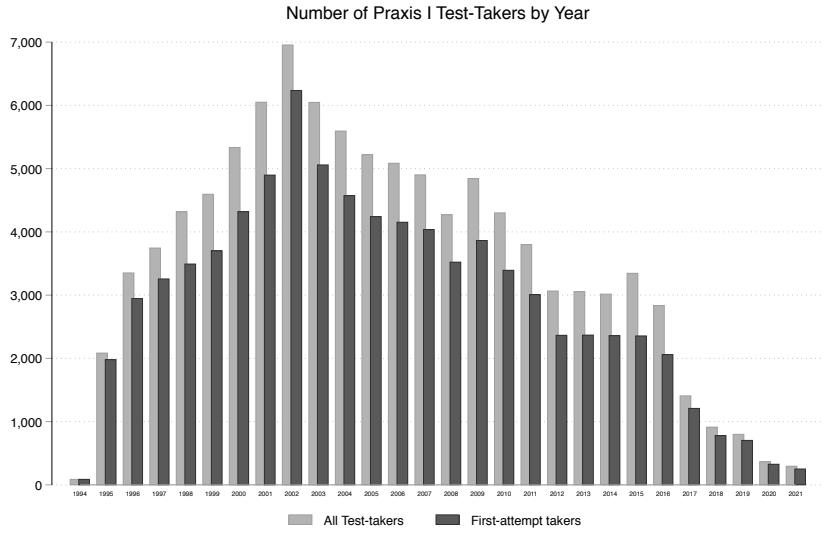
A Online Appendix

Licensure Tests and Teacher Supply

Alexis Orellana & Marcus A. Winters

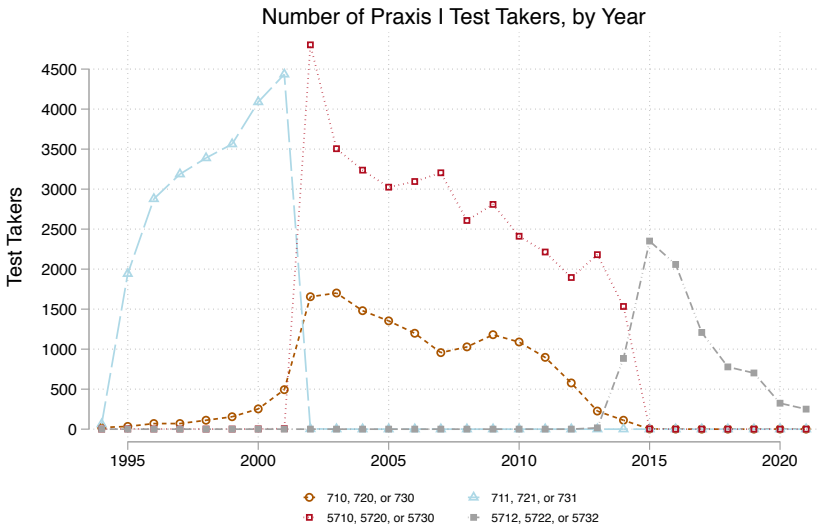
A.1 Additional Figures

Figure A.1: Praxis I Test-takers



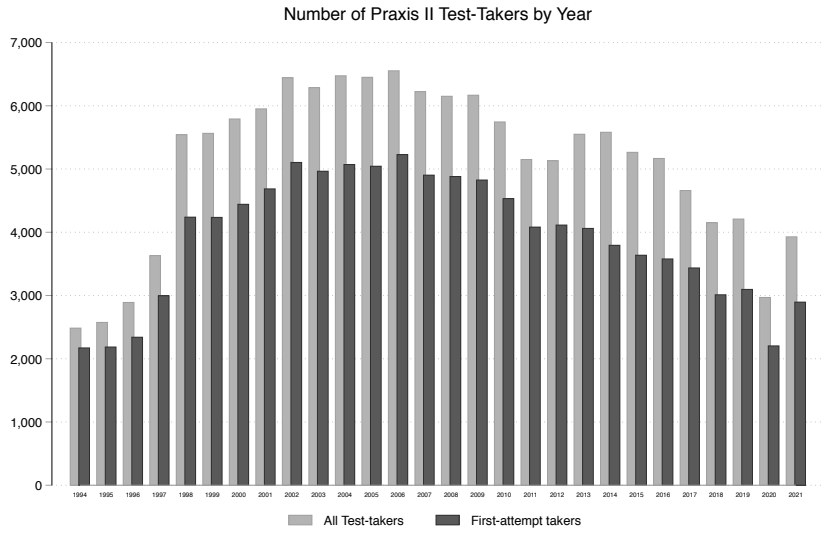
Notes: This plot shows the number of applicants who took a Praxis I test between 1995 and 2021. Gray and black bars represent the total number of test-takers and the total number of first-time applicants in each year, respectively.

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.

