Sauer, R. M. (2005). Doctors Without Borders? Relicensing Requirements and Negative Selection in the Market for Physicians. *Journal of Labor Economics*, 23(3):437–465.

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.

Learning Policy Institute (2024). The State of the Teacher Workforce. Accessed: 10 February 2025.

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.

Orellana, A. and Winters, M. A. (2023). Licensure Tests and Teacher Supply. Technical report, WEPC Policy Brief.

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

Ramseyer, J. M. and Rasmusen, E. B. (2015). Lowering the Bar to Raise the Bar: Licensing Difficulty and Attorney Quality in Japan. *The Journal of Japanese Studies*, 41(1):113–142.

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.

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.

Notes

1. Its value as a screen for teacher quality also depends on the extent to which the requirement more substantially deters those who would later prove to be ineffective teachers from even attempting the test. Our analysis focuses exclusively on the deterrence effect of failing among test-takers.

2. See online Appendix Table A.1.

3. See online Appendix Table A.2

4. The state also has alternative routes to licensure available for those entering from outside a state educator preparation program, such as those who have previously taught at a non-public school, completed a post-baccalaureate certification program, or completed an education preparation program out of state. Each of these pathways requires the candidate to pass the relevant licensure test for the position they are entering before they are employed except for those entering to fill Urgent Staffing Need or those who completed an out-of-state degree within a state that had a different testing requirement. In each of these cases, the individual must pass the relevant testing requirement to earn a permanent credential.

5. 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

6. This information is obtained from the ETS website: https://www.ets.org/praxis/site/epp/state-requirements/score-requirements.html

7. Washington used a passing score of 142 for math and 158 for writing, while North Dakota used a passing score of 160 for writing

8. Table A.4 in the online Appendix shows a few endorsements require an additional test, *Foundations of Reading*, which is not administered by ETS. We do not consider this subtest in our analyses.

9. An Initial Educator Certificate is a 3-year certificate for those who have either completed a preparation program or have at least 20 school-months of teaching experience in a non-public school. A Provisional Educator Certificate is an 8-year certification for those who have at least 10 school-months of experience under a different certificate type or at least 30 school-months of appropriate experience in a non-public school.

10. An Interim Educator Certificate is a nonrenewable certificate issued to those who have not fully completed either the testing or coursework requirements to obtain an Initial Educator Certification.

11. 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.

12. We estimate the model using the rdrobust command in STATA and report as our primary results estimates from the Robust specification.

13. We observe a first-time test-taker already teaching in 1.4% and 4.1% of cases in our Praxis I and Praxis II samples, respectively. In addition to tests taken prior to the years included in our sample, this group might include, among others, teachers who move from a different state to Connecticut and haven't completed the state certification requirements or individuals entering the teacher workforce through alternative certification programs.

14. Goldhaber and Hansen (2010) employ balance tests to assess differences in race and gender between applicants who fail and pass Praxis II tests in North Carolina. They do not find evidence of discontinuities at any of the cut scores they analyze.

15. We only illustrate select outcomes within the main text for space considerations. See Figure A.5 in the online Appendix for graphical description of the remaining outcomes.

16. Table A.6 in the online Appendix further investigates whether failing a licensure test altered the credential obtained to enter the classroom. On Praxis I, columns (1) and (2) of Panel A show the probability of obtaining different certification types. We find a reduction of 6 ppts (11%) on the probability of earning a standard teaching certification and no effect on the probability of getting a provisional or interim certification. Columns (3)-(5) display estimates for the probability of being observed as a teacher conditional on the certification status and obtaining certification but not teaching. We find a decrease of 3.7 ppts (8%) and 2.5 ppts (18%) on the likelihood of being observed as a certified teacher and getting certification but not obtaining a teaching position, respectively. Panel B shows that failing Praxis II reduced the likelihood that an individual obtained a standard teaching certification, but did not impact the probability of obtaining a provisional or interim certification, nor did it affect the likelihood that we observed the individual teaching despite not identifying their receipt of a certification. Those who failed Praxis II were about 4.5 ppts (21.4%) less likely to obtain a certification but not a teaching position, though this result is likely mechanical given that failing the test reduces the likelihood that the individual obtains a certification.

17. Table A.7 in the Online Appendix presents similar estimates for the effect of failing Praxis I on the likelihood of taking and passing Praxis II.

18. See online Appendix B for a detailed description of our approach to estimating teacher value-added in math and ELA.

19. In a related policy brief (Orellana and Winters, 2023), within our sample we find modest associations between test score value-added and Praxis II scores for math and ELA teachers, similar to what previous literature has shown (Clotfelter et al., 2006, 2007, 2010).

