Department of Economics
Working Paper Number 2023-04

Teacher Licensing, Teacher Supply, and Student Achievement: Nationwide Implementation of edTPA

Abstract: The recent controversial roll-out of the educative Teacher Performance Assessment (edTPA) - a performance-based exam - raises the bar of initial teacher licensure and makes teacher recruitment difficult. We leverage the quasi-experimental setting of different adoption timing by states and analyze multiple data sources containing a national sample of prospective teachers and students of new teachers in the US. With extensive controls of concurrent policies, we find that the edTPA reduced prospective teachers in undergraduate programs, less-selective and minority-concentrated universities. Contrary to the policy intention, we do not find evidence that edTPA increased student test scores.

Authors:

Bobby W. Chung, University of South Florida

Jian Zou, UIUC

Available Online: October 2023

Teacher Licensing, Teacher Supply, and Student Achievement: Nationwide Implementation of edTPA*

Bobby W. Chung†

Jian Zou‡

September 29, 2023

Abstract

The recent controversial roll-out of the educative Teacher Performance Assessment (edTPA) - a performance-based exam - raises the bar of initial teacher licensure and makes teacher recruitment difficult. We leverage the quasi-experimental setting of different adoption timing by states and analyze multiple data sources containing a national sample of prospective teachers and students of new teachers in the US. With extensive controls of concurrent policies, we find that the edTPA reduced prospective teachers in undergraduate programs, less-selective and minority-concentrated universities. Contrary to the policy intention, we do not find evidence that edTPA increased student test scores.

Keywords: teacher licensing, edTPA, occupational licensing, teacher supply
JEL Classification: I28, J2, J44, K31, L51

*Data availability: This paper uses confidential data from the National Center for Education Statistics. The data can be obtained by filing a request directly with the US Department of Education (https://nces.ed.gov/pubsearch/licenses.asp). The authors are willing to assist. We are thankful for the comments from Dan Bernhardt, Peter Blair, Robert Bruno, Dan Goldhaber, Morris Kleiner, Brad Larsen, Benjamin Marx, Mindy Marks, Rigissa Megalokonomou, Charles Peck, Elizabeth Powers, Edward Timmons, and Russell Weinstein, as well as seminar participants at BE-Lab, UIUC, South Florida, West Virginia University, William & Mary, CSOR conference, SEA 2021, AEA 2022, RIDGE labor conference 2022, SOLE 2022, IHS Regulation Symposium, and the 2021 Carolina Region Empirical Economics Day. All errors are ours.

†Department of Economics, University of South Florida, FL, 33620; Research affiliate at HCEO and Knee Center. bobbywchung@usf.edu.

‡Department of Economics, University of Illinois at Urbana-Champaign, 1407 West Gregory Drive, Urbana, IL, 61801. jianzou2@illinois.edu.

1 Introduction

The earliest call for teacher entry requirements in the US dates back to the 1960s following a concerning trend in student test scores (Rudner and Adelman, 1987). The underlying belief is that a minimum standard for public school teachers can enhance student learning. After decades of development, teacher licensure became the primary guarantee of teacher quality in U.S. public schools. Public school teachers also become the largest licensed profession in the US (Gittleman et al., 2018). Although teacher licensing is universal in the U.S. public sector, the requirements have been determined by the state legislature and they have varied substantially across jurisdictions (Kleiner, 2010). The complex historical development and a lack of concurrent national data create challenges to evaluating the impacts of licensure exams on teachers and their students on a nationwide scale.

The net effect of license exams is unclear: license requirements increase entry costs that reduce teacher availability and may distort investments; but a minimum standard of teachers may improve student learning by eliminating incompetent teachers or training teacher skills. Since 2014, the educative Teacher Performance Assessment (edTPA) – a performance-based examination to evaluate the teaching readiness of prospective teachers – has gained popularity across the nation. By 2018, edTPA had become a mandatory testing component for both program completion and initial teacher licensure in eight states. The rollout of edTPA provides a contemporaneous quasi-experimental setting to evaluate the effectiveness of teacher license exams.[1]

EdTPA offers two novel features compared to the existing written licensure tests. In terms of the required effort, unlike the traditional one-time licensure tests, edTPA is a semester-long project involving lesson plans, classroom videos, and follow-up reports. The required money and time investment create an additional barrier to entry, potentially exacerbating the existing teacher shortages (Bergstrand Othman et al., 2017; Goldhaber

[1] In total, 18 states recognized edTPA as a test option for initial teacher licensure in 2018. We define treatment states as those with edTPA being the only option. See Section 2 about the policy timing.

et al., 2017; Petchauer et al., 2018; Gilbert and Kuo, 2019).[2] In terms of the format, edTPA stresses the performance-based structure that attempts to link the test component to the actual delivery of instruction. It is a revolutionary development in the education community since 2010s to establish a more practical way to evaluate teaching readiness (Sato, 2014). The uniqueness of edTPA then connects broadly to classical debates in economics about whether occupational licensing is welfare-improving (Friedman, 1962; Leland, 1979; Shapiro, 1986; Kleiner and Soltas, 2019). A higher requirement filters pre-service teachers at the lower tail of quality distribution, but may lead to negative sorting where higher ability candidates opt for better outside options (Goldhaber, 2007; Larsen et al., 2020).[3] The complementarity between the test content and quality of teaching is also a key for the new standard to benefit student learning.

This paper provides the first causal evidence about the effects of edTPA on teacher supply and student outcomes. We build on extant qualitative or case-specific analyses in education literature, providing a quantitative evaluation of edTPA using a national sample of new teachers and their students.[4] Controlling for an extensive set of concurrent policies, our identification strategy leverages different policy timing of edTPA, that compares the outcomes of interest in treatment states with other states before and after the implementation of edTPA. Our analysis not only applies to the ongoing debate about the implementation/revocation of edTPA, but also speaks to the current challenge of teacher shortage and the economics of occupational licensing in general. Although the implementation timing and the data structure limit our conclusion to shorter-run impacts, our discussion remains informative for policymakers to assess the immediate challenge of edTPA adoption in some states.

We first examine the number of graduates from teacher preparation programs – an

[2] A common alternative for pedagogy testing is Praxis PLT test, which costs around $150. By comparison, the administration cost of edTPA is $300, with additional $100-$300 if a retake is needed.

[3] Kugler and Sauer (2005) also documented that licensing induced negative selection in the physician profession.