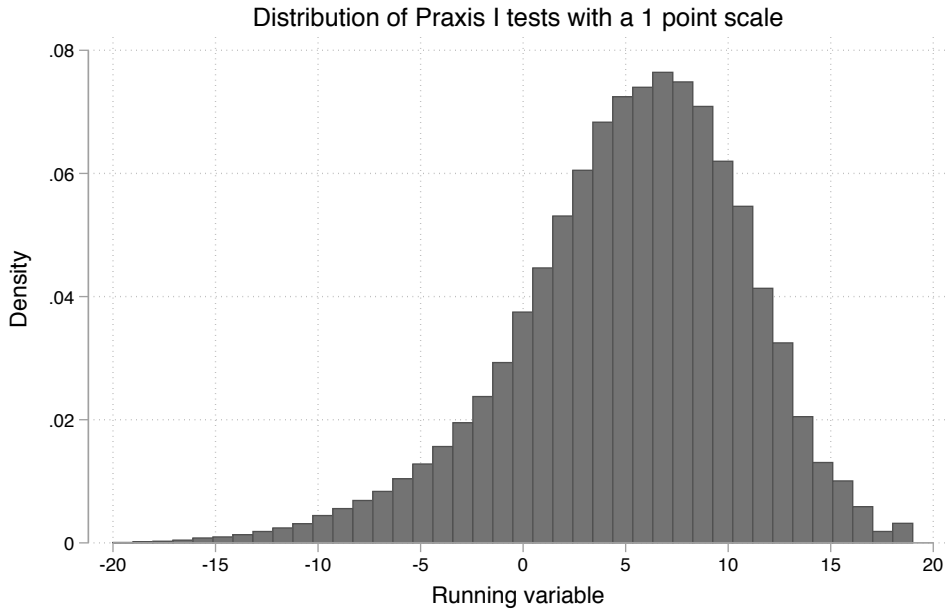
Figure A.3: Praxis II Test-takers



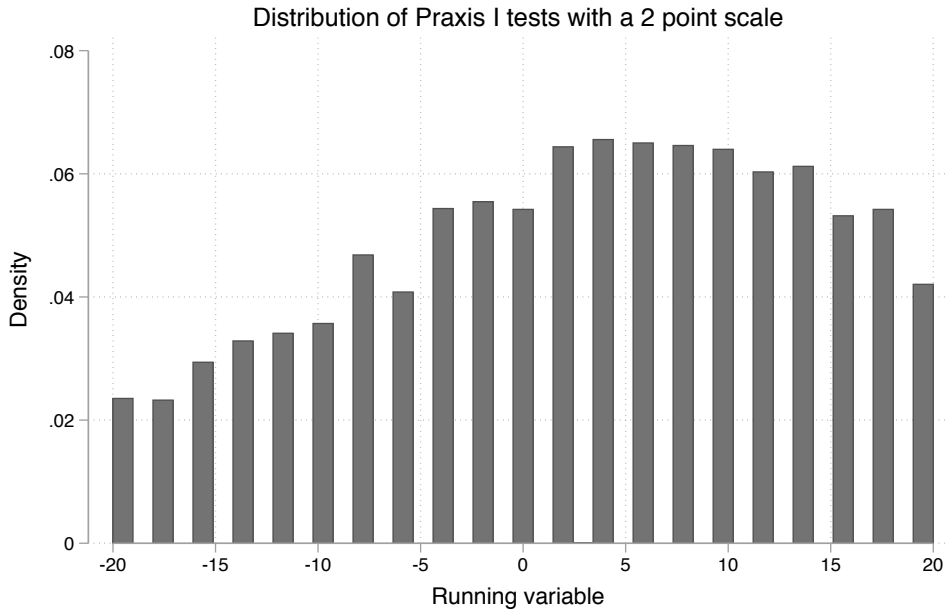
Notes: This plot shows the number of applicants who took a Praxis II test between 1995 and 2021. Gray and black bars represent the total number of test-takers and the total number of first-time applicants in each year, respectively.

Figure A.4: Distribution of Praxis I Scores

(a) Tests using a 1-point scale



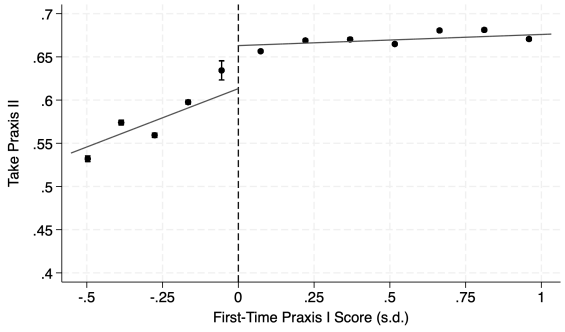
(b) Tests using a 2-point scale



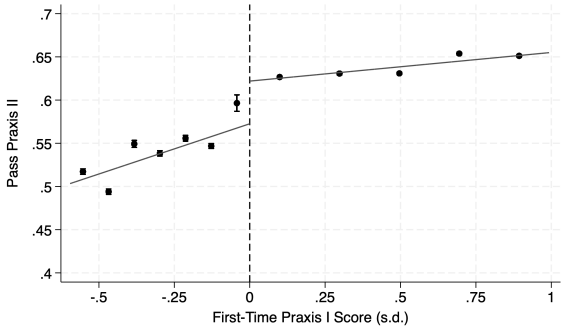
Notes: This figure shows the distribution of the running variable for Praxis I tests between 1995 and 2021. Each panel shows the distribution for tests using a one-point and two-point scale, respectively.