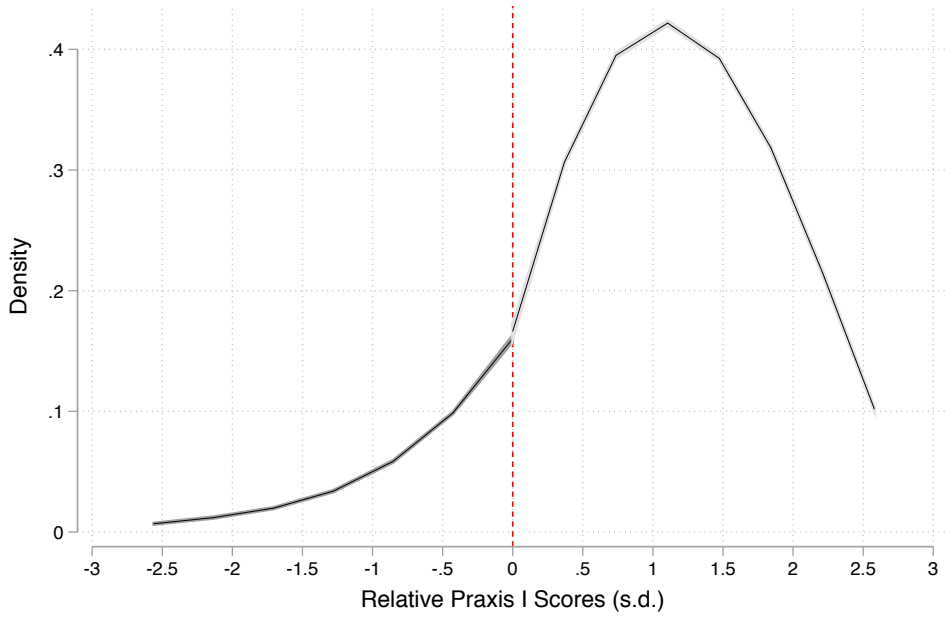
20. We present a description of the most relevant endorsements held by teachers in these sub-samples in online Appendix Table A.10.

21. We describe these mechanisms in much greater detail and provide examples in a simulation exercise in online Appendix C.