[4] Related edTPA studies from education scholars include Greenblatt (2016), Goldhaber et al. (2017), Hébert (2019), and Gitomer et al. (2019).

important source of new teachers in the US public schools – documented in the Integrated Postsecondary Education Data (IPEDS).[5] Analyzing graduation years from 2011 to 2019, we find that edTPA reduced teacher graduates by a magnitude between 6.3% to 8.7%, depending on the empirical specifications. We also find that the negative effect primarily occurs in undergraduate programs, in less selective universities, and in minority-concentrated universities, suggesting issues associated with equity concerns and entry barriers created by edTPA. We are one of the first to document the employment/labor supply effect of teacher license exams (Kleiner and Petree, 1988; Larsen et al., 2020).[6]

We then assess the impact on students whose teachers likely went through edTPA. We analyze the restricted student data from 2009 to 2019 in the National Assessment of Educational Progress (NAEP) that contains the test scores of a national sample of students in the US. The NAEP is the largest nationally representative assessment in core subjects that provides a common yardstick to compare student progress in different states. Importantly for our analysis, the dataset links students to the years of experience of their corresponding subject teachers. This unique feature allows us to minimize measurement errors in identifying students of new teachers. We explore various specifications, sample criterion, and heterogeneity by school and student type. In all attempts, we do not find edTPA increased student test scores – either the direct impact by reform-relevant teachers (newly licensed), or the overall spillover effect on other teachers. Our confidence intervals can dismiss benchmark positive effects identified in other teacher-impact studies (Clotfelter et al., 2006; Goldhaber and Brewer, 2000; Kraft, 2020). The current finding about the edTPA effect on students supplements the discussion about the potential merit/defect of this on-going debated policy (Gitomer et al., 2019; Goldhaber et al., 2017). In general, the null impact on students echoes the study by Buddin and Zamarro (2009), who find that

[5] We find a similar result using the initial licensure data in the Title II. In our context, IPEDS has fewer measurement errors to identify the relevant teacher population. More discussion in the Data section.

[6] In the Schools and Staffing Survey (SASS) and the National Teacher and Principal Survey (NTPS), we also find that edTPA reduced the number of teachers who hold a regular license.

teacher licensure test performance is not related to student performance.[7]

We document the extent to which the license policy is related to teacher shortages. The supply of new teachers has been declining in the recent 10 years (King and James, 2022). The significance of evolving license requirements on new teacher supply complements commonly-discussed factors, including monetary incentives (Goldhaber et al., 2015; Feng and Sass, 2018), work environment (Carter and Carter, 2000; Carroll et al., 2000), support from teacher programs (Liu et al., 2004), and other education reforms (Guarino et al., 2006; Kraft et al., 2020). We find a significant decline in teacher candidates who studied the traditional-route programs, which is the major source of new teachers in the US public schools (National Center for Education Statistics, 2022).

Our results also offer important empirical updates about occupational licensing. License regulations have become a major labor institutions in the US that affects one-third of the workers (Kleiner, 2010). Researchers generally found that licensing reduces employment (Blair and Chung, 2019; Chung, 2022), increases price/wage (Kleiner, 2000; Kleiner and Krueger, 2013; Thornton and Timmons, 2013), and has minimal improvement on quality (Carpenter and Dick, 2012; Kleiner et al., 2016; Farronato et al., 2020).[8] Most empirical work on licensing uses cross-sectional variation or historical data. As a socially-influential workforce and the largest licensed profession in the US, economists have endeavored to quantify the effects of teacher license exams. Results are mixed, which reflects the differences in research design and policy context.[9] For example, Goldhaber and Brewer (2000) analyze a national sample of 12th-grade teachers with their individual certificate status and find that students perform better under teachers who hold a standard license (compared to alternative types of certification). Kleiner and Petree (1988) exploit cross-sectional variation of state license requirements in the 70s and find mixed effects of licensing on teachers' ability. Larsen et al. (2020) find that the license policies across the US during the 90s filtered

[7]As a supplementary analysis, in SASS/NTPS, we also observe that edTPA reduced subjective readiness of new teachers.

[8]The study by Anderson et al. (2020) is among the few to find a positive quality effect.

[9]State-specific studies include Clotfelter et al. (2007, 2010), Kane et al. (2008), and Sass (2015).

lower-quality teachers.[10] We offer complementary evidence that can be applied to the current teacher licensure reform, with a sharper identification, by looking at a controversial licensure initiative in recent years. The heterogeneity by the race of teacher candidates and program type also speaks to the distributional effect of licensing by population characteristics (Law and Marks, 2009; Blair and Chung, 2021, 2022; Xia, 2021).

Lastly, our results speak to the unintended consequences of high-stake teacher assessments. The goal of performance-based evaluations in public schools is to improve teacher performance by providing incentives. Unfortunately, studies have found that high-stakes on-the-job evaluations exerted pressure on teachers, hampering teacher recruitment and retention (Reback et al., 2014; Dee and Wyckoff, 2015; Sartain and Steinberg, 2016; Kraft et al., 2020; Cullen et al., 2021). We evaluate a new performance-based assessment for pre-service teachers and offer complementary findings that high-stake assessments dampen new teacher supply.

2 Background of edTPA

Licensure exams for prospective teachers in the US mostly cover three areas: basic skills (such as reading, writing, grammar, mathematics), subject matter, and pedagogical knowledge (Larsen et al., 2020). For pedagogical knowledge, the education community in the 1990s started to recognize the need for performance-based evaluation rather than written examinations to guarantee the teaching readiness of prospective teachers (Sato, 2014).

The earliest attempt to incorporate a performance evaluation process into the teacher licensure system was in 1998 in California.[11] Borrowing from the experience and models in California, the American Association of Colleges of Teacher Education (AACTE), which is the leading organization representing educator preparation programs in the US, cooperated

[10]Teacher quality is measured by the college background of the teachers. Angrist and Guryan (2008) also adopt a similar approach and find the state-mandated testing has no effects on teacher quality.

[11]The legislation is 'CA Senate Bill 2042'. Among a variety of models, popular options include the California Teaching Performance Assessment (CalTPA) and the Performance Assessment for California Teachers (PACT)

with the Stanford Center for Assessment to develop a standardised assessment called the educative Teacher Performance Assessment (edTPA) for nation-wide adoptions. EdTPA is now administered by Pearson Education.

Unlike the usual form of written examinations, edTPA requires candidates to show competency in preparing classes by submitting detailed lesson plans, delivering instruction effectively by recording the lesson during the internship, and properly assessing student performance to guide future instruction via a thorough analysis of student learning outcomes. The experts at Pearson then score a candidate's materials in three areas: 'Planning for Instruction and Assessment', 'Instructing and Engaging Students in Learning', and 'Assessing Student Learning'.[12] Preparation for edTPA takes place alongside the teaching internship. The entire process can take months.