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

(a) Take Praxis II

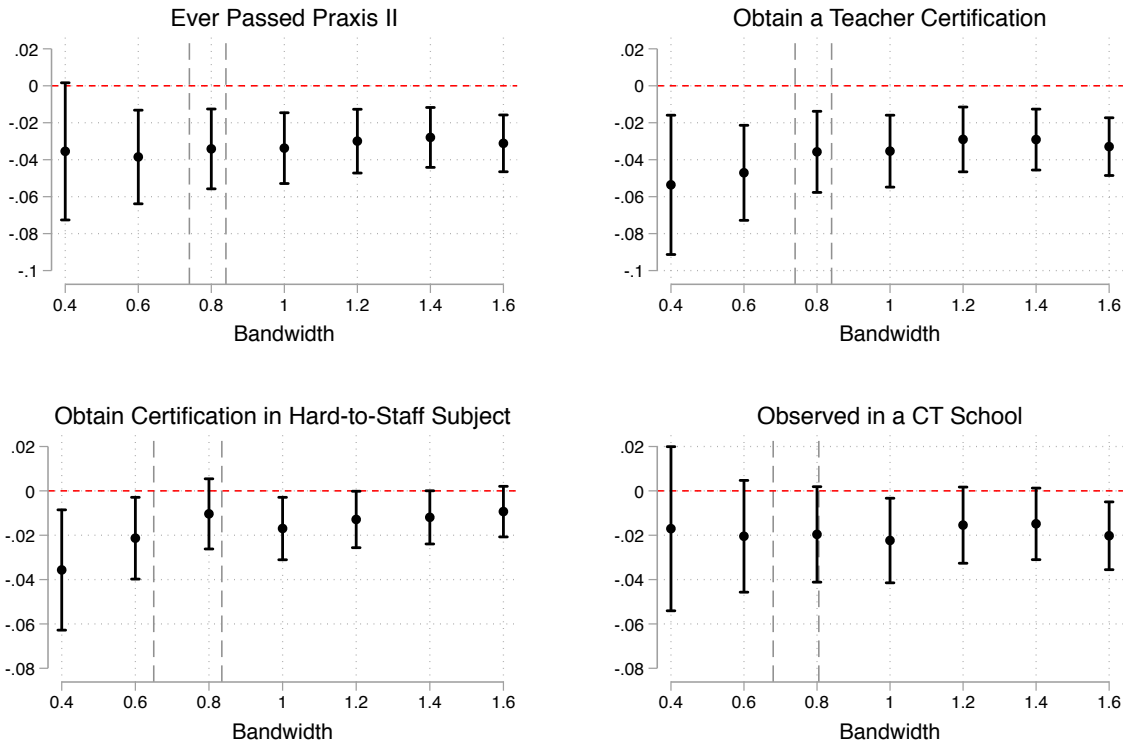


(b) Pass Praxis II



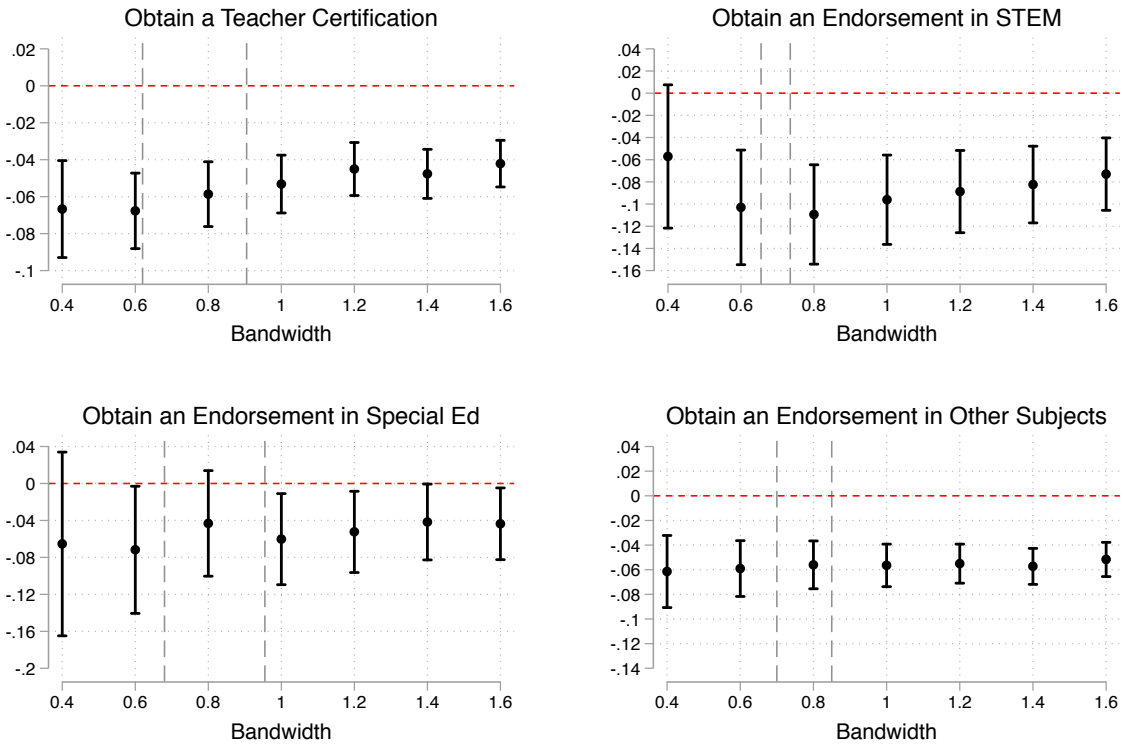
Notes: This figure includes additional outcomes to the ones presented in Figure 2 showing the relationship between failing a Praxis I test and subsequent outcomes. Each regression employs CCT optimal bandwidths (Calonico et al., 2014) and a triangular kernel. Observations binned according to the IMSE-optimal 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 Praxis I: Alternative Bandwidth



Notes: This figure presents estimates from Equation (1) using different bandwidth choices. The x-axis corresponds to different bandwidths used to compute each estimate. The points illustrate the estimated effect, and the lines denote the 95% confidence intervals. The dotted lines show the absolute value of the CCT optimal bandwidths (Calonico et al., 2014).

