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 I especially deters those who would be less effective teachers, but failing Praxis II disproportionately pushes out relatively effective potential teachers.

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 Boston University PREREQ seminar, the annual conferences for the Association of Education Finance and Policy, and the Society of Labor Economists. 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 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

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 weed-out ineffective teachers by posing a disproportionately large barrier for them to obtain certification, and push-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) call 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).

We apply a sharp regression discontinuity design to more than two decades of administrative data from Connecticut to investigate the impact of failing a licensure test on one's 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). Our results shed new light on the impacts that licensure test requirements have on the number and quality of the teacher workforce.

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 that an initial test-taker obtains a teaching certification by about 11.2% and reduces the probability they eventually teach within a Connecticut public school by 8.2%. The effect of failing the first attempt at Praxis II differs substantially by the subject area in which the candidate is seeking endorsement. When considering all Praxis II tests together, failing the first attempt reduces the likelihood of obtaining any teaching certification and of eventually teaching within the state by 8.4% and 3.5%, respectively, though the latter estimate is not statistically significant. But the deterrent effect from failing Praxis II is especially large for those seeking an endorsement to teach within the shortage areas. In particular, failing the first administration of the Praxis II test to obtain endorsement to teach within a STEM subject reduces the likelihood that the individual is eventually observed as a Connecticut public school teacher by about 15.4%, but failing a test for certification in an area other than STEM or special education does not alter the likelihood of becoming a public school teacher.

We then investigate the extent to which deterrence from failing a licensure test differs by the test-taker's latent value-added, by which we mean the contribution they would make to student test scores if they were to become a teacher in the state. From a policy perspective we are primarily concerned with whether failing a licensure test disproportionately deters those who would prove to be ineffective teachers if they entered the classroom.[1] However, prior studies measuring the correlation between licensure score and value-added cannot speak directly to this issue because they do not observe the value-added of licensure-test-takers who never enter the classroom (Clotfelter et al., 2006, 2007, 2010; Goldhaber et al., 2017). We address this challenge by looking for discontinuity at the passing threshold in the

[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.

relationship between initial licensure score and value-added for those we observe as teachers. Our results suggest that failing Praxis I more substantially deters those who would have proven to be relatively less effective math teachers, but failing Praxis II disproportionately pushes out those who would have been relatively more effective teachers. We posit that the different impacts associated with latent value-added across the tests could be explained by the substantial differences in the skills they evaluate and their timing on the teacher preparation pipeline.

Finally, we investigate the extent to which deterrance from failing a licensure test impacts different types of schools by applying our strategy to measure the relationship between failing the first attempt at a licensure test and the characteristics of the school in which the individual first teaches. We find no significant relationship between failing Praxis I and the characteristics of the schools in which the individual is first employed. However, failing Praxis II more substantially deters candidates who would have otherwise taught in a school with smaller shares 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.

Our estimates by subject area are unique within the literature and highly relevant for policy. Prior studies that rely on aggregated data 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 matter 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). 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 institution 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 from a different compelling identification strategy applied to administrative data in which we can measure differential effects on a direct measure of teacher value-added.

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).

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 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). 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). 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 (Larsen et al., 2022; Ramseyer and Rasmusen, 2015; Bhattacharya et al., 2019).

The remainder of the paper proceeds as follows. Section 2 describes the data and licensure testing requirements in Connecticut. We investigate the effect of failing the first administration of a licensure test on progress towards becoming a teacher in Section 3, and we investigate differential treatment effects by latent value-added and the characteristics of the school in which they are first employed as a teacher in Section 4. Section 5 summarizes and concludes.

2 Data

2.1 Licensure Tests

Similar to other states, in Connecticut the typical certification process 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.

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.

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] 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-

[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

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.

Praxis requirements used in Connecticut are similar to those used in other states. Table A.2 in the Online Appendix shows Praxis I passing scores used by Connecticut before 2016 and summary statistics of current passing scores in other states.[3] 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.2, with the exception of Washington and North Dakota.[4]

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.1 in the Online Appendix shows the link between some of the endorsement codes offered in Connecticut and the Praxis II tests required. Some

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

[4]Washington used a passing score of 142 for math and 158 for writing, while North Dakota used a passing score of 160 for writing

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.[5]

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.3 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.2 and A.3 show that states do not vary substantially in terms of their test score requirements. In addition, these cutoffs are typically invariant 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 score in 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, including the candidate's race/ethnicity and gender.

For our analyses, we define a certified teacher as one who has obtained

[5]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 do not consider this subtest in our analyses.

a renewable Initial or Provisional Educator Certificate.[6] 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.[7] 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

[6] 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.

[7] 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.

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 in this sub-sample to a single classroom teacher.

2.5 Summary Statistics

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 of 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.[8] 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 sub-tests. We account for changes in the tests over time and differences across subject-area tests by including fixed effects for year and specific test administered.

Our main analyses are based on a sharp regression discontinuity design using the following specification:

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

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

[8] 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.

slope at the cutoff value.[9] 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.

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

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

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 there to exist a systematic discontinuity at the passing threshold in the characteristics of test-takers.[10]

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 to 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 associated with obtaining each endorsement rather than the first attempt at any Praxis II test.

The figures illustrate several notable patterns. First, 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.

Notable from a policy perspective, the fairly flat slope in the relation-

[10] 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.

ship between licensure score and the respective outcome on the passing side of the threshold suggest that beyond the effect of fulfilling the passing requirement there is little to no relationship between licensure score and the pathway to becoming a teacher. Those who barely passed Praxis II are as likely as those who scored a standard deviation above the passing threshold to eventually teach within Connecticut.

[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. For Praxis I we also report estimates for the effect of failing on the likelihood the individual moves forward on the teacher pipeline by taking and passing Praxis II, which also serves as a proxy for completing an educator preparation program.

The results for Praxis I are reported on the table's top panel.[11] For those scoring at the threshold, failing the basic-skills test required for entry into an educator preparation program reduced the likelihood of eventually obtaining 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 also reduced the likelihood of eventually obtaining an endorsement to teach within a hard-to-staff subject by 3.1 ppts (17.2%).

[TABLE 2 ABOUT HERE]

Estimates for the causal effect of failing the first attempt at Praxis II are reported on the bottom panel of the table. Failing the first attempt at Praxis II reduced

[11] 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.

the likelihood of obtaining any teaching certification by about 6.6 ppts (8.4%). The effect appears especially large for those seeking an endorsement to teach within a shortage area. Failing the first attempt at the required version of Praxis II reduces the likelihood of obtaining an endorsement to teach a STEM subject by 8.9 ppts (13%), in special education by 10.7 ppts (13%), and for other subjects by 4.7 ppts (6.7%).

Table 3 reports regression discontinuity estimates for the effect of failing a particular licensure test on the likelihood that an individual is eventually observed teaching in a Connecticut public school. Those who failed Praxis I were 3.8 ppts (8.2%) less likely to become a Connecticut public school teacher than if they had passed. The effect of failing Praxis II differs substantially by the subject area in which the candidate is seeking endorsement. Considered together, failing Praxis II on average reduces the likelihood the test-taker is observed as a Connecticut public school teacher by about 2.1 ppts (3.5%), but the estimate is not statistically significant. However, candidates who fail their fist attempt at the Praxis II subject-matter test for endorsement to teach within a STEM subject are about 9.7 ppts (15.4%) less likely to become a public school teacher in the state than had they passed the test. We also find some evidence that failing the Praxis II test for endorsement to teach within special education reduces the likelihood of becoming a public school teacher, but this effect is estimated imprecisely. In contrast, failing the first attempt at a Praxis II test in a subject other than STEM or special education does not impact the likelihood that an individual becomes a teacher.

[TABLE 3 ABOUT HERE]

Table A.5 in the Online Appendix 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 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 effect 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 obtain a teaching position, though this result is likely mechanical given that failing the test reduces the likelihood that the individual obtains a certification.

4 Investigating Heterogeneous Treatment Effects

4.1 Estimating Differential Effects by Latent Value-Added

In this sub-section our goal is to investigate the extent to which the deterrent effect of failing the first administration of a licensure test differs by the test-taker's latent value-added. The primary challenge is that we observe value-added only for those who eventually teach within a tested grade and subject. Our strategy uses changes in observed value-added at the threshold to infer differences related to latent value-added in the effect of failing on the likelihood of becoming a teacher.