Some education scholars contend that this performance-based format better reflects the complexity of teaching better than written examinations and prepare teachers to focus on student learning (Darling-Hammond and Hyler, 2013). However, ample qualitative evidence suggests that edTPA discourages new teachers from entering the teaching profession. Gilbert and Kuo (2019) find that the test fee together with miscellaneous expenses add a significant burden to students who have already struggled financially. Bergstrand Othman et al. (2017) find that time commitment and the uncertainty about passing the exam created mental stress to the teacher candidates. Besides, Greenblatt (2016) and Shin (2019) suggest that teacher candidates often found themselves focusing too much on catching up the scoring rubrics and deadline at the expense of teaching opportunities. Worse still, the negative impacts fall disproportionately on minority and lower-income candidates (Greenblatt and O'Hara, 2015; Goldhaber et al., 2017; Petchauer et al., 2018).[13]

By 2018, eight states had implemented edTPA to evaluate teaching effectiveness for

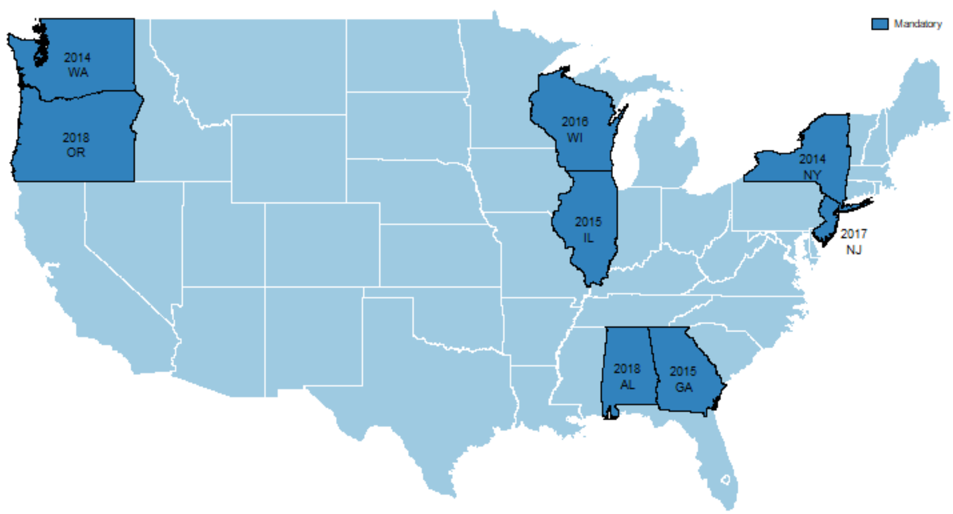
[12] Interested readers can refer to the official edTPA document (`http://www.edtpa.com/Content/Docs/edTPAMGC.pdf`) for a more detailed description on the assessment scheme.

[13] From the 2019 official statistics of edTPA (`https://edtpa.org/resource_item/2019AR`), the average pass rate is between 72% and 93%. The pass rate for ethnic minorities is significantly lower than for their white counterparts.

prospective public school teachers (see Figure 1).[14] Washington and New York were among the earliest states mandated edTPA as a necessary component for program completion and initial teacher licensure in January and May 2014, respectively. Prospective teachers have to satisfy a cutoff score to graduate from the teacher preparation program and qualify for a teacher license.[15] Later, the mandatory nature of edTPA expanded to Georgia (September 2015), Illinois (September 2015), Wisconsin (September 2016), New Jersey (September 2017), Alabama (September 2018) and Oregon (September 2018).[16]

Not all states consider edTPA as the sole assessment choice. By 2018, ten other states had added edTPA as an assessment option. Since teacher candidates in these states may opt for existing options other than edTPA, we do not include the optional states in our analysis.[17]

Figure 1: States mandated edTPA as an program completion and initial licensure requirement, Snapshot in 2018



Notes: In 2018, eight states have already introduced edTPA as the only assessment option for program completion and initial teacher licensure.

[14]Official document can be found here: `https://edtpa.org/resource_item/StatePolicyOverview`. We cross-check the mandatory nature in the official websites of state education departments.

[15]The cutoff scores vary by states and subjects. For a typical 15-Rubric criteria with a full score of 75, passing scores range from 35 to 42.

[16]New Jersey did not require a cutoff score until September 2019, and our results are robust to dropping NJ.

[17]The optional states include Arkansas, California, Delaware, Hawaii, Iowa, Maryland, Minnesota, North Carolina, South Carolina, West Virginia. Ohio, and Texas. We also do not observe the timing of edTPA in the optional states.

3 Data

3.1 IPEDS

We measure the teacher supply response to the implementation of edTPA by the number of graduates from teacher preparation programs in post-secondary institutions. The data is obtained from the Integrated Postsecondary Education Data (IPEDS), which contains rich information about the characteristics of post-secondary institutions in the entire US.[18] We exploit the detailed statistics of program completion by majors and identify graduates in teacher preparation programs (both undergraduate and postgraduate degrees) from school year 2010/2011 to 2018/2019.[19] The majors include 'Education, General', 'Bilingual, Multilingual, and Multicultural Education','Curriculum and Instruction', 'Special Education and Teaching', 'Teacher Education & Professional Development, Specific Levels and Methods', 'Teaching English or French as a Second or Foreign Language", and 'Education, Other'.[20] We then aggregate the number of teacher graduates at the institution level. In the sample (excluding optional states), we have a panel of 858 post-secondary institutions that offer teacher preparation programs (undergraduate or/and postgraduate).[21]

In Panel A of Table 1, in addition to the outcomes of interest — the number of teacher graduates and the breakdown by white and non-white candidates — we report time-varying

[18]An alternative data to measure the change in new teacher supply is the state-level initial licensure issuance documented in the Title II. It is less suitable than IPEDS in our context because Title II does not differentiate whether a license type requires edTPA. For example, in Washington, the aggregate count in Title II collapses 'Conditional certificate' and 'Residency certificate', where only the latter requires the edTPA score. The Washington districts can issue the conditional certificate for an individual who has not completed all the requirements for the regular certificate. In general, the state-level statistics in the Title II mix together temporary licenses (which do not require edTPA) with full licenses (which require edTPA). The measurement error in the outcome variable potentially attenuates the edTPA estimate. Nonetheless, in Table B1 of Appendix, we provide a supplementary result (excluding WA due to the aforementioned issue) using the Title II data and find a marginally significant effect.

[19]To become a licensed public school teacher in the US, a prospective teacher from the traditional route goes through training in a teacher preparation program. Alternatively, a person with a degree from non-education major can opt for the alternative route to complete an approved postgraduate program.

[20]IPEDS defines the major of a program using CIP codes. We follow the definition of teacher preparation programs recommended by Kraft et al. (2020) in their Appendix C.

[21]Including optional states, we have a total of 1,243 post-secondary institutions. The summary statistics are presented in Table A1 of appendix.

institution characteristics to account for concurrent changes in student demographics and the quality of institutions. The variables include the number and percent of minority of graduates in non-education majors, the submission rates and percentile scores of SAT/ACT, first-year full-time enrollment, par-time to full-time faculty ratio, and the amount of and the percent of students receiving federal grants/loans.