Figure A.7: Nonparametric Estimates for the Effect of Failing First Attempt on Praxis II: Alternative Bandwidth

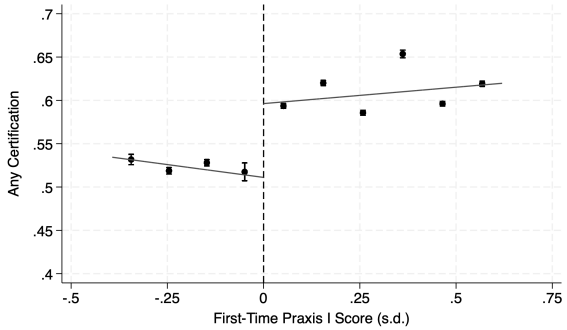


Notes: This figure presents estimates from Equation (1) using different bandwidth choices. The x-axis corresponds to different bandwidths used to compute each estimate. The points illustrate the estimated effect, and the lines denote the 95% confidence intervals. The dotted lines show the absolute value of the CCT optimal bandwidths (Calonico et al., 2014).

Figure A.8: Nonparametric Estimates for the Effect of Failing First Attempt on Licensure Test: Using an Uniform Kernel

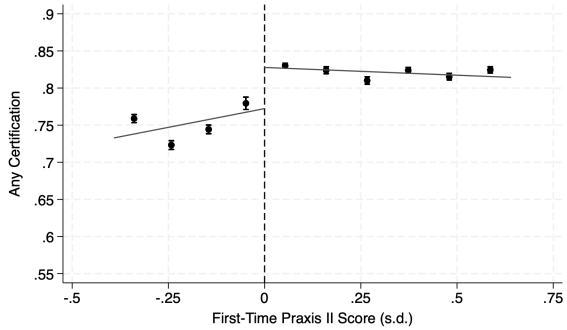
Praxis I:

(a) Obtain Any Certification

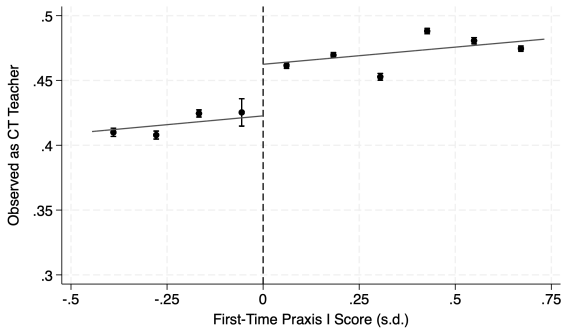


Praxis II:

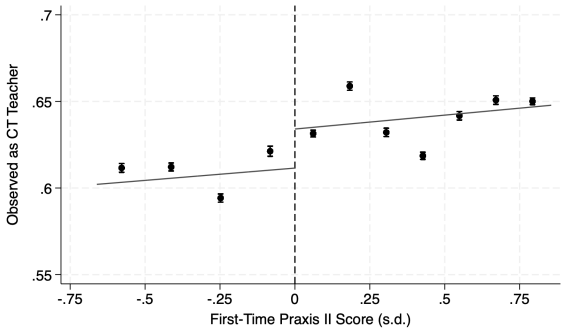
(b) Obtain Any Certification



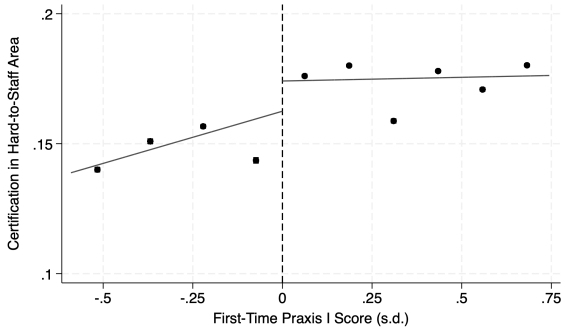
(c) Teach Within a Connecticut Public School



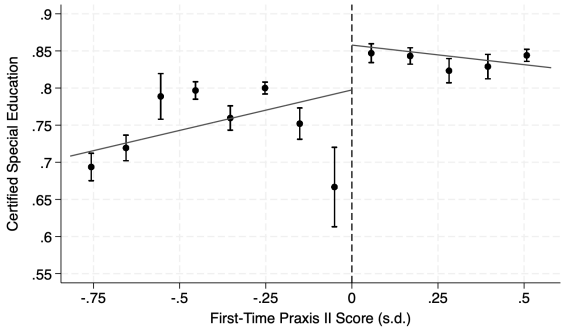
(d) Teach Within a Connecticut Public School



(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.

A.2 Additional Tables

Table A.1: Characteristics of the Connecticut Teacher Workforce

	(1) CT	(2) CT Quintile Among States	(3) US Mean	(4) US Std. Dev.
Starting Salary	$45,840	3rd	$44,530	$4,109.2
Wage Competitiveness	81.9%	4th	73.6%	6.9%
Stayed Teaching in Same School	83.3%	3rd	84.1%	5.7%
Left Teaching	7.8%	2nd	7.9%	2.8%
School Vacancies Unfilled or Hard to Fill	47%	3rd	46.9%	10.4%
Uncertified Teachers	1.3%	5th	3.7%	5.4%
Change in TPP Completers Past 5 Years	-2.1%		4.1%	26.8%
Change in Enrollment Past 5 Years	-2.5%		-2.3%	3.0%
Change in # of Teachers Past 5 Years	2.8%		1.7%	3.1%

Notes: This table presents different statistics of the Connecticut (CT) teacher workforce compared to the national level. Source: Learning Policy Institute (2024)

Table A.2: Shortage Areas Reported by Connecticut: 2023-24

Subject Matter	# States Reporting Shortage
English as a Second Language	17
Special Education	43
Science	39
Mathematics	37
Career & Technical Ed	30
World Languages	26
Social Studies	23

Notes: This table lists the subjects identified by Connecticut as teacher shortage areas for the 2023-24 school year, along with the number of states that reported the same areas. Source: U.S. Department of Education: Teacher Shortage Areas Report: https://tsa.ed.gov/#/home/