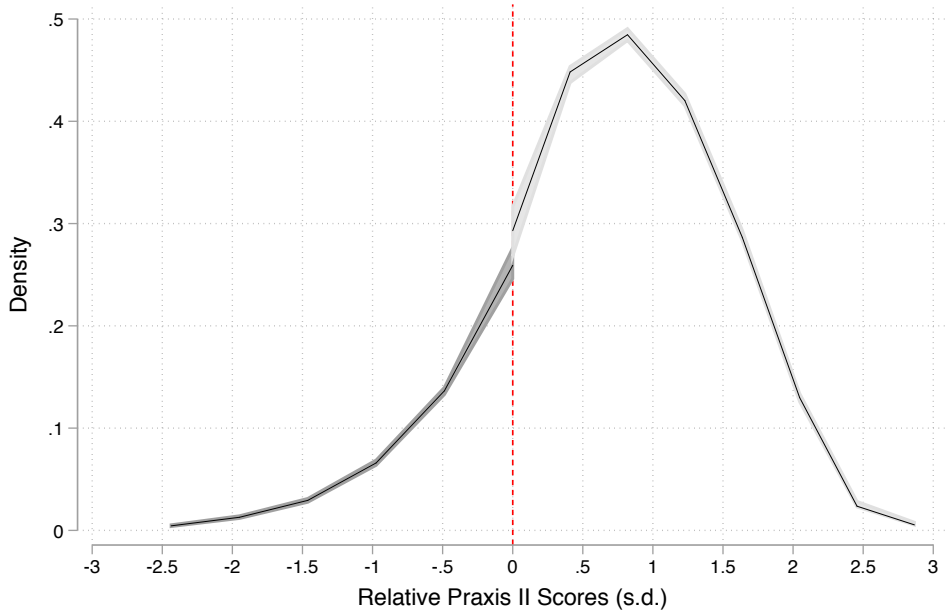
6 Figures and Tables

Figure 1: Density of Praxis Tests Around the Threshold

(a) Praxis I



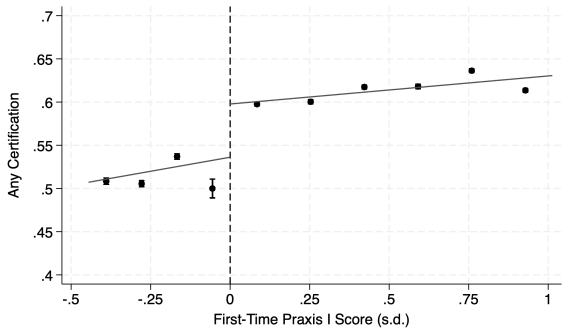
(b) Praxis II



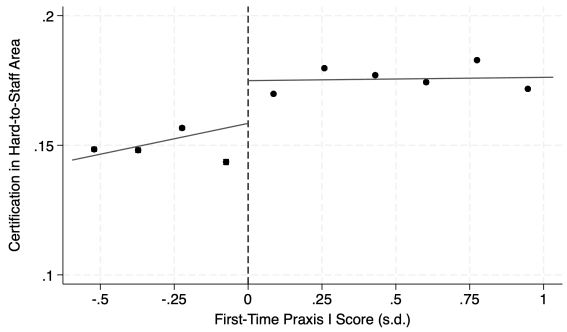
Notes: This figure illustrates the density of standardized Praxis scores around the threshold. The density and 95% confidence intervals at each side of the cutoff were estimated following Cattaneo et al. (2018). The discontinuity test has a p-value of 0.54 for Praxis I and a p-value of 0.35 for Praxis II. These values imply there is no statistical evidence to reject the null hypothesis of no discontinuity at the threshold.

Figure 2: Nonparametric Estimates for the Effect of Failing First Attempt on Licensure Test on Likelihood Obtain Certification or Endorsement

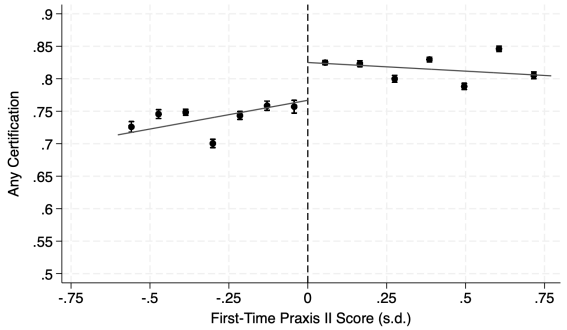
(a) Any Certification (Praxis I)



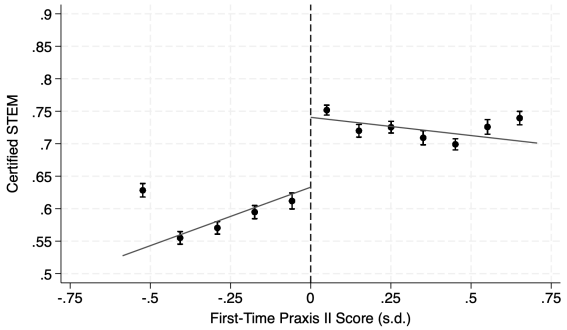
(b) Endorsement in Hard-to-Staff Subject (Praxis I)



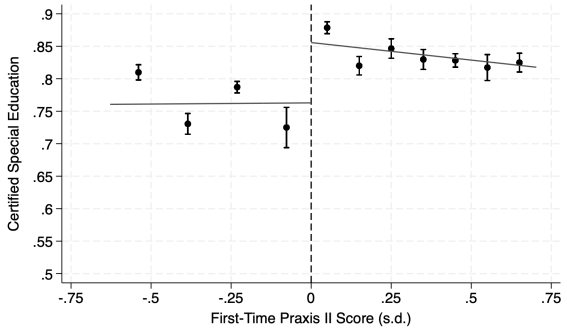
(c) Any Certification (Praxis II)



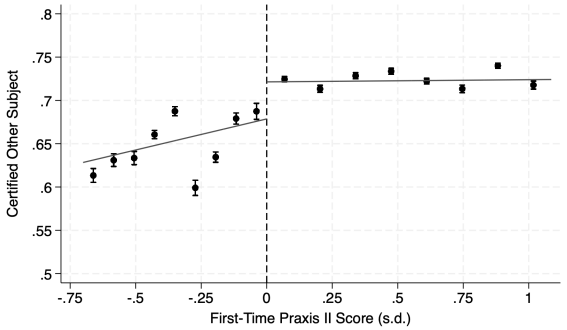
(d) Endorsement in STEM (Praxis II)



(e) Endorsement in Special Education (Praxis II)



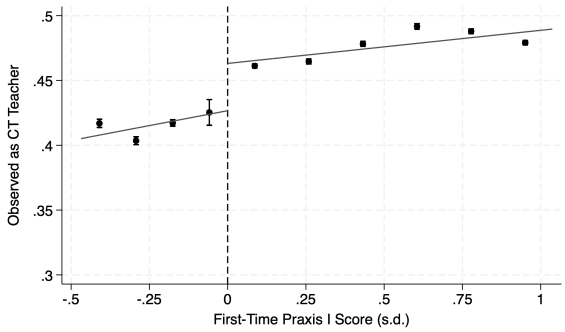
(f) Endorsement in Other Subjects (Praxis II)