Table 1: Summary statistics (IPEDS) - Estimation sample

	Mean	SD	Min	Max
A. Outcomes:				
Education graduates	142.31	188.28	0.00	3041.00
Education graduates (white)	107.42	139.71	0.00	1763.00
Education graduates (non-white)	34.89	66.76	0.00	1968.00
B. Time-varying controls:				
Graduates (non-education majors)	1594.51	2160.49	1.00	16364.00
Minority graduates (% of non-education majors)	16.57	17.68	0.00	100.00
SAT submission rate	51.22	34.20	0.00	100.00
ACT submission rate	52.81	32.12	0.00	100.00
SAT 25 percentile score	476.10	65.18	215.00	740.00
SAT 75 percentile score	583.41	63.97	349.00	800.00
ACT 25 percentile score (cumulative)	20.38	3.31	3.00	33.00
ACT 75 percentile score (cumulative)	25.58	3.18	8.00	35.00
First-year FT enrollment	1056.10	1315.98	9.00	9082.00
Part-time/full-time faculty ratio	0.03	0.11	0.00	2.32
Grant (% student)	76.19	16.74	16.00	100.00
Grant (dollar amount, thousands)	44407.21	48785.28	326.33	397711.80
Loan (% student)	58.80	16.66	0.00	99.00
Loan (dollar amount, thousands)	21099.54	24175.53	0.00	256364.16

Sources: IPEDS 2011-2019.
Notes: This table shows summary statistics of estimation sample for teacher supply using IPEDS. Optional states are excluded. Summary statistics for all states are presented in Table A1 of appendix.

3.2 NAEP

To assess the effect of edTPA on student achievement, we analyze the biennial restricted data of the National Assessment of Educational Progress (NAEP) administered by the U.S Department of Education and the Institute of Education Sciences from 2009 to 2019. The assessment is a nationwide test in the US that measures the knowledge of a representative sample of students in various core subjects.[22] The standardized nature of the test enables us to compare student achievement across the country using a common measurement. We standardize the assessment scores by first averaging the composite values of five (or twenty) assessment items within each year-grade-subject and then standardize the averaged assessment scores over the estimation sample to have a zero mean and one standard deviation within the same year-grade-subject level.[23]

In addition to the standardized nature, an important feature of NAEP is that it links students to characteristics of the corresponding subject teacher. A key characteristic is the years of experience.[24] The question on years of experience contains continuous measures in survey year 2009 and 2011 and categorical responses in year 2013, 2015, 2017, and 2019. The categorical responses are listed as the following: Less than 1 year, 1-2 years, 3-5 years, 6-10 years, 11-20 years, 21 or more years, omitted, and multiple responses. Since edTPA only applies to new teachers during their preparation programs, one of the exercise we do is to focus on students whose teachers have less than two years of experience ("new teachers" thereafter) and estimate the direct effect.

NAEP also provides important characteristics of students and schools, which enable more precise estimations by including them as controls. They allow us to conduct balance tests by regressing these predetermined variables on edTPA policy variances in our later analyses. The student controls include student's race and gender, if the student needs an Individualized Education Program (IEP), and if the student is an English-language learner. The school controls include share of black students, indicators for charter school, urban area, eligibility of lunch programs, and whether school enrollment is larger than 500 students.[25]

[22] The subjects include reading, mathematics, science, writing, arts, civics, geography, economics, U.S. history, and technology & engineering literacy.

[23] In survey year 2009 and 2011, NAEP uses a five-item scale to measure the composite values of students' math and reading assessment in grades 4 and 8. In survey year 2013, 2015, 2017, and 2019, NAEP uses a twenty-item scale for math and reading assessment in grades 4 and 8.

[24] We also use the licensure background of the teachers to employ heterogeneity analysis in Table 10.

[25] While most control variables employed in this study share consistent measures across the two subjects and grades, one exception is the school enrollment. For students at grade 4, we use enrollment larger than 500 to indicate magnitude of schools. However, for students at grade 8, we use enrollment larger than 600 in year 2009, 2011, 2013, 2017, and 2019, as data in these years use a different category for student enrollment.

As far as the data provides, we assess the overall effects using all students and the direct effects using students with new teachers, excluding samples in the optional states.[26] Combining the NAEP from different cohorts yields a repeated cross-sectional sample of students. To address the concern that changes in student and school characteristics may affect teacher assignments and contaminate the causal estimates, we control for student and school characteristics presented in Table 2.

Table 2: Summary statistics (NAEP) - Estimation sample

	All teachers		**New teachers**	
	Math (1)	Reading (2)	Math (3)	Reading (4)
A. Outcomes:				
Assessment score (raw)	246.60 (34.01)	238.05 (40.13)	234.36 (28.62)	231.89 (40.33)
B. Student controls:				
White	0.49 (0.50)	0.50 (0.50)	0.42 (0.49)	0.44 (0.50)
Black	0.14 (0.34)	0.14 (0.34)	0.16 (0.37)	0.16 (0.37)
Hispanic	0.25 (0.43)	0.25 (0.43)	0.29 (0.46)	0.28 (0.45)
Female	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)
Individualized Education Program (IEP)	0.12 (0.33)	0.12 (0.33)	0.13 (0.33)	0.12 (0.33)
English learner	0.07 (0.26)	0.06 (0.24)	0.10 (0.30)	0.08 (0.27)
C. School controls:				
Charter school	0.04 (0.19)	0.04 (0.19)	0.07 (0.25)	0.07 (0.25)
Urban area	0.76 (0.42)	0.76 (0.43)	0.78 (0.41)	0.77 (0.42)
Share of black student	16.46 (26.55)	16.40 (26.40)	19.95 (28.92)	19.95 (28.88)
Lunch program	0.51 (0.50)	0.50 (0.50)	0.59 (0.49)	0.56 (0.50)
Student enrollment (≥ 500)	0.36 (0.48)	0.47 (0.50)	0.45 (0.50)	0.48 (0.50)
Number of Student	1,403,080	1,391,580	63,180	122,660

Sources: NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
Notes: This table shows summary statistics of student achievement, student characteristics, and school characteristics using NAEP. The mean is shown in the cell and the standard deviation is shown in the parenthesis. Columns (1) and (2) use all students, and columns (3) and (4) use students with new teachers (defined as having less than 3 years of teaching experience). The odd/even column presents samples of the Math/Reading assessment. Column (3) only includes Grade 4 Math due to the lack of teaching experiences in Grade 8 Math for some survey years. Raw assessment scores are reported in the summary statistics. The number of observations is rounded to the nearest 10 per IES disclosure guidelines.

[26]For student samples with a new teacher, we assess student performances in the mathematics score at grade 4, and the reading scores at grades 4 and 8. The restricted data also tracks the mathematics scores at grade 8. However, it does not contain teacher experience in 2017 survey year and cannot identify new teachers.