We estimate a regression discontinuity design similar to our primary analysis, but using the individual's observed value-added score as the dependent variable.[12] Formally:

$$\theta_i = \tau + f(x_{ij}) + \lambda \mathbb{1}(x_{ij} < 0) + \phi_j + \phi_t + \epsilon_{ij} \tag{2}$$

where θ_i is individual i's estimated value-added score, and all other variables are as previously defined. The estimation sample is necessarily restricted to include only licensure-test-takers with an observed value-added score. We are primarily interested in λ, which represents the conditional difference in average observed value-added at the passing threshold.

A conventional regression discontinuity design would seek to interpret λ as the causal effect of failing a licensure test on a prospective teacher's latent value-added. However, such an interpretation is likely inappropriate because failing reduces the likelihood that an individual becomes a teacher, and thus that we observe their value-added, and the magnitude of this effect could correlate with latent value-added (Larsen et al., 2022). In that case, λ would incorporate both the direct effect of failing the test on latent value-added and selection bias due to differential attrition associated with latent value-added on the failing side of the threshold.

However, our goal is to investigate whether the effect of failing the test on the likelihood an individual's value-added is observed differs by latent value-added. That is, what would be considered worrisome selection bias in a conventional regression discontinuity design is our specific focus.

We impose the additional assumption that failing the first attempt at a licensure test does not directly impact an individual's latent value-added. For example, this assumption would be violated if preparing to pass a retake led to systematic improvements in an individual's potential effectiveness as a teacher. Though

[12] See Online Appendix A.2 for a detailed description of our approach to estimating teacher value-added in math and English Language Arts (ELA).

we cannot directly test this assumption, we argue that it is highly plausible given the nature of the tests, the availability of multiple retakes, and that licensure scores correlate weakly with the later value-added of those who eventually become teachers.[13] Under this assumption, a finding that $\lambda \neq 0$ can be interpreted as evidence that the effect of failing on the likelihood of becoming a teacher differs by latent value-added. Further, if this assumption is violated it is notable that to the extent that failing does increase the latent value-added of those who eventually become teachers it would tend to bias λ upwards.

Let us formalize the approach in order to fix ideas. Let θ_i^* be the latent value-added of test-taker i, which we allow to correlate with their initial licensure score, x_i. We observe test-taker i's value-added, θ_i, if and only if we observe them as a teacher, $T_i = 1$. We allow the probability of observing i's value-added to correlate directly with initial licensure test score, though the patterns illustrated on Figure 3 suggest this may not be the case. In addition, failing the first administration of a licensure test reduces the likelihood one's value-added is observed, and this deterrent effect may differ by i's latent value-added.

$$\theta_i = \begin{cases} \theta_i^* & T_i = 1 \\ . & T_i = 0 \end{cases} \tag{3}$$

$$T_i = \begin{cases} 1 & \gamma_0 + \gamma_1 x_i + \gamma_2 \mathbb{1}(x_i < 0) + \gamma_3 \mathbb{1}(x_i < 0) \times \theta_i^* + \mu_i > 0 \\ 0 & \text{Otherwise} \end{cases} \tag{4}$$

$$\theta_i^* = \delta_0 + \delta_1 x_i + \epsilon_i \tag{5}$$

For any point x on the distribution of initial licensure scores, the expected

[13] 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).

value for observed value added is average latent value added weighted by the probability the individual is observed as a teacher: $E(\theta_i \mid x_i = x) = \sum_{i=1}^{N} T_i(x) \times \theta_i^*(x)$. For passing scores, T_i is independent of θ_i^*, and thus $E(\theta_i \mid x \geq 0) = E(\theta_i^* \mid x \geq 0)$. But the situation is different on the failing side of the initial test score distribution. If $\gamma_3 = 0$ then T_i would remain independent of θ_i^*, and average observed value-added would remain an unbiased estimate for average latent value-added. However, if $\gamma_3 \neq 0$ then T_1 would depend on θ_i^* and consequently $E(\theta_i \mid x < 0) \neq E(\theta_i^* \mid x < 0)$.

The intuition with which we apply the above relationships in order to interpret $\lambda \neq 0$ in equation (2) as evidence that the deterrent effect of failing a licensure test differs by latent value-added is as follows. As a baseline characteristic, θ_i^* is balanced at the threshold, and thus if θ_i were observed for all test-takers it would similarly balance at the threshold. But in practice we only observe θ_i for a subset of test-takers. For any given passing score the likelihood an individual's θ_i is observed is as-good-as-randomly determined with respect to θ_i^*, and thus average θ_i is an unbiased estimate of average θ_i^*. If the effect of failing the test on the likelihood that an individual's value-added score is observed does not depend on θ_i^*, then average θ_i is an unbiased measure of average θ_i^* on the both sides of the threshold and average θ_i would similarly balance. However, if those with more/less θ_i^* are more substantially deterred by failing the test, then the resulting selection bias in observed value-added as a measure for latent value-added will appear as a systematic drop/increase in average θ_i on the failing side of the threshold, and thus we would find $\lambda \neq 0$.[14]

[14] Let's consider two illustrative examples. First, consider the scenario where the treatment reduces the probability of observing value-added equally across the distribution of latent value-added ($\gamma_3 = 0$). In this case, exposure to the treatment would not alter the relationship between θ_i^* and T_i. Since in this case average θ_i is an unbiased measure of θ_i^* on both sides of the threshold, and the relationship between x and θ_i^* is smooth at the threshold, average θ_i would also be balanced at the threshold.

4.1.1 Results

Figure 4 illustrates the relationship between initial licensure score and average observed value-added in ELA and mathematics. The results reported in Columns (1) and (2) of Table 4 coincide with the respective figures.

On Praxis I, to the left of the threshold we observe an upward jump in average math value-added of about 0.053σ, which is significant at the 10% level. This pattern suggests that failing the first administration of the basic skills test more substantially reduced the likelihood of observing the value-added of test-takers with lower latent value-added. The estimated impact from failing Praxis I on observed value-added in ELA is in the opposite direction (-0.02σ), but not statistically significant at any conventional level.

On Praxis II, for both subjects we find a small discontinuous drop in average observed value-added scores (-0.02σ) on the failing side of the threshold, implying that the treatment disproportionately reduced the likelihood of observing the value-added scores of test-takers with higher latent value-added. Thus, our results suggest that failing Praxis II disproportionately deterred those who would have been relatively higher performing math teachers if they were to have entered the classroom.

[FIGURE 4 ABOUT HERE]

Now consider the scenario where $\gamma_3 < 0$, indicating that failing the licensure test more substantially deterred those who would be relatively less effective teachers were they to have entered the classroom. Those who passed the test are not exposed to the treatment, and thus average θ is the same as in the previous example. However, in this case for any given failing score those with lower θ_i^* would be less likely to have an observed value-added score than if the treatment did not exist. Consequently, average θ_i for a given x on the failing side of the threshold would be systematically higher than it would be in absence of the treatment effect. In this case, average θ_i is an unbiased measure of θ_i^* on the passing side of the threshold, but is an upwards-biased estimate of θ_i^* on the failing side of the threshold. Since the relationship between x and θ_i^* is smooth at the threshold, this pattern would appear as a discontinuous jump upward in observed value-added on the failing side of the threshold, and thus we would find $\lambda > 0$.

[TABLE 4 ABOUT HERE]

4.2 Estimating Differential Effects by School Type

Finally, in order to speak to the impact of test requirements on the distribution of teachers across schools, we analyze the extent to which failing the initial attempt at a licensure test alters the type of school in which an individual is first employed as a teacher. As in the prior analysis, our challenge is that we observe school placements only for those test-takers who eventually obtain a teaching position. We again address this challenge by estimating Equation 2, but using as the dependent variable a select characteristic for the school in which the individual teaches. The sample is necessarily restricted to include only those who are ever observed teaching within a Connecticut public school. As outcomes we consider the proportion of students within the teacher's first school who are Black or Hispanic, the percent classified as an English Language Learner (ELL), and percent eligible for free or reduced-priced lunch.