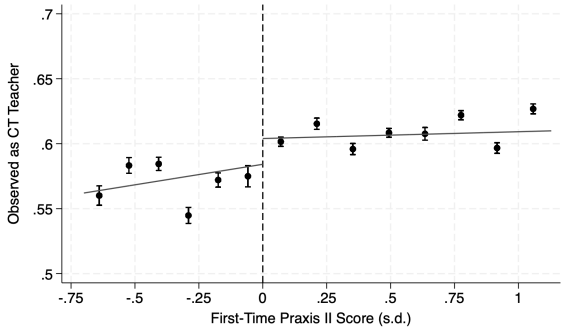
Notes: This figure illustrates the relationship between obtaining endorsement or teaching within Connecticut and first-time scores on the Praxis I (subplots (a)-(b)) and Praxis II (subplots (c)-(f)) tests. Figures created using the rdplot command in STATA. Each regression employs CCT optimal bandwidths (Calonico et al., 2014) and a triangular kernel. Observations binned according to the IMSE-optimal evenly-spaced method using polynomial regression; dots illustrate average within bin and whiskers illustrate the 95% confidence interval. Only select outcomes illustrated for space. See Figure A.5 in the online Appendix for other measured outcomes.

Figure 3: Nonparametric Estimates for the Effect of Failing First Attempt on Licensure Test on Likelihood Teach in a Connecticut Public School

