Licensure Tests and Teacher Supply in Connecticut:

Technical Appendix

Alexis Orellana and Marcus A. Winters

BU

Boston University Wheelock College of Education & Human Development
Wheelock Educational Policy Center

WEPC
WHEELOCK EDUCATIONAL POLICY CENTER

Appendices

A Data

A.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.

A.1.1 Subject-Matter Certification Test: Praxis II

We focus our analysis on 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 1 shows the link between some of the endorsement codes offered in Connecticut and the Praxis II tests required. Some endorsement codes involve passing more than one test (for example, *Elementary Grades, K-6*). In these cases, we group all sub-tests and employ the minimum score as the forcing variable in the analysis described in Section D.[1]

A.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 a renewable Initial or Provisional Educator Certificate.[2] 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

Table (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.

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.[3] 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.

A.3 Employment Records

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

A.4 Additional Teacher and Student Administrative Data

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

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

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

B Summary of Data Matching Process

To link students and teachers we match the TCS course offering and TCS student course grade datasets. To merge both datasets, we construct a classroom identifier as a unique *year-school-NCES code-section-start date-end date* combination. We do not consider year 2019-20 when the Smarter Balanced assessments were not administered.

We restrict our analysis to the set of classrooms assigned to educators who hold a valid *ein* 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.

Table 2 summarizes the number of students taking the math and language Smarter Balanced assessments (SB) between 3rd and 8th grades and the proportion of students we successfully match to a classroom with one teacher.

C Estimating Relationship Between Licensure Scores and Value-Added

We use a two-step approach to investigate the relationship between a teacher's score on the respective Praxis exam and their later impact on student achievement. The first stage estimates the teacher's independent contribution to student test scores, on average, commonly referred to as the

Table (2) Distribution of matches between 2014-2020

Year	Not Matched	Matched	Students taking ELA or math SB tests
2014	5.2%	94.8%	235,497
2015	3.1%	96.9%	234,993
2016	2.8%	97.2%	234,759
2017	3.1%	96.9%	233,465
2018	5.7%	94.3%	231,109
2019			0
2020	13.9%	86.1%	214,291
Total	5.5%	94.5%	1,384,114

teacher's "value-added". The second stage then measures the association between the teacher's value-added as estimated from the first stage and their Praxis score.

For the first-stage analysis, we use a conventional value-added approach to produce an estimate for each teacher's impact on student test scores. The general model takes the form:

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

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.

For each teacher we capture $\hat{\phi}_j$, which is our estimate for each teacher's contribution to student test scores conditional on the other covariates. A common challenge with value-added approaches is that for any individual teacher, the sample size used to identify the relevant fixed effect may be quite small and hence estimated with a substantial degree of noise. We address this issue by employing the Bayesian Shrinkage adjustment, as is typical in the value-added literature.

That we control for prior test scores focuses the model on estimating teacher impacts on student test scores gains. 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.

Equation 1 represents our base value-added model, which we use as our primary estimate for teacher value-added for which we report results in the main body of the paper. However, results are similar from models that incorporate various fixed effects for schools or school-by-year and from models that remove the function for prior test scores and rather incorporate a student fixed-effect.

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 teacher's estimated value-added from the first stage, $\hat{\phi}_j$, and independent variables include the teacher's score on the respective Praxis exam (P_j) and a vector of time-invariant teacher characteristics (gender, race/ethnicity, and education level indicators) represented by Z'_j. Formally:

$$\hat{\phi}_j = P_j\gamma + Z'_j\delta + \eta_j \tag{2}$$

We are primarily interested in the estimate for γ, which represents the relationship between the teacher's score on the Praxis II exam and their estimated value-added contribution to student test scores. We estimate equation 2 separately for Language, Math, and elementary teachers.

Figure 1 shows the association between licensure score and our empirical Bayes estimates of test score value-added by subject and test-type. For both subjects, we find a small, positive relationship. For ELA teachers, a standard deviation increase in Praxis II score associates with a gain of 0.0052σ in test scores. For math, a standard deviation increase in Praxis II scores associates with a gain of 0.0081σ.

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

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 a given test that occurs at the designated cutoff.