When interpreting the results from this analysis, we might be additionally concerned with the underlying assumption that failing the first attempt at a licensure test does not directly impact the type of school in which an individual first teaches. It is unlikely that failing the first administration of a licensure test alters the school's perception of a candidate because at the time of hire schools typically do not observe an applicant's licensure score or the the number of times they took a test. However, it is possible that failing the first attempt at Praxis II could delay an applicant's entry into the teacher job market as a certified candidate, thus limiting their access to schools with more advantaged students, which face fewer challenges with filling teaching positions (Hanushek et al., 2004; Boyd et al., 2013). It is difficult to imagine plausible scenarios such that failing the first administration of Praxis II would directly improve the likelihood that a particular candidate is first employed within a school with more historically advantaged students. Thus, when considering the likelihood of first teaching in a school with more historically disadvantaged students, to the extent that failing has a direct impact on the outcome we would expect the estimate for λ is likely biased upwards relative to our preferred interpretation that the impact is driven entirely by selective attrition.

Our results are reported on Table 4 and illustrated on Figure 5. On Praxis I, we find no significant discontinuity at the threshold in the characteristics of the schools in which an individual teaches. However, among those observed teaching within the state, those who failed their first attempt at Praxis II on average taught in schools with a 4.6 ppts and 2.3 ppts higher proportion of Black-Hispanic and ELL students, respectively. Column (5) shows a similar increase of 4.7 ppts in the fraction of students receiving subsidized lunch. Applying the intuition we described in the previous section, our results suggest that failing the first attempt at Praxis II more substantially deterred test-takers who would have on average taught in schools with relatively fewer historically disadvantaged students.

[FIGURE 5 ABOUT HERE]

5 Discussion and Conclusion

We find evidence that licensure test requirements at different points on the teacher preparation pipeline serve as a barrier to entering the teaching profession that can have unintended consequences for the teacher workforce. Failing the first attempt at a basic skills test to gain entry into an educator preparation program (Praxis I) disproportionately weeds out relatively less effective teachers, and thus may have value as a screen for teacher quality. However, failing the first attempt at the Praxis

II subject-matter test required to obtain a certification as a candidate prepares to apply for teaching positions on average pushes out relatively higher quality candidates and serves as an especially large barrier for those seeking an endorsement to teach within the persistent shortage areas of STEM and special education.

Among our unique contributions is to investigate differences in the impact of a stringent licensure requirement associated with a direct value-added measure of teacher effectiveness. Prior related studies have employed aggregated data and used the competitiveness of the colleges from which teachers are drawn as a plausible but arguably limited proxy for teacher quality.

Our finding that the relationship between the deterrence effect of failing and latent value-added differs across the two tests is worthy of future consideration. However, we argue that it is perhaps not especially surprising that tests measuring different capabilities and occurring at very different points in a prospective teacher's pre-service preparation would deter different types of candidates. For example, failing Praxis I early in college might especially discourage those who are only marginally interested in pursing teaching as a career, and level of commitment could be associated with latent quality. In contrast, those taking Praxis II have demonstrated sufficient interest in a teaching career to have nearly completed an educator preparation program. However, if latent teacher quality correlates with possessing skills valued in other professions, then we might expect that among those who fail their first attempt at Praxis II those with higher latent value-added could have more attractive employment opportunities available to pursue outside of teaching.

Our analysis suffers from a few notable limitations that are worthy of future research. First, we lack the necessary demographic data to investigate whether failing a licensure test poses an especially large barrier to entering the classroom for potential-teachers of color. Considering differential treatment effects by the race/ethnicity and other characteristics of prospective teachers is an obvious area for future research. It is notable, however, that given that teachers of color are more likely to teach in schools that serve larger proportions of historically disadvantaged students, our finding that failing Praxis II more substantially deters those who would have otherwise taught in schools with smaller proportions of such students counters the pattern we would expect to see if failing was a larger barrier for prospective teachers of color.

An important limitation of our analysis of differential effects by latent value-added is that we are only able to analyze differences in mean teacher quality. Larsen et al. (2022) convincingly argue that analyzing the impact of the stringency of licensure requirements at the mean could mask effects at tails, which is where theory would predict such impacts should occur. Unfortunately, interpreting the results from applying our strategy for considering heterogeneous effects requires much stronger assumptions when the outcome variable is categorical instead of continuous, and thus we are not able to produce compelling estimates for whether failing a licensure test disproportionately deters candidates with latent value-added that would fall within the bottom quartile, for example.[15] Nonetheless, it is notable that unlike Larsen et al. (2022) we find some significant and substantial differences in impacts in mean latent teacher quality, and thus in our case considering impacts at the mean has not prohibited us from identifying an effect, though it could mask the true magnitude of the effect at the tails.

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 aspect

[15] Interpreting the results of such an analysis as a difference in the causal effect of failing on the likelihood of becoming a teacher related to the quartile of latent value-added requires the seemingly strong and untestable assumption that in absence of the treatment effect test-takers with latent value-added in the bottom quartile are equally likely to become a teacher as test-takers with latent value-added that does not fall within the bottom quartile.

of 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 increasing the stringency of coursework requirements increases 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, 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.

Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, 31(1):83–117.

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 Papers 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., Kain, J., and Rivkin, S. (2004). Why Public Schools Lose Teachers. *Journal of Human Resources*, 39(2).

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.

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.

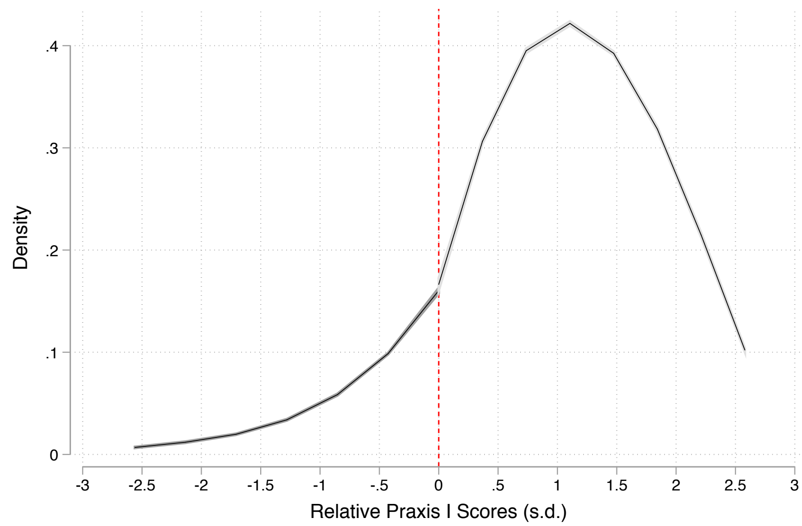
Shuls, J. V. (2018). Raising the Bar on Teacher Quality: Assessing the Impact of Increasing Licensure Exam Cut-Scores. *Educational Policy*, 32(7):969–992.