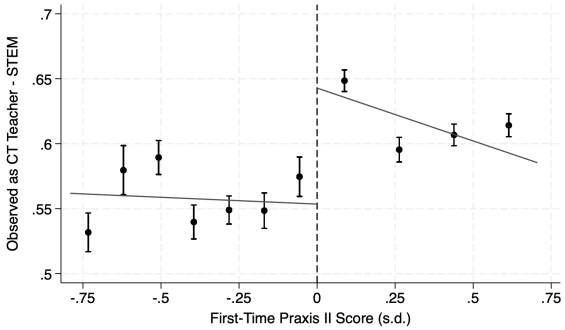
(a) Praxis I



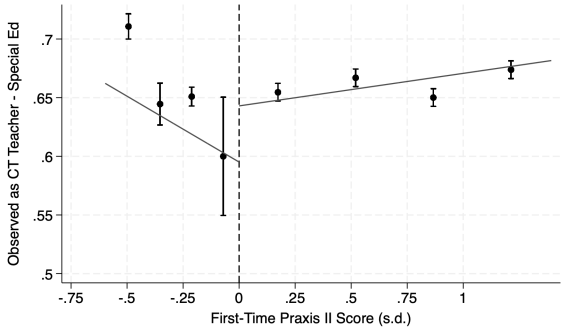
(b) Praxis II: Any Subject



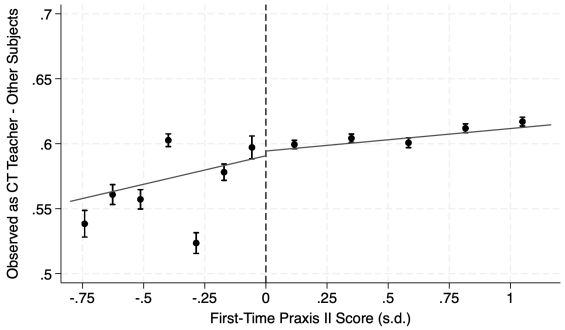
(c) Praxis II: STEM Subject



(d) Praxis II: Special Education

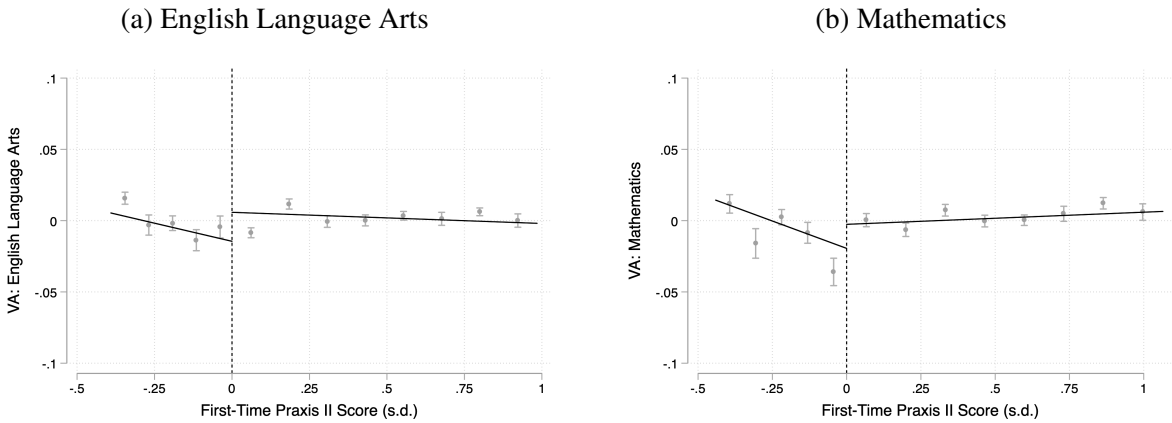


(e) Praxis II: Other Subject



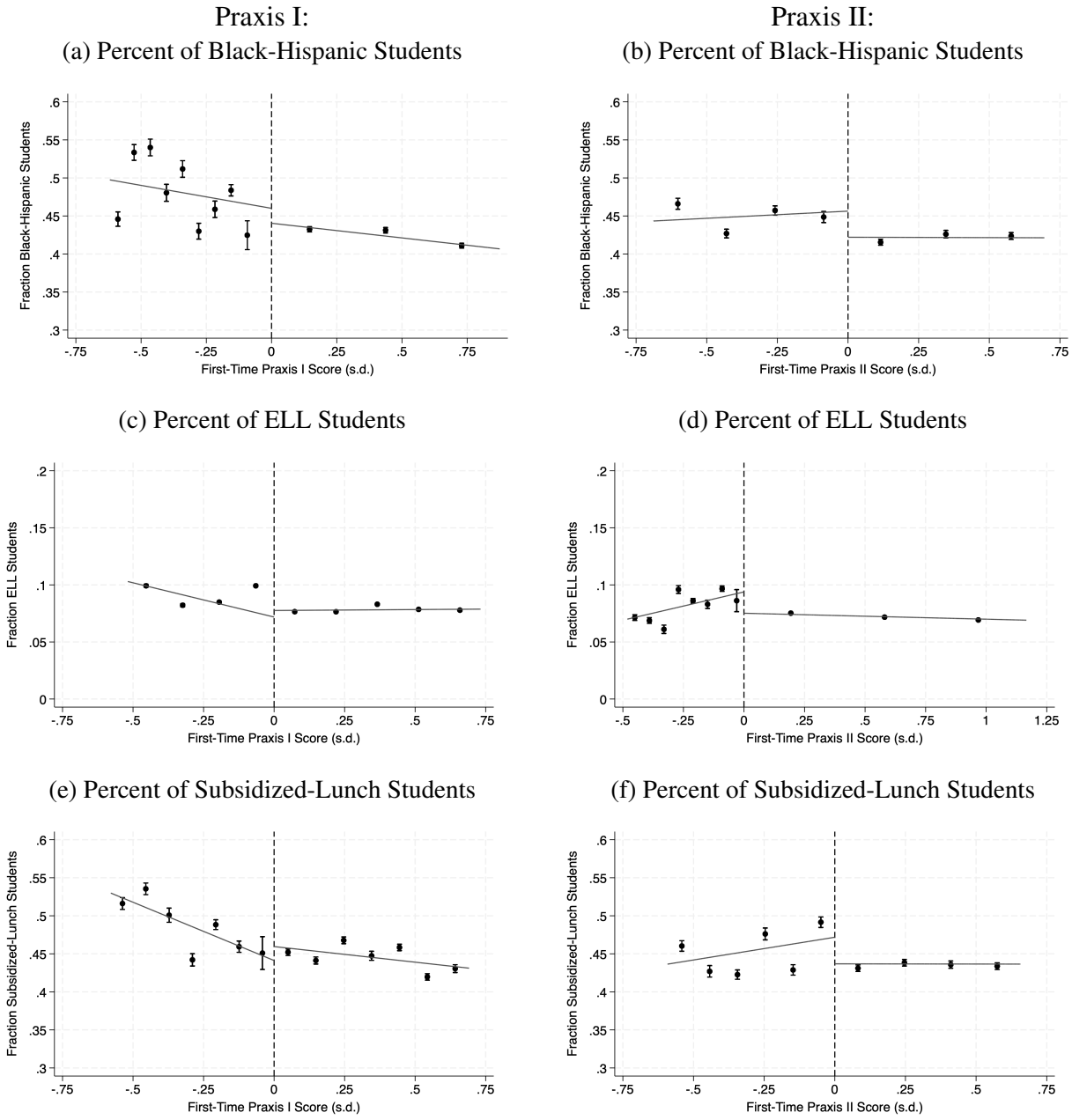
Notes: This figure illustrates the relationship between being observed as a teacher in a Connecticut public school and first-time scores on the Praxis I (subplot (a) and Praxis II (subplots (b)-(e)) tests. Figures created using the rdplot command in STATA. Each regression employs CCT optimal bandwidths (Calonico et al., 2014) and a triangular kernel. Observations binned according to the IMSE-optimal evenly-spaced method using polynomial regression; dots illustrate average within bin and whiskers illustrate the 95% confidence interval. Only select outcomes illustrated for space. See Figure A.5 in the online Appendix for other measured outcomes.