4 Identification Strategy

4.1 Teacher Supply

We estimate the effects of the mandatory edTPA requirement on teacher and student outcomes using a difference-in-differences framework with the leads and lags of treatment. Formally, for teacher supply analysis, we employ the following specification:

$$Y_{u,s,t} = \sum_{k \neq -1} \beta_k edTPA_{s,t} \mathbb{1}(t = t^* + k) + \mathbf{X}_{u,s,t}\Gamma + \mathbf{Z}_{s,t}\theta + \alpha_u + \alpha_t + \epsilon_{u,s,t} \tag{1}$$

where $Y_{u,s,t}$ refers to the log of the number of teacher graduates from institution u in state s in year t. To differentiate the edTPA effects on teacher supply by race, we run separate regressions on the number of white and non-white candidates. $edTPA$ is a dummy indicator equals 1 after a state mandated edTPA as the initial licensure requirement in the graduation year t^*. In the above non-parametric model, the omitted period is the graduation year right before the policy took effect. For example, the effective date in Illinois is September 2015. Its omitted year is the 2014/2015 school year. Then, $\beta_{(k>-1)}$ measures the edTPA effect on teacher supply in a given post-policy year, whereas $\beta_{(k<-1)}$ detects any deviation in trends in the pre-policy period between the edTPA and non-edTPA states. $X_{u,s,t}$ refers to a vector of time-varying controls at the institution level presented in Table 1. $\mathbf{Z}_{s,t}$ refers to a series of education policy indicators studied by Kraft et al. (2020) to control for potential confounds on the teacher supply response. The policies include the accountability reforms, the elimination of teacher tenure, the increase in probationary period, the elimination of mandatory union dues, the adoption of Common Core Standards, and changes in the licensure contents.[27] α_u and α_t are institution and year fixed effects, respectively. To account for serial correlation within a state, we cluster the standard errors at the state level.

[27] We code the policy year based on Table A1 of Kraft et al. (2020).

4.2 Student Outcomes

To estimate the impacts of edTPA on student achievement, we exploit the same policy variation in which edTPA becomes consequential in the educator licensing process shown in Figure 1 using the NAEP data. We employ the same differences-in-differences framework with a repeated cross-sectional sample of students. Formally, we estimate the following model:

$$Y_{i,j,s,t} = \sum_{k=-5,k\neq-1}^{k=2} \beta_k edTPA_{s,t} \mathbb{1}\left(t = t^* + k\right) + \mathbf{X}_{i,j,s,t}\Gamma + \mathbf{Z}_{s,t}\theta + \alpha_s + \alpha_t + \epsilon_{i,j,s,t} \tag{2}$$

where Y_{ist} is the reading/mathematics score of student i's in school j in state s sampled in period t. We again included the leads and lags of treatment indicators ($edTPA$) to check the parallel-trend assumption and also capture the dynamic effects. t_s^* is the policy implementation year for the eight treated states.

Continuing with the control variables, $\mathrm{X}_{i,j,s,t}$ is a vector of student and school characteristics listed in Table 2. $\mathbf{Z}_{s,t}$ refers to the same set of policy controls as in equation 1 that is studied in Kraft et al. (2020). α_s and α_t are the state fixed effects and year fixed effects, respectively. Standard errors are clustered at state level, which is the level the edTPA was implemented.

While this specification is almost identical to the one for teacher supply above, there is a difference on the time period because of the data structure. NAEP is a biennial assessment and the NAEP data we obtained is from 2009 to 2019, the time period of this specification ranges from -5 to 2 with each period t represents two academic years.

5 Results - Teacher Supply

5.1 Main Pattern

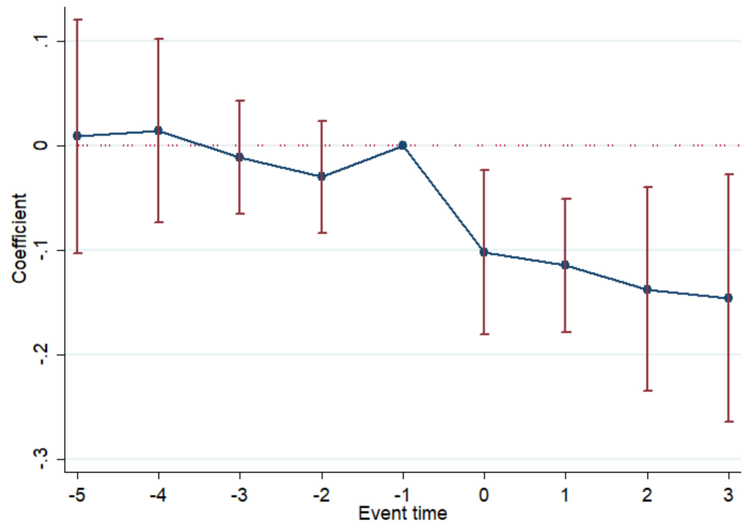
In Figure 2, we plot the event study dummies with the corresponding 95% confidence interval, conditional on institution and year fixed effects. The pre-treatment periods show that there is no systematic deviation in pre-trends. This validates the difference-in-differences model in producing a reliable post-edTPA counterfactual. The raw post-treatment pattern indicates a sudden decrease in teacher graduates that captures the immediate impact of candidates failing the test (and therefore fail to meet the graduation requirement). The effect size also grows over time, suggesting that the test both screens out students who failed the test and dissuades students from enrolling into teacher programs.[28] Since we have an unbalanced panel, the estimates of the two periods before and after the policy year are more precisely estimated. The growing effect (with less precision) beyond the second period is mainly driven by earlier-treated states.[29]

In Table 3, we present the estimates from the diff-in-diff strategy in various specifications. With the basic time-varying controls, fixed effects, and regional fixed effects, Column 1 shows that edTPA reduced the number of teacher graduates by 8.7%. In Column 2, we discern the edTPA from concurrent policies in the preK-12 public schools. We include the set of policy controls suggested by Kraft et al. (2020) that may influence new teacher supply. More importantly, the policy controls include changes in state licensure standards (basic skills, subject knowledge, and pedagogy). This help distinguish the effect of edTPA and other licensing changes that may affect the passing rate and degree choices. The negative effect drops by about a standard deviation, but the coefficient remains statistically significant at the 1% level.

[28]In the result not presented, we also find that edTPA reduced teacher-degree fall enrollment using IPEDS data from 2010 to 2018. Because the data is only biennial and the pre-trend is not stable, we do not over-interpret the enrollment finding.

[29]For the post-treatment year $t + 2$, the contributing states are the five early-treated units (WA, NY, IL, GA, WI); for the post-treatment year $t + 3$, the contributing states are WA, NY, IL, and GA.

Figure 2: No significant deviation of the pre-trend



Notes: This figure plots the estimates of the event study dummies and the corresponding 95% confidence interval. The regression in this figure includes year and institution fixed effects. No control variables are added. The endpoints are binned up to show a balanced window.

In Column 3, we further control for the implementation of public accountability reforms. Kraft et al. (2020) find that the on-the-job high-stake evaluations created pressure on teaching and impeded new teacher supply. This reform had an overlapping implementation schedule with the roll-out of the edTPA in the eight treatment states. Washington and Illinois implemented the reform one year after the roll-out of edTPA, while six of them implemented the reform prior to edTPA.