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

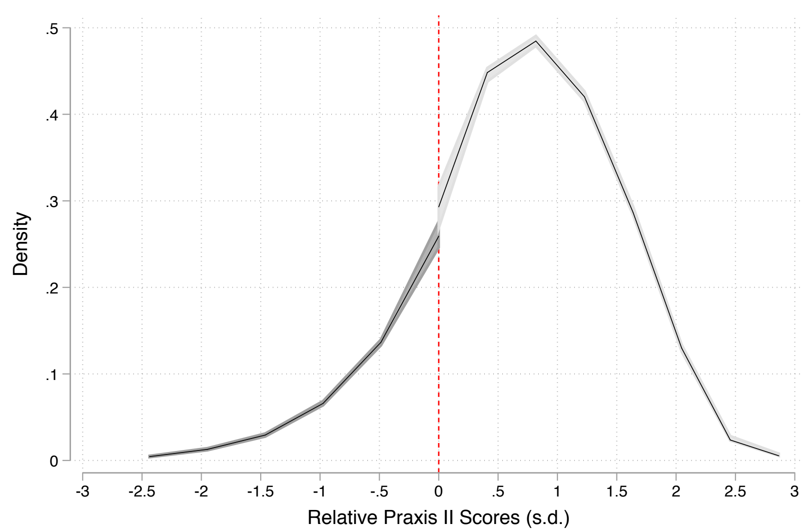
6 Figures and Tables

Figure 1: Density of Praxis Tests Around the Threshold

(a) Praxis I



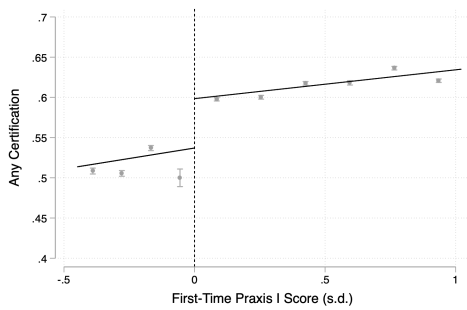
(b) Praxis II



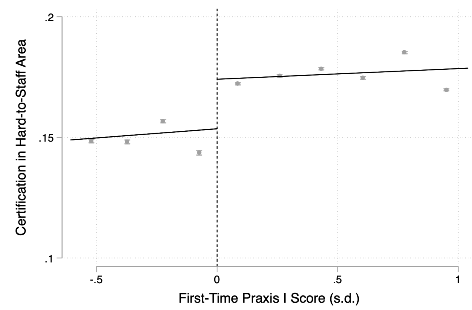
Notes: This figure illustrates the density of standardized Praxis scores around the threshold. The density and 95% confidence intervals at each side of the cutoff were estimated following Cattaneo et al. (2018). The discontinuity test has a p-value of 0.54 for Praxis I and a p-value of 0.35 for Praxis 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

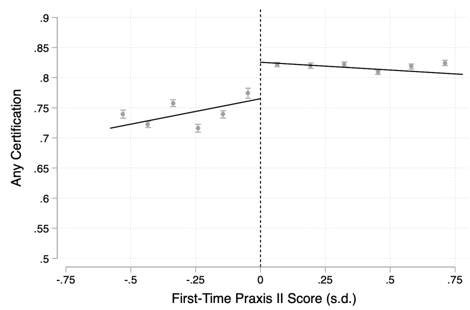
(a) Any Certification (Praxis I)



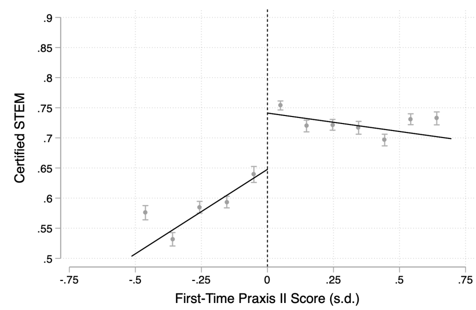
(b) Endorsement in Hard-to-Staff Subject (Praxis I)



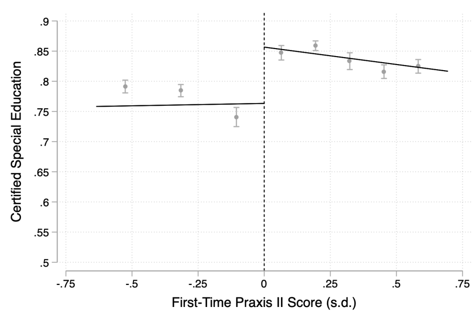
(c) Any Certification (Praxis II)



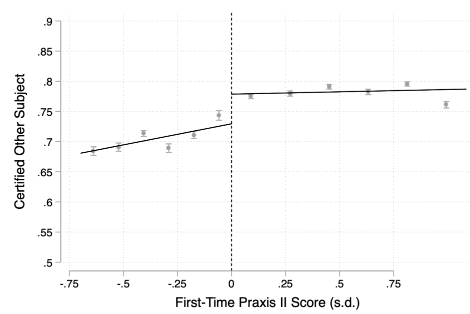
(d) Endorsement in STEM (Praxis II)



(e) Endorsement in Special Education (Praxis II)



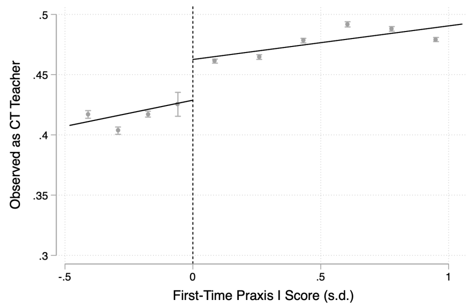
(f) Endorsement in Other Subjects (Praxis II)



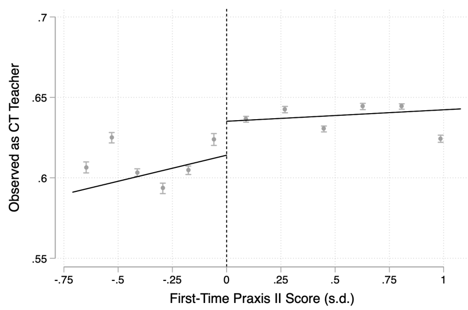
Notes: This figure illustrates the relationship between obtaining endorsement or teaching within Connecticut and first-time scores on the Praxis I (subplots (a)-(b)) and Praxis II (subplots (c)-(f)) tests. Figures created using the rdplot command in STATA. Each regression employs CCT optimal bandwidths (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

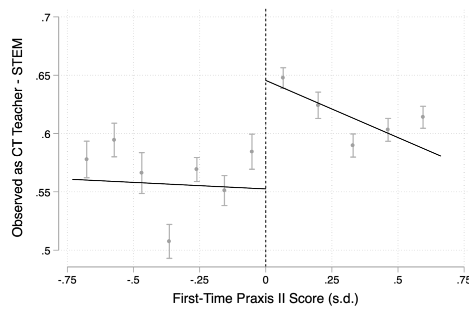
(a) Praxis I



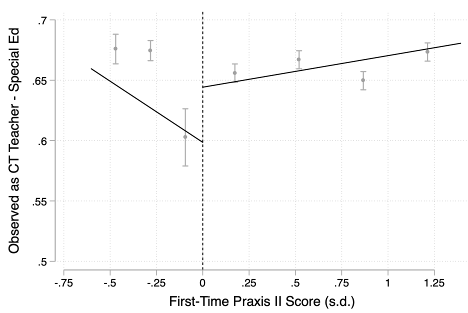
(b) Praxis II: Any Subject



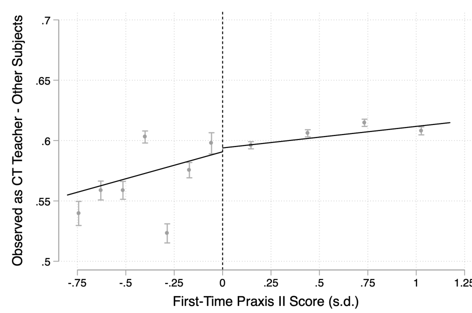
(c) Praxis II: STEM Subject



(d) Praxis II: Special Education



(e) Praxis II: Other Subject

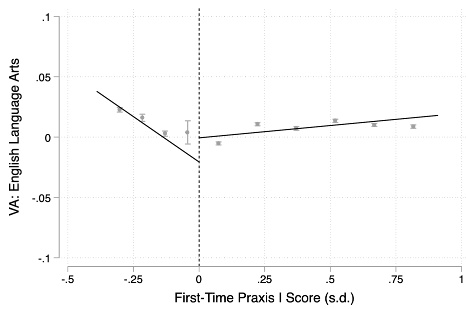


Notes: This figure illustrates the relationship between being observed as a teacher in a Connecticut public school and first-time scores on the Praxis I (subplot (a) and Praxis II (subplots (b)-(e)) tests. Figures created using the rdplot command in STATA. Each regression employs CCT optimal bandwidths (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

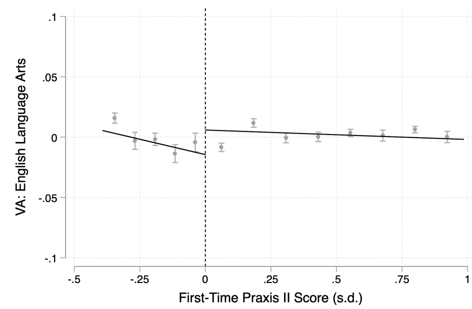
Praxis I:

(a) English Language Arts

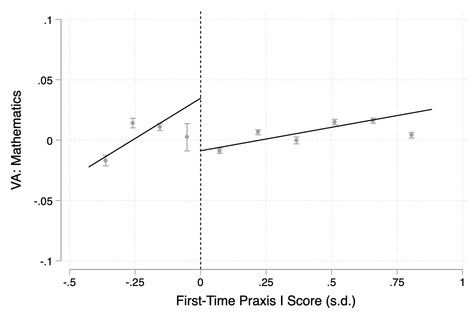


Praxis II:

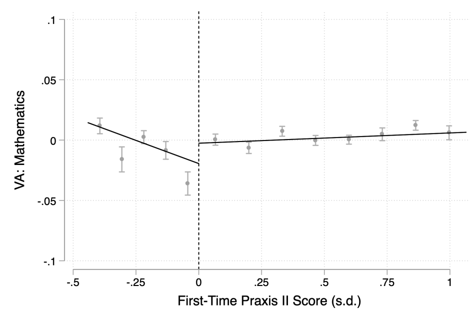
(b) English Language Arts



(c) Mathematics



(d) Mathematics

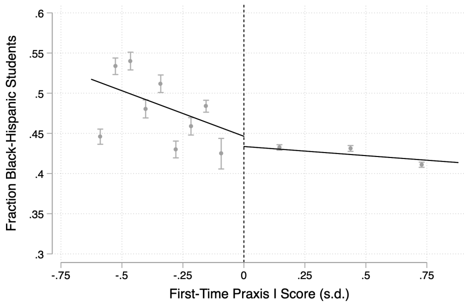


Notes: This figure illustrates the relationship between estimated value-added in ELA (top panel) and math (bottom panel) and first-time score on the Praxis I (left panel) and Praxis II (right panel) tests. Figures created using the rdplot command in STATA. Each regression employs optimal bandwidths following 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

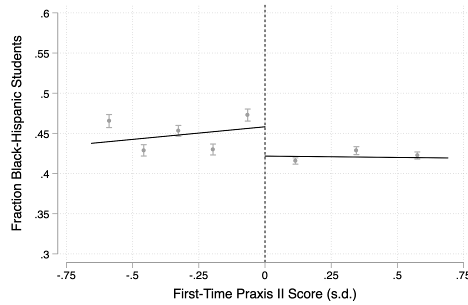
Praxis I:

(a) Percent of Black-Hispanic Students

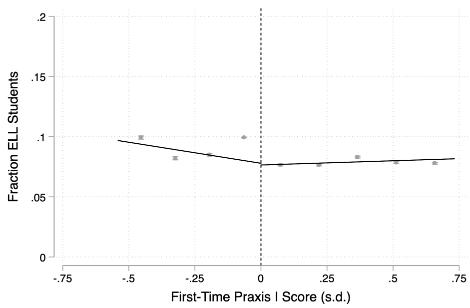


Praxis II:

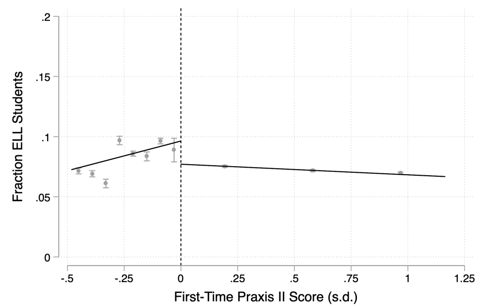
(b) Percent of Black-Hispanic Students



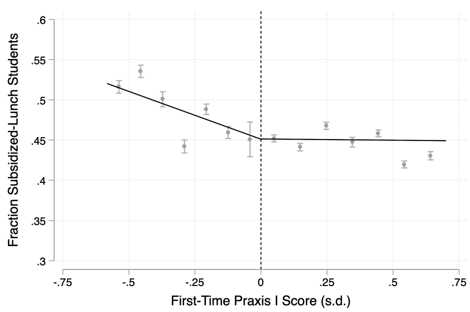
(c) Percent of ELL Students



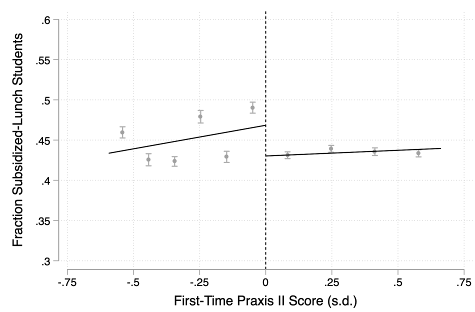
(d) Percent of ELL Students



(e) Percent of Subsidized-Lunch Students



(f) Percent of Subsidized-Lunch Students



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

	Panel A: Praxis I			
	(1) Take Praxis II	(2) Pass Praxis II	(3) Any Certification	(4) Hard Staff
Failed	-0.056***	-0.051***	-0.067**	-0.031**
	(0.019)	(0.018)	(0.023)	(0.015)
Average Outcome	0.65	0.61	0.60	0.18
Bandwidth	(-0.57,1.04)	(-0.60,1.00)	(-0.45,1.02)	(-0.54,1.00)
N	42,314	42,015	40,423	40,512

	Panel B: Praxis II			
	Any Certification	STEM	Special Ed	Other Subjects
Failed	-0.066***	-0.089***	-0.107**	-0.047***
	(0.013)	(0.033)	(0.042)	(0.015)
Average Outcome	0.79	0.68	0.82	0.70
Bandwidth	(-0.58,0.78)	(-0.51,0.69)	(-0.63,0.69)	(-0.70,1.08)
N	34,307	6,207	3,425	36,501

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 for the Praxis I analysis are indicators for whether the individual later attempted Praxis II, ever passed Praxis II, ever obtained any teaching certification, and ever obtained a teaching certification in a hard-to-staff subject. Dependent variables for the Praxis II analysis are indicators for whether the individual ever obtained any teaching certification, obtained an endorsement to teach within a STEM subject, within special education, and in a subject other than STEM or special education. Analyses of STEM and special education endorsement are restricted to the first administration of a test associated with that particular endorsement, rather than the first Praxis II attempt. 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 3: RD Estimates for Effect of Failing First Administration of Licensure Test on Likelihood of Teaching

	(1)	(2)	(3)	(4)	(5)
	Praxis I	Praxis II			
		Any Subject	STEM	Special Ed	Other Subjects
Failed	-0.038*	-0.021	-0.097***	-0.047	-0.006
	(0.021)	(0.015)	(0.030)	(0.051)	(0.014)
Average Outcome	0.46	0.60	0.63	0.69	0.61
Bandwidth	(-0.50,1.05)	(-0.71,1.18)	(-0.79,0.59)	(-0.61,1.15)	(-0.82,1.28)
N	41,653	42,846	6,413	5,641	42,062

Notes: This table presents estimates of the effects of failing the first attempt at Praxis I and Praxis II on the likelihood of being observed as a Connecticut public school teacher. Columns (1) and (2) show the estimate of failing Praxis I and Praxis II on being ever observed teaching in any subject. Columns (3)-(5) shows separate estimates of failing the first attempt at Praxis II by subject. 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 4: 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
			Panel A: Praxis I		
Failed	-0.020	0.053*	0.017	0.002	0.003
	(0.019)	(0.028)	(0.035)	(0.012)	(0.033)
Average Outcome	0.01	0.01	0.44	0.08	0.46
Bandwidth	(-0.41,0.91)	(-0.41,0.88)	(-0.62,0.88)	(-0.54,0.73)	(-0.58,0.70)
N	1,459	1,200	4,342	3,801	3,517
			Panel B: Praxis II		
Failed	-0.024**	-0.023	0.046*	0.023**	0.047**
	(0.011)	(0.016)	(0.025)	(0.009)	(0.022)
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 estimates investigating discontinuities at the passing threshold of Praxis I (top panel) and Praxis II (bottom panel) in the relationship between licensure score on the first attempt and estimated value-added and school characteristics where they are first observed teaching. The sample is restricted to include only those observed employed as a teacher within a Connecticut public school. The outcome in columns (1) and (2) is the estimated teacher value-added in the respective subject following the methodology described in Online Appendix A.2. The outcome in columns (3)-(5) is each school's respective fraction of students. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