Figure 4: Illustrating Discontinuity in Observed Value-Added



Notes: This figure illustrates the relationship between estimated value-added in ELA and math and first-time score on the Praxis II tests. Figures created using the rdplot command in STATA. Each regression employs optimal bandwidths following Calonico et al. (2014) and a triangular kernel. Observations binned according to the IMSE-optimal evenly-spaced method using polynomial regression; dots illustrate average within bin and whiskers illustrate the 95% confidence interval.

Figure 5: Illustrating Discontinuity in Characteristics of School First Employed as a Teacher



Notes: This figure illustrates the relationship between the first-time score on the Praxis I (left panel) and Praxis II (right panel) tests and the characteristics of schools where test-takers are observed teaching for the first time. Each regression uses optimal bandwidths following Calonico et al. (2014) and a triangular kernel. Observations binned according to the IMSE-optimal evenly-spaced method using polynomial regression; dots illustrate average within bin and whiskers illustrate the 95% confidence interval.

Table 1: Summarizing Pathway to Becoming a Teacher

	All Praxis I Takers			Obtained Certification	
	(1) Pass	(2) Fail		(4) Pass	(5) Fail
Take Praxis II	0.667	0.494***			
Pass Praxis II	0.640	0.452***			
Teaching Certification	0.603	0.423***			
Interim Certification	0.015	0.028***			
Observed Teacher	0.462	0.347***		0.754	0.782***
Teaching (No Certificate)	0.004	0.007***			
N	50,283	20,218		30,317	8,561
	All Praxis II Takers		Ever Passed Praxis II	Obtained Certification	
	(1) Pass	(2) Fail	(3) Fail	(4) Pass	(5) Fail
Retake Praxis II		0.784			
Ever Pass Praxis II		0.744			
Teaching Certification	0.811	0.673***	0.802***		
Interim Certification	0.014	0.039***	0.029***		
Observed Teacher	0.605	0.535***	0.628***	0.734	0.765***
Teaching (No Certificate)	0.005	0.008***	0.005		
N	57,082	23,958	17,833	46,308	16,119

Notes: This table presents the probability of advancing in a prospective applicant's teaching path for different sub-samples of the universe of test-takers. Top panel reports results for Praxis I sample and bottom panel reports results for Praxis II sample. Each cell shows the average of test-takers observed in the respective category, conditional on whether they passed or failed their first attempt. Columns (1) and (2) show differences by passing status for all test takers. Column (3) is restricted to test-takers who ever passed a Praxis II test. Columns (4) and (5) condition on individuals who obtained a teaching certification. Significant differences derived from t-tests compare Columns (2) and (3) against Column (1) and compare Column (4) to Column (5). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: RD Estimates for Effect of Failing First Administration of Licensure Test on Certification & Likelihood of Teaching: Full Sample

	Panel A: Praxis I			
	(1) Take Praxis II	(2) Pass Praxis II	(3) Any Certification	(4) Teaches Any Subject
Failed	-0.056***	-0.051***	-0.067***	-0.038*
	(0.019)	(0.018)	(0.023)	(0.021)
Average Outcome	0.65	0.61	0.60	0.46
Bandwidth	(-0.57,1.04)	(-0.60,1.00)	(-0.45,1.02)	(-0.50,1.05)
N	42,314	42,015	39,873	41,668
	Panel B: Praxis II			
			(5) Any Certification	(6) Teaches Any Subject
Failed			-0.066***	-0.021
			(0.013)	(0.015)
Average Outcome			0.79	0.60
Bandwidth			(-0.58,0.78)	(-0.71,1.18)
N			34,307	42,846

Notes: This table presents estimates of the effects of failing the first attempt at Praxis I (top panel) and Praxis II (bottom panel) scores on different outcomes. Dependent variables are indicators for whether the individual later attempted Praxis II, ever passed Praxis II, ever obtained any teaching certification, and ever taught in a Connecticut public school. CCT optimal bandwidths (computed using the methodology proposed by Calonico et al. (2014)) are reported at the bottom of the respective analysis. Each regression controls for the difference between the individual's initial licensure score and the passing score for the respective test within a linear function allowing for changes in the slope at the threshold, as well as both year and test fixed effects. Heteroskedastic robust standard errors reported in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: RD Estimates for Effect of Failing Praxis I on Certification and Teaching in Connecticut