Despite the potential competing effects, we believe our estimate of edTPA does not pick up the influence of the accountability reform. First, the high-stake evaluation reform and the edTPA affect new teachers at different margins. Kraft et al. (2020) finds a reduction in initial issuance of licensure, while our focus is the number of new teacher graduates which captures the potential supply. New teachers are possible to pursue other occupations and to not obtain a teacher license upon graduation.[30] The two policies affect different margins because edTPA applies directly to the graduation requirement, while the high-stake evaluation reform

[30]While both margins are important, as we discussed in the data section, our focus is the effect on potential supply because the IPEDS provides a cleaner measurement than the licensure data.

can be seen as a cost at the occupation choice margin. Therefore, the accountability reform likely affected the occupational choice more than the degree completion rate. Consistent with the fact that the two polices affect new teacher at different margins, when we control for the accountability reform in Column 3, our estimate only changes slightly. The negative effect of the edTPA on degree completion rate remains strong at 1% level.

To further show that the edTPA estimate does not pick up the influence of the teacher accountability reforms, we explore two additional exercises. Our first check is to leverage the possible heterogeneity by program type. As shown by Kraft et al. (2020) (in their Section 8.1), the degree completion of postgraduate programs is more appropriate to capture the effects by high-stake evaluation reforms because the occupational choice of postgraduates is more responsive than those in an undergraduate program. By contrast, passing edTPA is tied to the graduation requirement of undergraduate programs but not all postgraduate programs. Also, some postgraduate degrees are designed for both not-yet licensed (requiring edTPA) and currently-licensed (not requiring edTPA). Therefore, on the degree completion margin, edTPA likely affects postgraduate programs less than undergraduate programs because it does not bind for some postgraduate students.[31]

We separate the analysis into undergraduate programs and postgraduate programs using the full specification in Column 3 of Table 3. Column 4 of Table 3 shows that edTPA significantly reduced the number of teacher graduates in the undergraduate programs and the magnitude is bigger than the average effect in Column 3. By contrast, we do not observe significant changes in the number of teacher graduates in post-graduate programs as shown in Column 5 of Table 3.

The null effect on postgraduates is a useful placebo test. When edTPA may not bind for some candidates in postgraduate programs, we observe insignificant impacts, both economically and statistically, on degree completion. Overall, the null impact on graduate

[31] An alternative explanation is differential ability of students in different programs. From the official Pearson report (https://edtpa.org/resource_item/2018BTN), we do not observe significant differences in edTPA scores between undergraduate and postgraduate programs.

degrees again shows our estimates of edTPA is less likely to be affected by the high-stake evaluations studied by Kraft et al. (2020), where postgraduate programs are more likely than undergraduate programs to be affected in their case.

Table 3: Diff-in-diff estimates with various specifications

	All programs			Undegraduate	Postgraduate
	(1)	(2)	(3)	(4)	(5)
edTPA	-0.0870***	-0.0651***	-0.0630***	-0.0830**	-0.0137
	(0.0285)	(0.0229)	(0.0221)	(0.0397)	(0.0279)
R-squared	0.394	0.398	0.398	0.445	0.504
Observations	7,168	7,168	7,168	6,453	5,605
Time-varying controls	X	X	X	X	X
Regional trend	X	X	X	X	X
Confounding policies#		X	X	X	X
Accountability Reform			X	X	X

Sources: IPEDS, 2011-2019.
Notes: Each column in each panel represents one regression. Dependent variable in each regression is the log of the number of teacher graduates. All regressions include time-varying controls in Panel A of Table 1, year and institution fixed effects. #Confounding policies are based on Table A1 of Kraft et al. (2020). All regressions are weighted by the pre-2014 average program size. Standard errors in the parenthesis are clustered at the state level. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

To further show that the edTPA estimate does not pick up the influence of the teacher accountability reforms, in Figure B1 of appendix, we perform a permutation test as a second check. The test runs 10,000 permutations with placebo treatments on the non-edTPA states in the conditional sample. In each round, we randomly assign placebo treatments to eight of the non-edTPA states that mimic the implementation timing of edTPA relative to the teacher accountability reform: two of them implemented edTPA 1 year prior; two implemented edTPA 1 year after; two implemented edTPA 2 years after; and the remaining two implemented edTPA 4 and 5 years after. If our edTPA estimate does pick up the effect of accountability reforms, the distribution of the placebo estimates should overlap with our DID estimates in Table 3. As shown in the first figure, our DID estimate (-0.063) in Column 3 does not overlap with the placebo distribution (p-value=0.087). In the second figure, we compares the effect on undergraduate programs in Column 4 of Table 3, which is the main driver of the average effect in Column 3 of Table 3, with the empirical placebo effects. The estimate is distinctively different than the placebo distribution (p-value=0.02). implying that the identified treatment effects less likely to pick up residual influences of the competing policy.

5.2 Robustness

We perform several tests to show that our identified effect on teacher supply does not capture other confounding factors. In Table 4, we find that the placebo treatment has essentially zero effects on the number of non-education graduates. This alleviates the concern that the drop in teacher graduates simply reflects state-specific shocks in tertiary education.

Table 4: Placebo test on non-education majors

	(1)	(2)	(3)
Placebo treatment	0.0218	0.0185	0.0184
	(0.0186)	(0.0129)	(0.0124)
R-squared	0.130	0.131	0.131
Observations	7,200	7,200	7,200
Regional trend	X	X	X
Confounding policies#		X	X
Accountability Reform			X

Sources: IPEDS, 2011-2019.
Notes: Dependent variable in each regression is the log of the number of non-education graduates (by race). All regressions include time-varying controls in Panel A of Table 1, year and institution fixed effects. All regressions are weighted by the pre-2014 average program size. Standard errors in the parenthesis are clustered at the state level. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

Because we are leveraging different policy timing across states, one concern is that the propensity to adopt edTPA is correlated with the regional teacher market conditions. Although previous licensing studies have pointed out that state variation in licensing policy is largely determined randomly by political forces, we perform a balancing test to show there are no systematic differences in observed characteristics between edTPA and non-edTPA states.[32] In all columns of Table 5, we regress an indicator equals 1 if a state adopted

[32]Relevant studies include Kleiner and Soltas (2019) and Larsen et al. (2020).

edTPA during the sample period on its pre-2014 attributes, including the level/growth of the number of teacher graduates, and average institution characteristics. Across columns, we use different measures of institution quality available in IPEDS to probe the sensitivity of the estimates. In all specifications, the only significant co-variate is the state number of education graduates that implies larger states tend to adopt edTPA. It is less of a concern because the state fixed effect takes into account time-invariant state differences and the growth in education graduates does not correlate significantly with the edTPA implementation. Overall, we do not find strong evidence that edTPA adoption was correlated with pre-policy characteristics of post-secondary institution or teacher graduates. This gives us credence about the quasi-random nature of edTPA implementations.