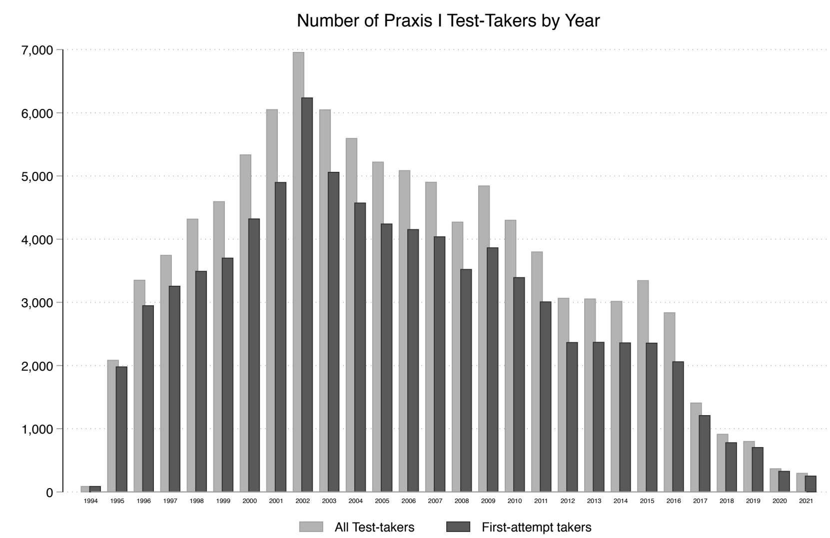
A Online Appendix

Licensure Tests and Teacher Supply

Alexis Orellana & Marcus A. Winters

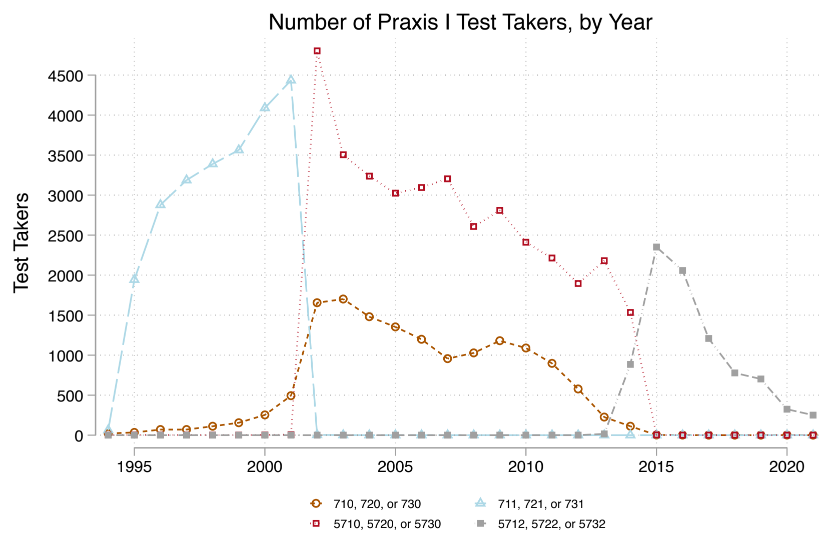
A.1 Additional Figures and Tables

Figure A.1: Praxis I Test-takers



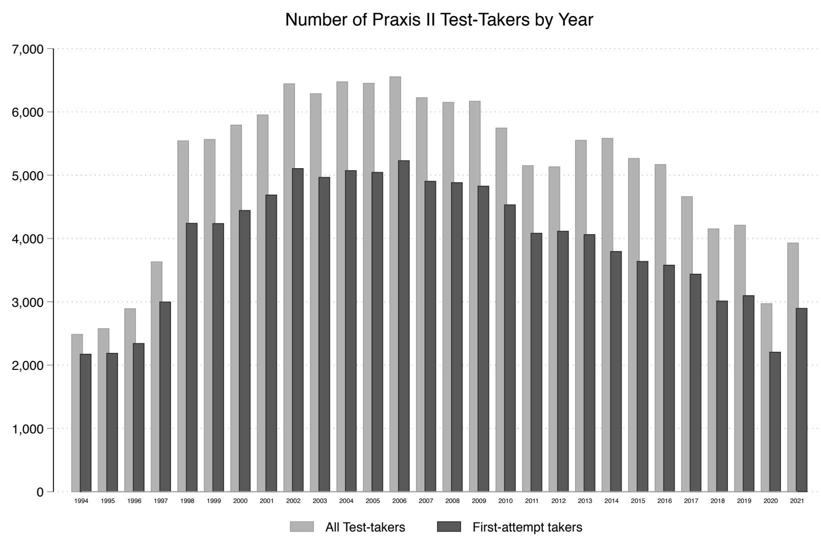
Notes: This plot shows the number of applicants who took a Praxis I test between 1995 and 2021. Gray and black bars represent the total number of test-takers and the total number of first-time applicants in each year, respectively.

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.

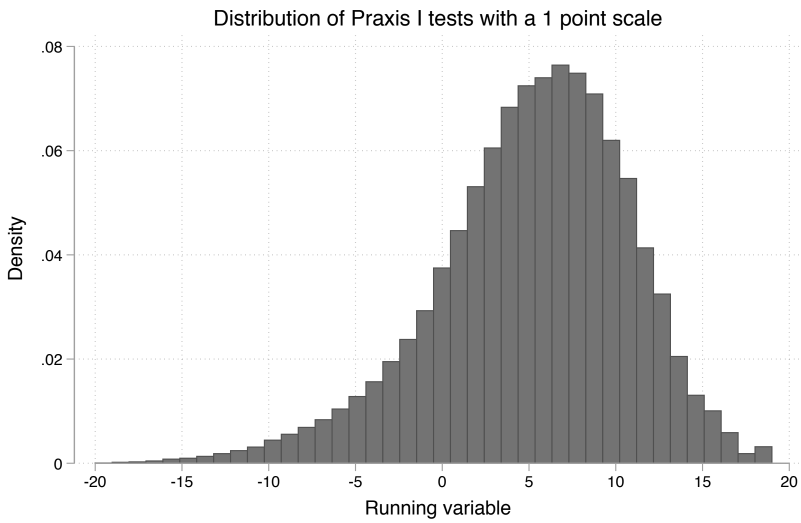
Figure A.3: Praxis II Test-takers



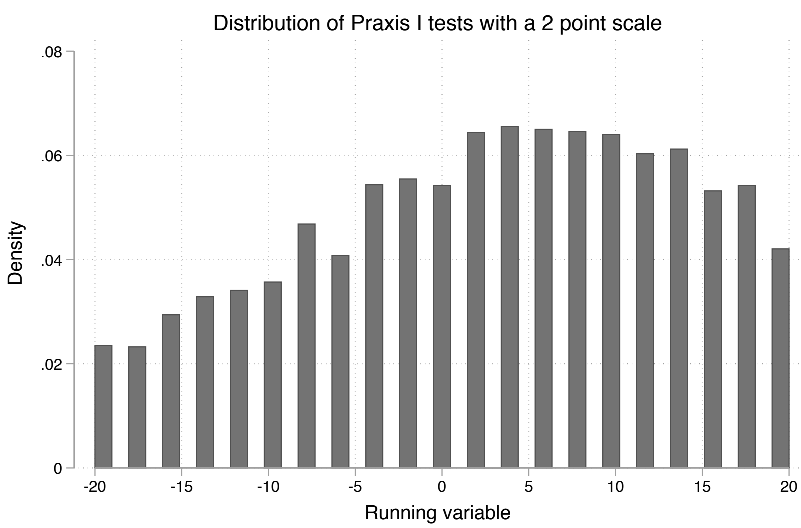
Notes: This plot shows the number of applicants who took a Praxis II test between 1995 and 2021. Gray and black bars represent the total number of test-takers and the total number of first-time applicants in each year, respectively.