	Panel A: Ever Obtain Teacher Certification									
	Any Endorsement	Special Education	Foreign Languages	ESOL	STEM	Elementary Grades	Art and Music	English	History	Other Subjects
Failed	-0.067***	0.004	-0.007	-0.007	-0.033***	-0.035*	0.016*	0.022**	-0.010	-0.036
	(0.023)	(0.015)	(0.007)	(0.006)	(0.008)	(0.020)	(0.009)	(0.009)	(0.009)	(0.023)
Average Outcome	0.595	0.107	0.022	0.016	0.052	0.283	0.036	0.041	0.044	0.424
Bandwidth					(-0.448,1.019)					
N					39,873					

	Panel B: Ever Observed Teaching									
	Any Subject	Special Education	Foreign Languages	ESOL	STEM	Elementary Grades	Art and Music	English	History	Other Subjects
Failed	-0.038*	0.004	-0.008	-0.001	-0.030***	-0.003	0.021***	0.008	-0.008	-0.023**
	(0.021)	(0.012)	(0.006)	(0.004)	(0.007)	(0.016)	(0.008)	(0.009)	(0.006)	(0.009)
Average Outcome	0.457	0.086	0.020	0.007	0.056	0.184	0.029	0.055	0.031	0.059
Bandwidth					(-0.500,1.054)					
N					41,668					

Notes: This table presents estimates of the effects of failing the first attempt at Praxis I on the likelihood of obtaining a certification (panel A) and teaching (panel B) in a Connecticut public school. Dependent variables are indicators for whether the individual later obtained teacher certification in an area and in each subject within Connecticut. Bandwidths are computed following Calonico et al. (2014). Each regression controls for the difference between the individual's licensure score and the passing score for the respective test within a linear function allowing for changes in the slope at the threshold, as well as both year and test fixed effects. Heteroskedastic robust standard errors reported in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: RD Estimates for Effect of Failing First Administration of Licensure Test: STEM Test Takers

	Praxis II	Endorsement			Teaching		
	Take Different Test	Any Subject	STEM	Non-STEM	Special Ed	Foreign	TESOL
Failed	-0.072**	-0.079***	-0.107***	0.028*	0.001	0.005	0.005
	(0.031)	(0.031)	(0.032)	(0.016)	(0.010)	(0.005)	(0.004)
Average Outcome	0.415	0.726	0.670	0.056	0.022	0.004	0.003
Bandwidth	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)
N	6,106	6,106	6,106	6,106	6,106	6,106	6,106

	Teaching						
	STEM	Elementary	Music	English	History	Other	No Teaching
Failed	-0.073**	0.003	0.003	-0.007	-0.011	-0.013	0.087***
	(0.034)	(0.016)	(0.003)	(0.008)	(0.008)	(0.012)	(0.033)
Average Outcome	0.574	0.051	0.002	0.013	0.012	0.033	0.369
Bandwidth	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)	(-0.614,0.63)
N	6,106	6,106	6,106	6,106	6,106	6,106	6,106

Notes: This table presents estimates of the effects of failing the first attempt at Praxis II on different outcomes. Dependent variables are indicators for whether the individual later took a Praxis II test in a non-STEM subject, ever obtained any teaching certification, ever obtained a teaching certification in a STEM or non-STEM subject, and if they were ever employed as a teacher in each subject within a Connecticut public school. CCT optimal bandwidths (computed using the methodology proposed by Calonico et al. (2014)) are reported at the bottom of the respective analysis. Each regression controls for the difference between the individual's licensure score and the passing score for the respective test within a linear function allowing for changes in the slope at the threshold, as well as both year and test fixed effects. Heteroskedastic robust standard errors reported in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: RD Estimates for Effect of Failing First Administration of Licensure Test: Special Education Test Takers

	Praxis II	Endorsement			Teaching		
	Take Different Test	Any Subject	Special Ed	Non-Special Ed	Special Ed	Foreign	TESOL
Failed	0.117**	-0.061	-0.106**	0.045	-0.111**	0.003	0.002
	(0.054)	(0.041)	(0.045)	(0.028)	(0.050)	(0.011)	(0.009)
Average Outcome	0.444	0.847	0.818	0.029	0.598	0.004	0.005
Bandwidth	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)
N	3,392	3,392	3,392	3,392	3,392	3,392	3,392