Let i denote an applicant taking test j for the first time. Each test j has a minimum passing score $\bar{x}_j$. We center scores around the corresponding cutoff and standardize them using the within-sample standard deviation.[4] We denote this variable x_{ij}. When a test j considers more than one subtest, we define x_{ij} as the minimum value across all sub-tests. We account for changes in the tests over time and differences between different 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:

$$y_{ij} = \alpha + f(x_{ij}) + \beta \mathbb{1}(x_{ij} < 0) + \phi_j + \phi_t + \epsilon_{ij} \tag{3}$$

The term $f(x_{ij})$ is a parametric function of the (normalized) score obtained by applicant i, which our primary model employs as quadratic and allows for changes in the slope at the cutoff value.[5] 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. The sample includes an individual's first observed score on the relevant licensure test, excluding first-time test-takers who we previously observe teaching within a Connecticut public school. 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 assumptions unlikely. Figure 2 shows no indication of manipulation around the cutoff values. We find no statistical evidence to reject the null hypothesis of continuity around the passing threshold. The p-value of the discontinuity test is 0.35.

The second threat to identification is the potential for discontinuities in the value for confounders around the threshold. To investigate the potential for this threat, authors typically look for balance in the value of observed baseline characteristics on either side of the threshold. Unfortunately, such 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.[6]

Table 3 reports regression discontinuity estimates for the effect of failing the first administration of Praxis II. For those scoring at the threshold, failing the first administration of Praxis II reduced the likelihood that a candidate obtained any teaching certification by about 6.6 percentage points. Failing the test required for endorsement to teach a STEM subject reduced the likelihood of obtaining the endorsement by about 8.9 percentage points. And failing the test for endorsement to teach within special education reduced the likelihood of obtaining that endorsement by 10.7 percentage points.