Figure A.4: Distribution of Praxis I Scores

(a) Tests using a 1-point scale



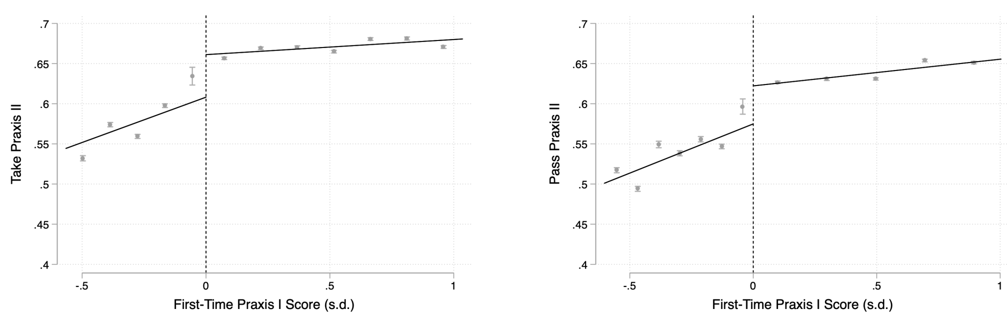
(b) Tests using a 2-point scale



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.

Figure A.5: Nonparametric Estimates for the Effect of Failing First Attempt on Licensure Test: Additional Outcomes

(a) Take Praxis II (b) Pass Praxis II

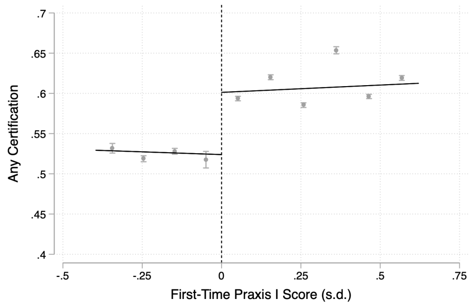


Notes: This figure includes additional outcomes to the ones presented in Figure 2 showing the relationship between failing a Praxis I test and subsequent outcomes. Each regression employs CCT optimal bandwidths (Calonico et al., 2014) and a triangular kernel. Observations binned according to the IMSE-optimal evenly-spaced method using polynomial regression; dots illustrate average within bin and whiskers illustrate the 95% confidence interval. Only select outcomes illustrated for space.

Figure A.6: Nonparametric Estimates for the Effect of Failing First Attempt on Licensure Test: Using an Uniform Kernel

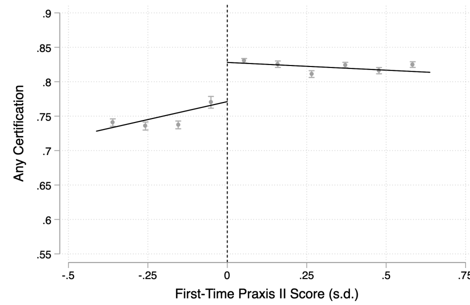
Praxis I:

(a) Obtain Any Certification

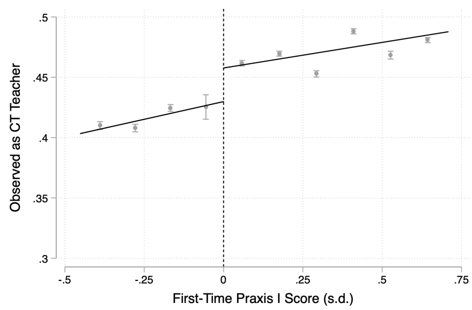


Praxis II:

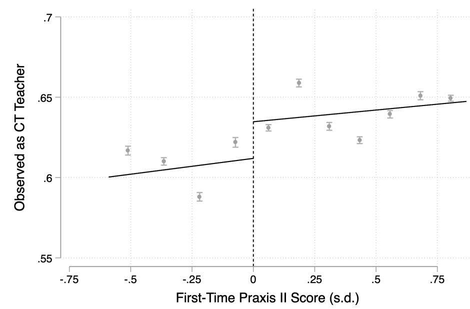
(b) Obtain Any Certification



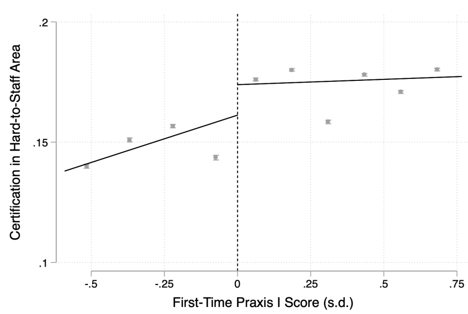
(c) Teach Within a Connecticut Public School



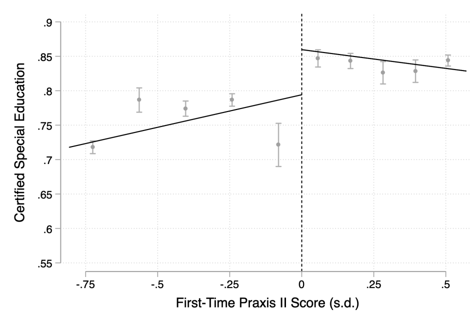
(d) Teach Within a Connecticut Public School



(e) Obtain Endorsement in Hard-to-Staff Subject



(f) Obtain Endorsement to Teach Special Education



Notes: This figure replicates the results from Figures 2 and 3 using an uniform kernel. Each regression employs CCT optimal bandwidths (Calonico et al., 2014). 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.

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: Praxis I Passing Scores in Connecticut and Other States

Test Code	Description	Connecticut	Other States			
		(before 2016)	Average	S.D.	Mode	Number
Core Academic Skills for Educators:						
5713	Reading Subtest	156	155.6	1.3	156	25
5723	Writing Subtest	162	161.4	1.6	162	25
5733	Mathematics Subtest	150	148.7	4.2	150	25

Notes: This table presents Praxis I passing scores employed in Connecticut before 2016 and current passing scores in other states. Column *Connecticut (before 2016)* displays the passing scores used by this state for Praxis I tests 5712, 5722, and 5732. These tests were replaced by the new versions 5713, 5723, and 5733 in 2019. The last four columns show summary statistics of passing scores in other states. Column *Number* shows the number of states using each test while columns *Average*, *S.D.*, and *Mode* present the average value, standard deviation, and modal passing score, respectively, among these states. Score requirements were obtained from the ETS website: https://www.ets.org/praxis/site/epp/state-requirements/score-requirements.html

Table A.3: Praxis II Passing Scores in Connecticut and Other States

Test Code	Description	Connecticut	Other States			
			Average	S.D.	Mode	Number
Elementary Education						
5002	Reading Subtest	157	156.4	1.9	157	22
5003	Mathematics Subtest	157	156.1	3.0	157	22
5004	Social Studies Subtest	155	154.3	2.2	155	22
5005	Science Subtest	159	158.3	2.4	159	22
Middle School						
5047	Middle School ELA	164	163.3	1.8	164	29
5089	Middle School Social Studies	160	152.6	5.4	149	28
5169	Middle School Mathematics	165	165	0	165	5
5442	Middle School Science	152	151.1	1.9	152	29
Secondary Education						
5039	ELA: Content and Analysis	168	167.1	2.1	168	11
5081	Social Studies: CK	162	154.6	3.9	155	25
5101	Business Education: CK	154	154.7	4.6	154	31
5122	Family and Consumer Sciences	153	152.9	1.7	153	32
5161	Mathematics: CK	160	158.4	3.6	160	5
5235	Biology: CK	152	148.8	3.6	150	28
5245	Chemistry: CK	151	149.6	5.4	151	28
5265	Physics: CK	141	137.9	6.7	141	27
5435	General Science: CK	157	152.2	4.8	152	22
5571	Earth and Space Sciences: CK	157	148.8	4.2	150	25
5652	Computer Science	149	148.1	2.4	149	24
K-12						
5051	Technology Education	159	158.7	2.4	159	29
5095	Physical Education: Content and Design	169	168.1	2.0	169	11
5114	Music: Content and Instruction	162	160	3.9	162	8
5135	Art: Content and Analysis	161	160	1.9	161	8
5551	Health Education	164	154.4	5.8	155	25
World Languages						
5362	ESOL	155	153.4	4.1	155	27
Special Education						
5543	Core Knowledge and Mild to Moderate Applications	158	156.4	3.2	158	12

Notes: This table presents Praxis II current passing scores employed in Connecticut and in other states. The last four columns show summary statistics of passing scores in other states for each test. Column *Number* shows the number of states using each test while columns *Average*, *S.D.*, and *Mode* present the average value, standard deviation, and modal passing score, respectively, among these states. Score requirements were obtained from the ETS website: https://www.ets.org/praxis/site/epp/state-requirements/score-requirements.html

Table A.4: RD Estimates for Effect of Failing First Administration of Licensure Test: Uniform Kernel