	Teaching						
	STEM	Elementary	Music	English	History	Other	No Teaching
Failed	-0.002	0.085**	0.003	0.007	0.021*	-0.009	0.051
	(0.012)	(0.041)	(0.010)	(0.017)	(0.013)	(0.016)	(0.052)
Average Outcome	0.019	0.122	0.005	0.034	0.017	0.019	0.314
Bandwidth	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)	(-0.606,0.686)
N	3,392	3,392	3,392	3,392	3,392	3,392	3,392

Notes: This table presents estimates of the effects of failing the first attempt at Praxis II on different outcomes. Dependent variables are indicators for whether the individual later took a Praxis II test in an area other than Special Education, ever obtained any teaching certification, ever obtained a teaching certification in a Special Education or non-Special Education subject, and if they were ever employed as a teacher in each subject within a Connecticut public school. CCT optimal bandwidths (computed using the methodology proposed by Calonico et al. (2014)) are reported at the bottom of the respective analysis. Each regression controls for the difference between the individual's licensure score and the passing score for the respective test within a linear function allowing for changes in the slope at the threshold, as well as both year and test fixed effects. Heteroskedastic robust standard errors reported in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: RD Estimates for Effect of Failing First Administration of Licensure Test: Elementary K-6 Test Takers

	Praxis II	Endorsement			Teaching		
	Take Different Test	Any Subject	Elementary	Non-Elementary	Special Ed	Foreign	TESOL
Failed	-0.007	-0.029*	-0.035**	0.006	-0.009	-0.001	-0.000
	(0.019)	(0.015)	(0.016)	(0.007)	(0.012)	(0.003)	(0.004)
Average Outcome	0.261	0.861	0.835	0.025	0.099	0.006	0.009
Bandwidth	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)
N	17,020	17,020	17,020	17,020	17,020	17,020	17,020

	Teaching						
	STEM	Elementary	Music	English	History	Other	No Teaching
Failed	-0.001	0.011	0.000	-0.014	-0.008*	0.001	-0.011
	(0.010)	(0.021)	(0.003)	(0.009)	(0.004)	(0.007)	(0.020)
Average Outcome	0.052	0.487	0.006	0.055	0.015	0.033	0.384
Bandwidth	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)	(-0.792,0.972)
N	17,020	17,020	17,020	17,020	17,020	17,020	17,020

Notes: This table presents estimates of the effects of failing the first attempt at Praxis II on different outcomes. Dependent variables are indicators for whether the individual later took a Praxis II test in an area other than Elementary Grades, ever obtained any teaching certification, ever obtained a teaching certification in a Elementary Grades or non-Elementary subject, and if they were ever employed as a teacher in each subject within a Connecticut public school. CCT optimal bandwidths (computed using the methodology proposed by Calonico et al. (2014)) are reported at the bottom of the respective analysis. Each regression controls for the difference between the individual's licensure score and the passing score for the respective test within a linear function allowing for changes in the slope at the threshold, as well as both year and test fixed effects. Heteroskedastic robust standard errors reported in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7: Evaluating Discontinuity in Value-Added and School Characteristics Among Observed Teachers

	(1) VAM: ELA	(2) VAM: Math	(3) % Black-Hispanic	(4) % ELL	(5) % Subsidized Lunch
Failed	-0.024**	-0.023	0.046*	0.023**	0.047**
	(0.011)	(0.016)	(0.025)	(0.009)	(0.022)
IV	-0.311*	-0.352	0.758	0.358**	0.813*
	(0.183)	(0.320)	(0.512)	(0.166)	(0.492)
Average Outcome	0	0	0.43	0.07	0.44
Bandwidth	(-0.39,0.99)	(-0.44,1.03)	(-0.66,0.69)	(-0.48,1.16)	(-0.59,0.66)
N	1,914	1,792	4,780	6,405	4,600

Notes: This table presents RD and IV estimates investigating discontinuities at the passing threshold of Praxis II in the relationship between licensure score on the first attempt and observed estimated value-added and school characteristics. The outcome in columns (1) and (2) is the estimated teacher value-added in the respective subject following the methodology described in Online Appendix B. The outcome in columns (3)-(5) is the fraction of students within the respective subgroup in the school where the teacher is first observed teaching. On each panel, the first row (Failed) reports results from a sharp RD regression estimated via OLS, and the second row (IV) reports the results from models that adjust first row estimate by the appropriate first-stage. RD estimates employ optimal bandwidths following Calonico et al. (2014). IV estimates are computed by resampling the data 1,000 times (with repetition) and estimating in each iteration the effect of failing the licensure test on the probability of obtaining a certification (first stage) and the effect on the corresponding outcome (reduced-form) using the specification described by Equation (1). The estimates and standard errors for the IV analyses correspond to the mean and standard deviation of the bootstrapped distribution, respectively. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.