Table A.3: 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.4: 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.5: 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. CK: content knowledge. 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.6: Effect of Failing Licensure Test on Certification and Teaching Type

	(1) Standard Certificate	(2) Provisional/Interim Certificate	(3) Teacher with Certification	(4) Certified but does not teach	(5) Teacher with no Certification
			Panel A: Praxis I		
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 I and 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7: RD Estimates for Effect of Failing Praxis I on Praxis II Outcomes

	Panel A: Ever take a Praxis II test									
	Any Praxis II	Special Education	Foreign Languages	ESOL	STEM	Elementary Grades	Art and Music	English	History	Other Subjects
Failed	-0.056***	0.007	-0.002	-0.005	-0.034***	-0.030*	0.019**	0.008	-0.010	-0.015
	(0.019)	(0.013)	(0.004)	(0.004)	(0.007)	(0.018)	(0.008)	(0.009)	(0.008)	(0.010)
Average Outcome	0.649	0.117	0.009	0.009	0.064	0.296	0.041	0.055	0.058	0.074
Bandwidth					(-0.566,1.036)					
N					42,314					

	Panel B: Ever passed a Praxis II test									
	Any Praxis II	Special Education	Foreign Languages	ESOL	STEM	Elementary Grades	Art and Music	English	History	Oher Subjects
Failed	-0.051***	0.003	-0.003	-0.004	-0.027***	-0.030*	0.018**	0.008	-0.005	-0.012
	(0.018)	(0.012)	(0.003)	(0.004)	(0.006)	(0.016)	(0.008)	(0.007)	(0.007)	(0.009)
Average Outcome	0.613	0.114	0.007	0.009	0.055	0.295	0.039	0.045	0.049	0.073
Bandwidth					(-0.603,1.004)					
N					42,015					

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

Table A.8: 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					
	(7) Any Certification	(8) STEM	(9) Special Ed	(10) Other Subjects	(11) Teach	(12) 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.9: RD Estimates for Years Observed Teaching

	Panel A: Praxis I			
	(1) Number of Years	(2) Tenure > 3 Years	(3) Tenure > 5 Years	(4) Observed Value-Added
Failed	-0.379	-0.050*	-0.048	-0.026*
	(0.397)	(0.029)	(0.035)	(0.015)
Average Outcome	9.55	0.90	0.82	0.11
Bandwidth	(-0.37,0.90)	(-0.37,0.80)	(-0.41,1.01)	(-0.78,0.54)
N	16,302	13,735	15,519	12,986
	Panel B: Praxis II			
Failed	-0.143	-0.021*	-0.006	-0.004
	(0.171)	(0.011)	(0.015)	(0.012)
Average Outcome	9.37	0.89	0.81	0.12
Bandwidth	(-0.80,0.79)	(-0.83,0.69)	(-0.81,0.88)	(-0.72,1.22)
N	23,625	20,022	22,244	31,416

Notes: This table presents RD estimates investigating discontinuities at the passing threshold of Praxis I and Praxis II in the relationship between licensure score on additional outcomes. Columns (1)-(3) show estimates for the number of years observed as a teacher and indicators if the number of years exceeds three and five, respectively. Column (4) show estimates of an indicator equals to one if the test-taker is included in the sub-sample used to estimate test score teacher value-added. All outcomes are conditional on being observed as a teacher in the sample. 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.10: Distribution of Endorsements for Different Samples

Endorsement	Valid Endorsement	Observed Teaching	English Value-Added	Math Value-Added
Elementary Grades K-6	39.3%	38.3%	71.3%	79.5%
English 7-12	9.5%	10%	21%	2.2%
English Middle School 4-8	1%	1.1%	3.7%	0.9%
Mathematics 7-12	6.6%	7.5%	1.3%	11.2%
Mathematics Middle School 4-8	1.9%	2.3%	1.9%	9.2%
History/Social Studies 7-12	9.3%	8.6%	3.3%	2.4%
History/Social Studies 4-8	0.8%	0.9%	1.1%	0.7%
Special Education K-12	16.5%	17.8%	7.3%	7.7%
All other STEM	7.6%	8.6%	2%	1.8%
Number of individuals	68,808	50,903	4,055	3,533

Notes: This table presents the distribution of certifications across different samples. Column (1) displays all test takers gaining certification, column (2) restricts this sample to individuals observed teaching in a public school in Connecticut. Columns (3) and (4) restrict the sample to individuals used to estimate teacher value-added on English and Math test scores, respectively. The category All other STEM includes Biology 7-12, Chemistry 7-12, Physics 7-12, Earth Science 7-12, and General Science 7-12.

Table A.11: RD Estimates for Effect of Failing First Administration of Licensure Test: English Test Takers

	Praxis II	Endorsement			Teaching		
	Take Different Test	Any Subject	English	Non-English	Special Ed	Foreign	TESOL
Failed	0.015	-0.141***	-0.151***	0.010	-0.018	-0.007	-0.007
	(0.027)	(0.030)	(0.032)	(0.017)	(0.011)	(0.006)	(0.005)
Average Outcome	0.209	0.759	0.712	0.047	0.029	0.009	0.008
Bandwidth			(-1.006,0.876)				
N			6,324				

	Teaching						
	STEM	Elementary	Music	English	History	Other	No Teaching
Failed	-0.001	-0.006	0.003	-0.034	0.014	0.032**	0.052
	(0.009)	(0.012)	(0.007)	(0.032)	(0.010)	(0.013)	(0.034)
Average Outcome	0.015	0.028	0.009	0.552	0.027	0.037	0.387
Bandwidth			(-1.006,0.876)				
N			6,324				