	Panel A: Praxis I					
	(1) Take Praxis II	(2) Pass Praxis II	(3) Any Certification	(4) Hard Staff	(5) Teach	(6) Teach > 5yr
Failed	-0.054***	-0.045***	-0.082***	-0.026**	-0.029	-0.022
	(0.016)	(0.016)	(0.023)	(0.013)	(0.019)	(0.014)
Average Outcome	0.64	0.61	0.53	0.18	0.45	0.32
Bandwidth	(-0.56,0.77)	(-0.59,0.75)	(-0.40,0.62)	(-0.57,0.77)	(-0.45,0.71)	(-0.59,0.63)
N	33,972	34,884	26,591	33,972	30,740	30,152

	Panel B: Praxis II					
	Any Certification	STEM	Special Ed	Other Subjects	Teach	Teach > 5yr
Failed	-0.065***	-0.137***	-0.076**	-0.053***	-0.033**	-0.037**
	(0.014)	(0.028)	(0.032)	(0.015)	(0.013)	(0.015)
Average Outcome	0.80	0.67	0.81	0.71	0.60	0.47
Bandwidth	(-0.41,0.64)	(-0.66,0.70)	(-0.81,0.57)	(-0.53,0.83)	(-0.57,0.83)	(-0.53,0.90)
N	27,086	6,579	3,349	28,165	35,439	33,459

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 for the Praxis I analysis are indicators for whether the individual later attempted Praxis II, ever passed Praxis II, ever obtained any teaching certification, ever obtained a teaching certification in a hard-to-staff subject, was ever employed as a teacher and taught for more than five years within a Connecticut public school. Dependent variables for the Praxis II analysis are indicators for whether the individual ever obtained any teaching certification, obtained an endorsement to teach within a STEM subject, within special education, and the subject in which the individual was tested in their first Praxis II administration, was ever employed as a teacher and taught for more than five years within a Connecticut public school. Analyses of STEM and special education endorsement are restricted to the first administration of a test associated with that particular endorsement, rather than the first Praxis II attempt. 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 A.5: Effect of Failing Licensure Test on Certification and Teaching Type

	Panel A: Praxis I				
	(1) Standard Certificate	(2) Provisional/Interim Certificate	(3) Teacher with Certification	(4) Certified but does not teach	(5) Teacher with no Certification
Failed	-0.060***	-0.000	-0.037*	-0.025**	0.005*
	(0.023)	(0.007)	(0.020)	(0.012)	(0.003)
Average Outcome	0.56	0.04	0.45	0.14	0.00
Bandwidth	(-0.46,0.98)	(-0.64,0.68)	(-0.51,1.05)	(-0.63,1.13)	(-0.56,0.92)
N	38,613	31,012	41,770	44,200	37,675
	Panel B: Praxis II				
Failed	-0.062***	0.010	-0.022	-0.045***	-0.000
	(0.013)	(0.007)	(0.013)	(0.011)	(0.001)
Average Outcome	0.76	0.05	0.59	0.21	0.00
Bandwidth	(-0.55,1.01)	(-0.60,0.92)	(-0.73,1.18)	(-0.75,1.14)	(-0.59,0.91)
N	40,757	39,177	48,244	47,822	39,083

Notes: This table presents RD estimates investigating discontinuities at the passing threshold of Praxis II in the relationship between licensure score on the first attempt and the likelihood of obtaining different licensure types. Columns (1)-(2) show estimates for standard and provisional/interim certification, respectively. Columns (3)-(5) show estimates for different combinations of teaching and certification categories. 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.2 Estimating Teacher Value-Added

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:

$$y_{ijst} = X'_{ijst}\beta + f(y_{ijst-1})\lambda + \phi_j + \xi_{ijst} \tag{6}$$

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; ϕ_j is a teacher fixed effect; ϵ_{ijst} is a stochastic term, and β and λ 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.7 shows the distribution of the raw teacher fixed-effects $\hat{\phi}_j$ and the empirical Bayes estimates.

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:

$$\hat{\phi}_j = \delta_0 + \delta_1 P_j + \eta_j \tag{7}$$

The estimate for δ_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.

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 (6) and assume these can be decomposed into a component attributable to teachers (ϕ_j), classroom-level shocks (θ_{ct}), and student-level idiosyncratic error (ϵ_{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:

$$\hat{\phi}_j^{EB} = \overline{\xi}_j \times \frac{\hat{\sigma}_\phi^2}{\hat{\sigma}_\phi^2 + \left(\sum_{m_j} \hat{\sigma}_{jt}^2\right)^{-1}} \tag{8}$$

Where:

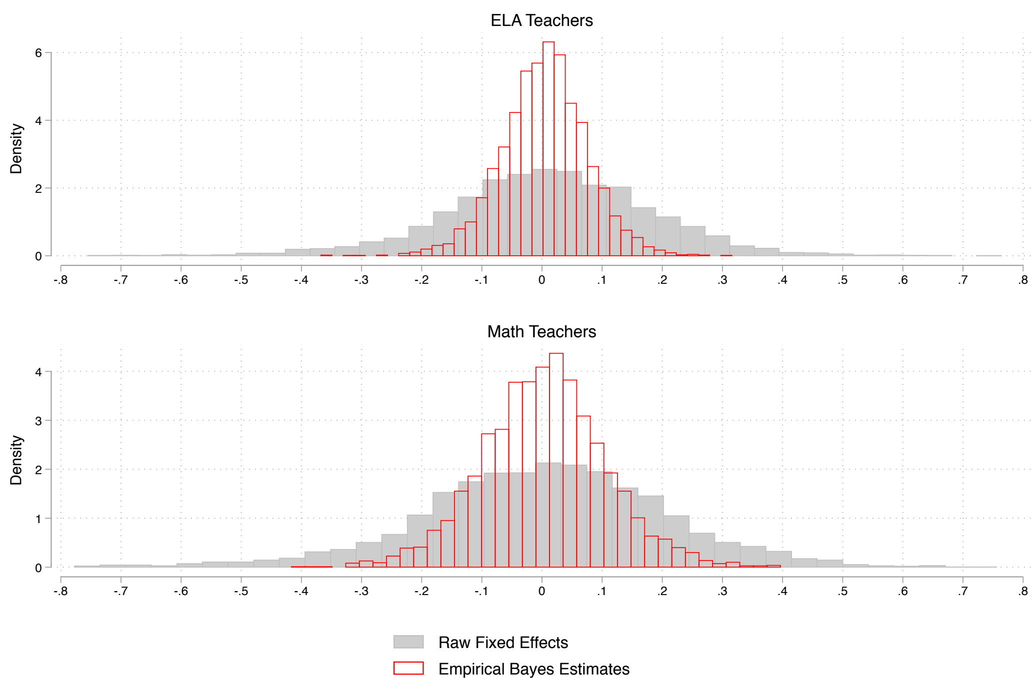
$$\overline{\xi}_j = \sum_t \overline{\xi}_{jt} \times \frac{\hat{\sigma}_{jt}^2}{\sum_l \hat{\sigma}_{jl}^2} \tag{9}$$

$$\hat{\sigma}_{jt}^2 = \left(\hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\xi^2}{N_{cj}}\right)^{-1} \tag{10}$$

In equations (8), (9), and (10), the teacher-level variance $\hat{\sigma}^2_\phi$ corresponds to the covariance in classroom-level average residuals for the same teacher over time $\hat{\sigma}^2_\phi = cov(\overline{\xi}_{jct}, \overline{\xi}_{jc't'})$. We estimate the student-level idiosyncratic variance $\hat{\sigma}^2_\epsilon$ 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: $\hat{\sigma}^2_\theta = Var(\xi_{ijst}) - \hat{\sigma}^2_\phi - \hat{\sigma}^2_\epsilon$.

Figure A.7 shows the distribution of the raw fixed effects ($\hat{\phi}_j$) and the Empirical Bayes estimates ($\hat{\phi}^{EB}_j$) for Math and ELA teachers.

Figure A.7: Distribution of Empirical Bayes Estimates



Notes: This figure shows the distribution of raw teacher fixed effects and shrunken empirical Bayes estimates obtained from equation (6). We construct empirical Bayes estimates following Kane and Staiger (2008). See section A.2 for details.