Next, in Table 6, we run a series of auxiliary fixed-effect models excluding the time-varying controls. As shown in Column 1 to 6, the edTPA treatment does not change institution characteristics, including first-year enrollment (all majors), faculty resources, and the financial background of students. In Column 7 of Table 6, we also find that there are no significant changes in teacher demand measured by public school enrollments.[33]

Lastly, the DID strategy with staggered timing may be susceptible to 'bad comparison' that the earlier-treated units are used as the comparison groups for later-treated units. The 'bad comparison' concern manifests as a problem when treatment effects evolve over time. Depending on the direction of the dynamic effects, the DID estimates with staggering timing may over- or under-state the average treatment effect. To check the robustness, we first assess if our estimation involves negative weights, which is the source of the bias (De Chaisemartin and d'Haultfoeuille, 2020). In the total of 1225 ATTs we have, only 1.5% of the weights are negative. The sum of negative weights equals -0.000904, which is negligible. Nonetheless, to check the sensitivity, we adopt the stacked regression estimator summarised by Baker et al. (2021).[34] We create event-specific data sets that pair the corresponding treatment states with only never-treated states. According to Figure 1, we have five treatment cohorts in

[33] We pool the state-level statistics (2011-2019) from the National Center for Education Statistics (NCES).

[34] A recent application is by Cengiz et al. (2019), who analyze the effect of minimum wage laws.

Table 5: Balancing test - Pre-2014 characteristics do not predict edTPA implementation

	(1)	(2)	(3)	(4)
Education graduates (level)	0.258*	0.258*	0.266*	0.261*
	(0.138)	(0.138)	(0.138)	(0.138)
Education graduates (growth)	0.0393	0.0479	0.0398	0.0406
	(0.161)	(0.161)	(0.164)	(0.161)
First-year FT enrollment (thousands)	-0.122	-0.120	-0.149	-0.141
	(0.261)	(0.260)	(0.255)	(0.255)
Part-time/full-time faculty ratio	-0.268	-0.278	-0.254	-0.261
	(0.370)	(0.372)	(0.374)	(0.371)
Grant (% student)	-0.00334	-0.00421	-0.00469	-0.00471
	(0.00957)	(0.00905)	(0.00903)	(0.00899)
Grant (dollar amount)	-6.43e-05	0.000448	0.00182	0.00139
	(0.00801)	(0.00697)	(0.00756)	(0.00658)
Loan (% student)	0.0113	0.0122	0.0114	0.0118
	(0.00864)	(0.00883)	(0.00866)	(0.00880)
Loan (dollar amount)	-0.00191	-0.00269	-0.00215	-0.00219
	(0.0141)	(0.0140)	(0.0147)	(0.0141)
SAT 25 percentile score	0.00155			
	(0.00355)			
SAT 75 percentile score		0.00145		
		(0.00298)		
ACT 25 percentile score			0.00790	
			(0.0694)	
ACT 75 percentile score				0.0192
				(0.0625)
Observations	51	51	51	51
R-squared	0.157	0.158	0.154	0.155

Sources: IPEDS, 2011-2014.
Notes: Dependent variable in all regressions is an indicator equals 1 if a state mandated edTPA after 2014. All regressors are pre-2014 averages.

total. Stacking all five into one single dataset, we run the full specification (including policy controls and regional trends) with set-specific institution- and year-fixed effects. This way we ensure the comparison group in each event cohort consists of 'clean control' states. With various specifications, Table 7 shows a similar reduction in new teachers with slightly bigger magnitudes compared to the main analysis in Table 3. The decrease again primarily occurs to four-year undergraduate programs.

Table 6: Changes in institution characteristics and teacher demand are not significant confounders

	(1) First-year enrollment	(2) PT/FT faculty ratio	(3) Grant (% students)	(4) Grant (amount)	(5) Loan (% students)	(6) Loan (amount)	(7) Teacher demand
Placebo treatment	0.00730	0.00219	0.670	1.475	-0.348	-0.0247	-0.00972
	(0.0237)	(0.00382)	(0.562)	(1.382)	(0.551)	(0.533)	(0.0116)
Observations	7,281	7,281	7,281	7,281	7,281	7,281	351
R-squared	0.025	0.002	0.023	0.219	0.118	0.093	0.172
Number of unit id	858	858	858	858	858	858	-
Number of state	-	-	-	-	-	-	39

Sources: IPEDS, 2011-2019 (Column 1 to 6); NCES (Column7).
Notes: All regressions include year and institution fixed effects. Standard errors in the parenthesis are clustered at the state level. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

Table 7: Robust to Addressing Staggered DID Issue - Stacked DID approach

	All programs			Undergraduate	Postgraduate
	(1)	(2)	(3)	(4)	(5)
edTPA	-0.0948**	-0.0950**	-0.0915**	-0.117**	-0.0384
	(0.0385)	(0.0370)	(0.0373)	(0.0469)	(0.0436)
R-squared	0.353	0.355	0.355	0.363	0.503
Observations	26,692	26,692	26,692	24,289	20,385
Regional trend	X	X	X	X	X
Confounding policies#		X	X	X	X
Accountability Reform			X	X	X

Sources: IPEDS, 2011-2019.
Notes: We adopted the stacked regression estimator summarised by Baker et al. (2021). We create event-specific data sets (a total of 5 in our case) that pair the corresponding treatment states with only never-treated states. In the final step, we stack the data sets and run the standard DID models with set-specific institution- and year-fixed effects. All regressions include time-varying controls in Panel A of Table 1, year and institution fixed effects. All regressions are weighted by the pre-2014 average program size. Standard errors in the parenthesis are clustered at the state level. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

5.3 Heterogeneity

The negative impacts on teacher supply fall disproportionately on candidates with a disadvantaged background, as the education literature suggests. To explore possible heterogeneity, we define the type of university in two aspects. First, we utilize the university-wide the 25^{th} SAT admission scores contained in IPEDS before 2014 to categorize institutions into two groups: more (top 50%) and less (less 50%) selective. In the regression, we interact the bottom 50% indicator with the treatment dummy to see if there exists heterogeneity by institution selectivity.[35] We present the results using the undergraduate sample in which the effect is concentrated.

In Panel A of Table 8, we present the heterogeneity result by ranking institutions based on their 25^{th} SAT percentile scores. In Column 1, while the total edTPA effect on teacher supply becomes less precise, there exists a marginally significant differential impact on less-selective programs. When we break down the analysis by the race of candidates, the differential effect by selectivity is mainly driven by Hispanics and candidates of 'other races', as shown in Column 4 and 5 respectively.

We then look at the heterogeneity by the racial composition of a university. In Panel B of Table 8, we categorize universities based on the percent of non-white graduates in non-education majors. The estimate of the interaction term in Column 1 is not precise, but the magnitude indicates a stronger negative impact on minority-concentrated universities. The combined effect (-14.6%) is also statistically significant at the 5% level (F-stat: 5.18). When we split the analysis by subgroup from Column 3 to 5, we also observe a stronger magnitude for non-white candidates than white candidates in Column 2.