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

Table A.12: RD Estimates for Effect of Failing First Administration of Licensure Test: Art-Music Test Takers

	Praxis II	Endorsement			Teaching		
	Take Different Test	Any Subject	Art-Music	Non-Art-Music	Special Ed	Foreign	TESOL
Failed	0.122**	-0.124*	-0.170**	0.046*	-0.007	0.004	0.008
	(0.057)	(0.073)	(0.075)	(0.027)	(0.011)	(0.009)	(0.008)
Average Outcome	0.129	0.811	0.8	0.01	0.009	0.002	0.002
Bandwidth			(-0.362,0.938)				
N			2,866				

	Teaching						
	STEM	Elementary	Music	English	History	Other	No Teaching
Failed	0.063*	0.019	-0.046	0.019	-0.001	0.031	-0.001
	(0.033)	(0.023)	(0.078)	(0.015)	(0.003)	(0.022)	(0.076)
Average Outcome	0.015	0.02	0.615	0.005	0.003	0.01	0.358
Bandwidth			(-0.362,0.938)				
N			2,866				

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

B 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:

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

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.

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 B.1 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:

$$\hat{\phi}_j = \delta_0 + \delta_1 P_j + \eta_j \tag{B.2}$$

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 (B.1) 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:

$$\hat{\phi}_j^{EB} = \overline{\xi}_j \times \frac{\hat{\sigma}_\phi^2}{\hat{\sigma}_\phi^2 + \left(\sum_{m_j} \hat{\sigma}_{jt}^2\right)^{-1}} \tag{B.3}$$

Where:

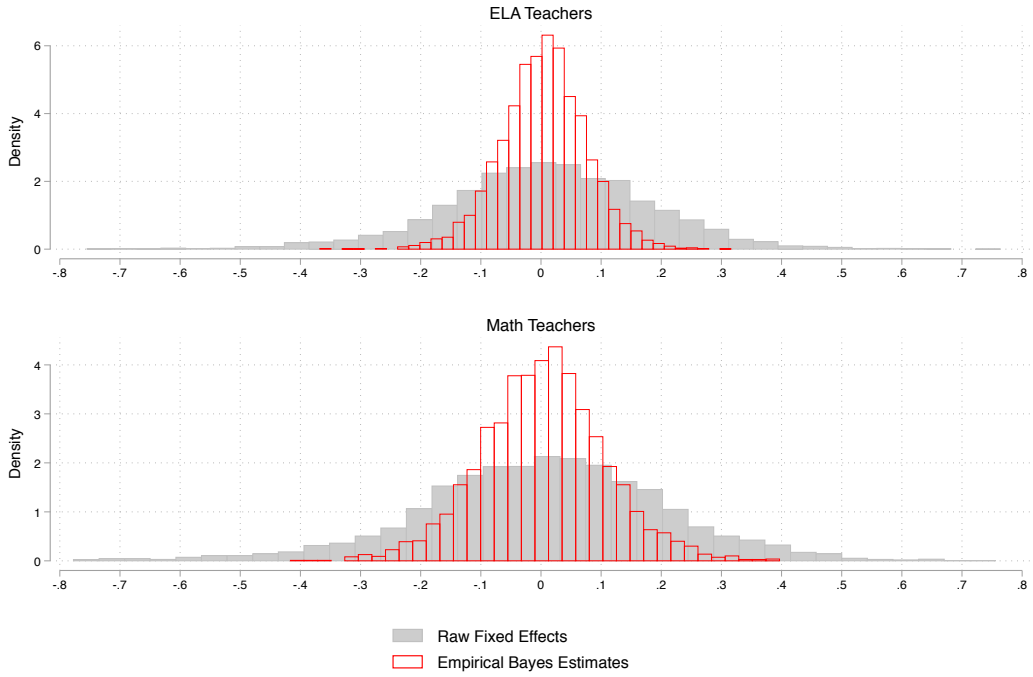
$$\overline{\xi}_j = \sum_t \overline{\xi}_{jt} \times \frac{\hat{\sigma}_{jt}^2}{\sum_l \hat{\sigma}_{jl}^2} \tag{B.4}$$

$$\hat{\sigma}_{jt}^2 = \left(\hat{\sigma}_\theta^2 + \frac{\hat{\sigma}_\xi^2}{N_{cj}}\right)^{-1} \tag{B.5}$$

In Equations (B.3), (B.4), and (B.5), the teacher-level variance $\hat{\sigma}_\phi^2$ corresponds to the covariance in classroom-level average residuals for the same teacher over time $\hat{\sigma}_\phi^2 = cov(\overline{\xi}_{jct}, \overline{\xi}_{jc't'})$. We estimate the student-level idiosyncratic variance $\hat{\sigma}_\epsilon^2$ as the variance in within-classroom deviations in student outcomes. Finally, we estimate the variance of classroom-level shocks as the remainder of the total variation: $\hat{\sigma}_\theta^2 = Var(\xi_{ijst}) - \hat{\sigma}_\phi^2 - \hat{\sigma}_\epsilon^2$.

Figure B.1 shows the distribution of the raw fixed effects ($\hat{\phi}_j$) and the Empirical Bayes estimates ($\hat{\phi}_j^{EB}$) for Math and ELA teachers.

Figure B.1: Distribution of Empirical Bayes Estimates



Notes: This figure shows the distribution of raw teacher fixed effects and shrunken empirical Bayes estimates obtained from Equation (B.1). We construct empirical Bayes estimates following Kane and Staiger (2008). See section B for details.