Figure (1) Association between Praxis II scores and Teacher Value-Added

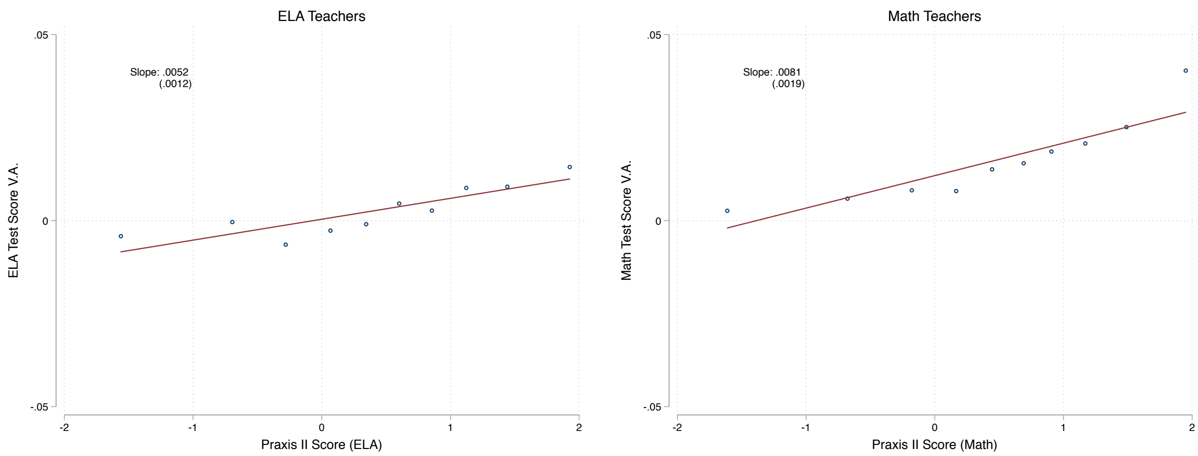
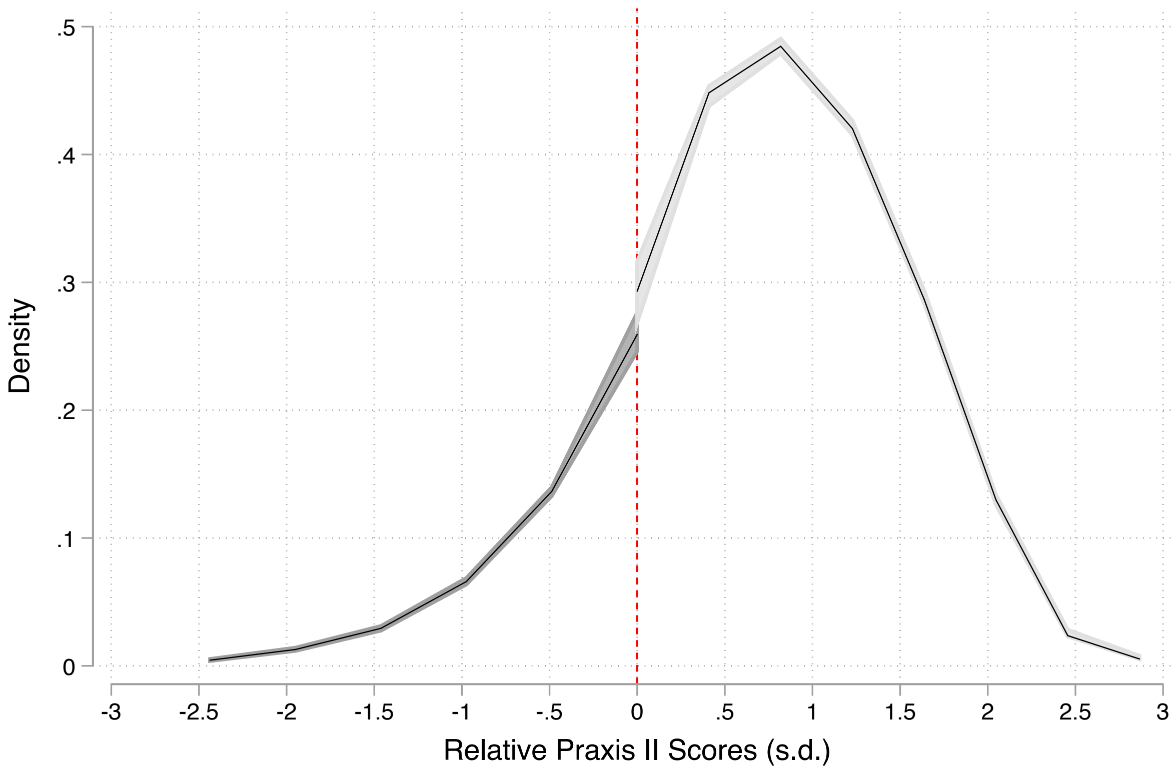


Table (3) RD Estimates for Effect of Failing First Administration of Licensure Test

	(1) Any Certification	(2) STEM	(3) Special Education
Failed Praxis II	-0.066***	-0.089***	-0.107**
	(0.013)	(0.033)	(0.042)
Average Outcome	0.79	0.68	0.82
Bandwidth	(-0.58,0.78)	(-0.51,0.69)	(-0.63,0.69)
N	34,307	6,207	3,425

Notes: This table presents estimates of the effects of failing the first attempt at Praxis II on the likelihood of eventually obtaining any teaching certification, obtaining an endorsement to teach within a STEM subject, and endorsement to teach special education. Analyses of STEM and special education endorsements are restricted to the first administration of a test associated with that particular endorsement, rather than the first Praxis II attempt. Bandwidths are selected following Calonico et al. (2014) and 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure (2) Density of Praxis II tests around the threshold



Notes: This figure illustrates the density of standardized Praxis II 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 0.35. These values imply there is no statistical evidence to reject the null hypothesis of no discontinuity at the threshold.

Notes

[1]Table 1 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.

[2]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 who 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.

[3]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.

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

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

[6]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.

References

Calonico, S., Cattaneo, M., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82:2295–2326.

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

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.

OUR MISSION

The Wheelock Educational Policy Center (WEPC) conducts and disseminates rigorous, policy-relevant education research in partnership with local, state, and federal policymakers and stakeholders to improve pk–20 educational opportunities and holistic outcomes for underserved students.

www.wheelockpolicycenter.org
wheelockpolicy@bu.edu

Boston University Wheelock College of Education & Human Development
Wheelock Educational Policy Center

WEPC
WHEELOCK EDUCATIONAL POLICY CENTER