C Potential for a Homogeneous Treatment Effect to Produce Differential Selection by Latent Value-Added

Each licensure test-taker i is endowed with some amount of latent value-added, θ_i^*, which is normally distributed with mean μ and standard deviation σ.

$$\theta_i^* \sim N(0,1) \tag{C.1}$$

Each licensure test-taker also achieves an initial licensure score, S_i, which is a noisy measure of θ_i^*. We allow for correlation between θ_i^* and S_i using the coefficient $\rho_{S\theta}$.

$$\begin{aligned} S_i = \theta_i^* \; + \; \sqrt{\sigma_S^2}\,\epsilon_i, \quad \epsilon_i \sim N(0,1) \text{ (independent of } \theta_i^*), \\ \text{where } \sigma_S^2 \; = \; \frac{1}{\rho_{S\theta}^2} \; - \; 1 \quad \big(\mathrm{Corr}(S_i, \theta_i^*) = \rho_{S\theta}\big). \end{aligned} \tag{C.2}$$

Individuals fail the test if their initial score is below a cutoff, κ.

$$\text{fail}_i = \mathbf{1}\{S_i < \kappa\} \tag{C.3}$$

The first key assumption underlying RD analysis is that among the population of test-takers the relationship between initial licensure score and latent value-added is smooth at the passing threshold, conditional on the forcing variable. Because licensure score is correlated with latent value-added, we would expect a naive comparison to find that average latent value-added for those who fail the test to be lower than average latent value-added among those who pass. However, since there are no factors correlated with both latent value-added and failing other than licensure score, there should be no difference in average latent value-added after conditioning on licensure score. That is, in the below equation we would anticipate to find $\beta_2 = 0$.

$$\theta_i^* = \alpha \; + \; \beta_1\, S_i \; + \; \beta_2\, \text{fail}_i \; + \; \varepsilon_i \tag{C.4}$$

However, not all licensure test-takers will become teachers and thus achieve an observed value-added score. Some percentage of licensure test-takers will not become a teacher, independent of the treatment effect. We operationalize such "natural attrition" with an index score, A_i. Importantly, we allow for the possibility that natural attrition is positively correlated with latent value-added, implying that without intervention initial test-takers with higher latent value-added are either as likely or less likely to eventually become a teacher. For instance, this would occur if those with higher latent value-added may have more attractive opportunities in the outside labor market. We model this relationship as:

$$A_i = \theta_i^* \; + \; \sqrt{\sigma_A^2}\,\eta_i, \quad \eta_i \sim N(0,1) \text{ (independent of } \theta_i^*, \epsilon_i), \tag{C.5}$$
$$\text{where } \sigma_A^2 \; = \; \frac{1}{\rho_{A\theta}^2} \; - \; 1 \quad \big(\mathrm{Corr}(A_i, \theta_i^*) = \rho_{A\theta}\big).$$

The individual's total attrition index score, ℓ_i, is the log-odds index for attrition, capturing both their natural attrition index and a uniform penalty, τ, imposed on all who fail the test.

$$\ell_i = A_i \; + \; \tau \cdot \text{fail}_i \tag{C.6}$$

Attrition takes the form of a Bernoulli draw from a logistic distribution.

$$p_i = \Pr(\text{attrit}_i = 1) = \Lambda(\ell_i) = \frac{1}{1 + \exp\big(-\ell_i\big)} \tag{C.7}$$

$$\text{attrit}_i \sim \mathrm{Bernoulli}(p_i) \tag{C.8}$$

We observe i's value-added, θ_i, if and only if they survive the process (i.e., do not attrit) and thus become a teacher.

$$\theta_i = \begin{cases} \theta_i^* & \text{if attrit}_i = 0 \\ . & \text{if attrit}_i = 1 \end{cases} \tag{C.9}$$

To apply the RD design we regress observed value-added on fail_i and S_i in the post-

attrition sample. This regression effectively gives us the difference in the average observed value-added for those with the same initial licensure score on the passing and failing side of the threshold. Note that for the purposes of this explanation we assume that we observe latent value-added directly, while in practice we observe an estimated value-added. The description holds as long as observed value-added is an unbiased measure of latent value-added.

$$\theta_i^* = \alpha \; + \; \beta_1 \, S_i \; + \; \beta_2 \, \text{fail}_i \; + \; \varepsilon_i, \quad \text{(for individuals with attrit}_i = 0) \tag{C.10}$$

However, unlike when comparing average latent value-added across the threshold among all licensure test-takers, even though the penalty to the log-odds of attrition from failing the test is uniform, the non-random attrition process described above can nonetheless lead us to find $\beta_2 < 0$. Such negative selection can occur in light of a homogeneous treatment effect because of the correlation between latent value-added and natural attrition, and the non-linearity of the attrition process.

Consider two broad sets of students who achieve a licensure score $S_i < \kappa$. Recall that the licensure score an individual achieves is a function of both their latent value-added and a stochastic error term. The first group is comprised of individuals with low latent value-added who fail "naturally" or with the help of a moderate or mild shock that pushes their score below the threshold. The second group is comprised of individuals with relatively high latent value-added who nonetheless achieved a failing score because they experienced an especially large negative shock.

Because of the positive correlation between latent value-added and natural attrition, those unlucky high-θ_i^* failers also tend to have a higher baseline natural attrition index. Because the logistic function is non-linear, when the failing penalty is applied equally to the log-odds of all who failed the test, it will tend to push those with higher natural attrition (and thus systematically higher latent value-added) into an attrition probability range that can be much higher than those with the same licensure score but lower natural attrition (and thus systematically lower latent value-added). The effect here can be that relative to a scenario where there is only natural attrition, the

introduction of a homogeneous treatment effect can pull downward the average latent value-added for those who failed but did not attrit.

The potential for a homogeneous treatment effect to disproportionately push out those with higher latent value-added hinges on the strength of the correlation between latent value-added and natural attrition, and the correlation between latent value-added and licensure score. A stronger correlation between latent value-added and licensure score will tend to attenuate negative selection among failers by reducing the number of "unlucky" high value-added individuals who fail due to a random shock. In contrast, a stronger correlation between latent value-added and natural attrition will tend to exacerbate negative selection by further magnifying the dropout probability for those with high latent value-added who nonetheless fail the test due to a random negative shock to their licensure score.

C.1 Differential Treatment Response Related to Latent Value-Added

Finally, notice that we would also observe negative selection if those with higher latent value-added are more responsive to the treatment. Working from the above framework, we add to the total attrition score an interaction between latent value-added and the indicator for failing the test.

$$\ell_i = A_i \; + \; \tau \, \text{fail}_i \; + \; \delta\big(\theta_i^* \times \text{fail}_i\big) \tag{C.11}$$

The additional term, $\delta\big(\theta_i^* \times \text{fail}_i\big)$, differentially shifts the attrition score, allowing for a heterogeneous treatment effect of failing by latent ability. Notice that $\delta > 0$ would increase the negative selection on observed value-added following attrition. It would also produce negative selection even if there were no natural attrition, or if natural attrition is not correlated with latent value-added.

C.2 Simulations

We conduct a Monte Carlo simulation where we estimate β_2 from Equation (10) under different assumptions of the correlation between value-added and natural attrition ($\rho_{A\theta}$) and between value-

added and licensure scores ($\rho_{S\theta}$). We perform 1,000 simulations, obtaining 10,000 individual draws of $\{\theta_i^*, S_i, A_i\}$ in each iteration conditional on the parameters of the data-generating process (κ, τ, δ). In the tables presented below, we show the mean and standard deviation of the distribution of estimated $\hat{\beta}_2$ under different values of $\rho_{S\theta}$ (rows) and $\rho_{A\theta}$ (columns).

C.2.1 Case 1: $\delta = 0$

We consider the following parameter values: $\kappa = -1, \tau = 0.75$

Table C.1: Mean (St.Dev.) of $\hat{\beta}_2$ from 1,000 Iterations of Simulation 1

	Correlation between VAM and Attrition ($\rho_{A\theta}$):			
Correlation between VAM and Scores ($\rho_{S\theta}$):	$\rho_{A\theta} = 0$	$\rho_{A\theta} = 0.1$	$\rho_{A\theta} = 0.5$	$\rho_{A\theta} = 0.9$
$\rho_{S\theta} = 0$	-0.002 (0.046)	-0.006 (0.047)	-0.058 (0.044)	-0.075 (0.043)
$\rho_{S\theta} = 0.1$	0.002 (0.046)	-0.005 (0.050)	-0.055 (0.045)	-0.071 (0.045)
$\rho_{S\theta} = 0.5$	0.000 (0.040)	-0.005 (0.043)	-0.043 (0.038)	-0.053 (0.038)
$\rho_{S\theta} = 0.9$	0.000 (0.019)	-0.002 (0.021)	-0.012 (0.018)	-0.014 (0.018)
No failing ($\tau = 0$)	-0.001 (0.043)	0.002 (0.049)	0.000 (0.037)	-0.003 (0.017)

C.2.2 Case 2: $\delta > 0$

We consider the following parameter values: $\kappa = -1, \tau = 0.75, \delta = 0.25$

It is notable that none of the relationships under alternative assumptions reported from Simulation 1 are statistically significant, though some are estimated imprecisely. However, a general pattern appears such that relative to failing have no effect, a heterogeneous treatment effect will disproportionately push out higher quality teachers (i.e. $\hat{\beta}_2$ is more negative) for larger correlations between value-added and natural attrition, and the influence of this selection declines for higher

Table C.2: Mean (St.Dev.) of $\hat{\beta}_2$ from 1,000 Iterations of Simulation 2

	Correlation between VAM and Attrition ($\rho_{A\theta}$):			
Correlation between VAM and Scores ($\rho_{S\theta}$):	$\rho_{A\theta} = 0$	$\rho_{A\theta} = 0.1$	$\rho_{A\theta} = 0.5$	$\rho_{A\theta} = 0.9$
$\rho_{S\theta} = 0$	-0.106 (0.045)	-0.042 (0.047)	-0.139 (0.044)	-0.163 (0.042)
$\rho_{S\theta} = 0.1$	-0.098 (0.045)	-0.039 (0.050)	-0.131 (0.044)	-0.150 (0.044)
$\rho_{S\theta} = 0.5$	-0.070 (0.038)	-0.029 (0.043)	-0.083 (0.038)	-0.093 (0.037)
$\rho_{S\theta} = 0.9$	-0.017 (0.018)	-0.008 (0.020)	-0.018 (0.018)	-0.020 (0.018)
No failing ($\tau = 0$)	-0.001 (0.043)	0.002 (0.049)	0.000 (0.037)	-0.003 (0.017)

correlations between value-added and licensure score. For Simulation 2, which models an interaction effect of failing the test, we detect significant differences in the quality of teacher pushed out by failing the test exept for cases where the correlation between value-added and licensure scores is exceptionally high.

C.3 Summary

The analyses described in the main text of this paper suggest that failing the test differentially pushes out licensure test-takers with higher latent value-added. However, there are two potential mechanisms that could explain this effect. We would observe this result if candidates with higher latent value-added are more responsive to the failing treatment. However, we could also observe this pattern of results from a homogeneous effect of failing the licensure test that exacerbates an underlying correlation between latent value-added and the likelihood a test-taker will become a teacher that is independent of their licensure score. Our analysis is not able to distinguish between the relative impact of these possible mechanisms.