Our heterogeneity results overall echo previous findings that minority teachers tend to have lower performances in teacher tests. For example, Angrist and Guryan (2008) also find that Hispanics have lower licensure scores than non-Hispanics candidates in state-mandated exams (basic skills or subject-matter) in the 90s. More recently, Goldhaber et al. (2017)

[35] The base indicators for 'bottom 50%' is time-invariant and is absorbed by institution-fixed effects.

also find that Hispanic candidates are more likely than non-Hispanics to fail edTPA in Washington.

Table 8: Heterogeneity by the type of university - Undergraduate programs

	(1) Total	(2) White	(3) Black	(4) Hispanic	(5) Other race
Panel A: X = University ranks at bottom 50% (SAT score 25^{th} percentile)					
edTPA	-0.0546	-0.0839*	-0.0266	0.0427	-0.0295
	(0.0445)	(0.0459)	(0.0631)	(0.0667)	(0.0561)
edTPA*X	-0.0654*	-0.0279	-0.0821	-0.180***	-0.167**
	(0.0328)	(0.0318)	(0.0905)	(0.0574)	(0.0786)
R-squared	0.446	0.465	0.092	0.074	0.168
Panel B: X = Minority students in non-education majors (above mean)					
edTPA	-0.0587*	-0.0662*	0.0213	0.0280	-0.0876*
	(0.0337)	(0.0360)	(0.0740)	(0.0510)	(0.0489)
edTPA*X	-0.0874	-0.0851**	-0.166**	-0.119*	-0.229**
	(0.0558)	(0.0390)	(0.0711)	(0.0629)	(0.104)
R-squared	0.445	0.494	0.088	0.067	0.179
Observations	6,453	6,453	6,453	6,453	6,453

Sources: IPEDS, 2011-2019.
Notes: Sample is restricted to undergraduate programs. Dependent variable in each regression is the log of the number of teacher graduates (by race). All regressions include time-varying controls in Panel A of Table 1, policy controls, year and institution fixed effects. We categorise institutions as bottom 50% according to their pre-2014 25^{th} percentile SAT in Panel A. In Panel B, we categorize institutions using the minority (non-white students) concentration in non-education majors. All regressions are weighted by the pre-2014 average program size. Standard errors in the parenthesis are clustered at the state level. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

6 Results - Student Achievement

The intention of edTPA is improving student learning through better equipping the teacher workforce. In this section, we test if edTPA significantly increases student test scores.

Since edTPA is a state-level policy on teachers, its treatment effects on students have two channels. When edTPA reduced new teacher supply as we document in the previous section, teacher composition may change because school districts need to fill vacancies, for example, by hiring teachers with a temporary/emergency license. Interacting with different types of teachers could affect student performance. In addition to this general effect, edTPA potentially changes the skill of new teachers and directly impacts their students accordingly.

6.1 Main Patterns

In Table 9, we first analyse the student-level achievement data in NAEP to document the overall effect using the full sample that contains all types of teachers. In Column 1, 3, and 5, conditional fixed effects, edTPA has a slight and insignificant negative effect on the math score, the reading score, and the total of the two. In Column 2, 4, and 6, when we run the full specification – with student, school, and policy controls – the coefficient turns positive but the effect size remains statistically insignificant. For the overall impact in Column 6, edTPA increased student test scores by 0.006 of a standard deviation on average.

Based on the findings on student standardized scores in other policy evaluation studies, we can compare our upper bound (adding 1.96 of S.E. to the estimate) and assess to what extent our confidence intervals is informative to dismiss small positive effects. Kraft (2020) surveyed 747 education policy research and recommended general guidelines on the effect size categories: 'small' if less than 0.05; 'large' if bigger than 0.2; and 'medium' if in between. According to our estimate in Column 6 of Table 9 and its confidence interval, we can rule out a positive effect of 0.054 of a standard deviation for overall test scores, that belongs to the lower end of the 'medium-effect' category.

Our estimates about the edTPA effect on student test scores are also smaller relative to the estimates in similar teacher value-added studies. For example, in Goldhaber and Brewer (1999) and Goldhaber and Brewer (2000), compared to teachers with no license, teachers with a regular license increase student math scores by 0.16 of a standard deviation. For specific licensure components, Goldhaber (2007) finds that teachers who pass Praxis II exam increased student math scores by 0.06 of a standard deviation. Besides certification backgrounds, experience of a teacher is a profound characteristic that impacts student learning. For example, Clotfelter et al. (2006) find that compared to first-year teachers, having experienced teachers increases student math scores by a range between 0.058 and 0.104 of a standard deviation, and increases student reading scores by a range between 0.046 and 0.092 of a standard deviation. Overall, the close-to-zero impact of edTPA on students we identify in Table 9 is an informative null compared to those in the landmark studies that investigate the impact of teacher characteristics.

Table 9: Impacts of edTPA on student achievement - All teachers

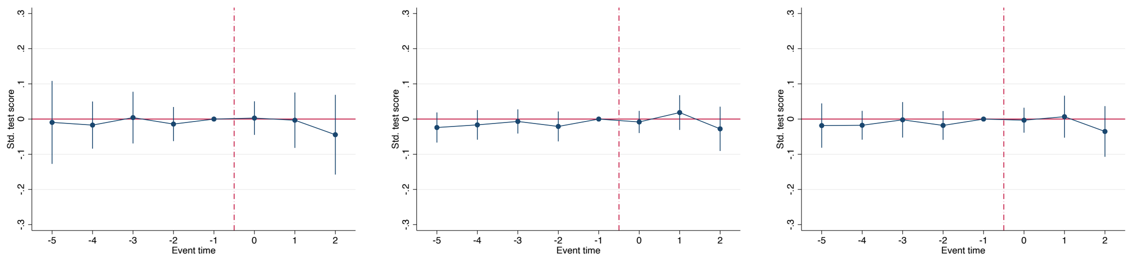
	Std. test score					
	Math		Reading		Total	
	(1)	(2)	(3)	(4)	(5)	(6)
edTPA	-0.033	0.003	-0.001	0.001	-0.017	0.006
	(0.036)	(0.029)	(0.020)	(0.022)	(-0.026)	(0.024)
R-squared	0.034	0.355	0.003	0.384	0.003	0.362
Observations	1,403,080	1,403,080	1,391,580	1,391,580	2,794,660	2,794,660
State-grade FE	X	X	X	X		
State-grade-subject FE					X	X
Year FE	X	X	X	X	X	X
Controls		X		X		X
Policy controls #		X		X		X

Sources: NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
Notes: The table shows estimates using the student level data in Grade 4 and Grade 8 in NAEP, containing eight states that adopted edTPA as the compulsory option and the control states that do not introduce edTPA. The test scores are standardized to a zero mean and one standard deviation. 'edTPA' refers to the treatment indicator. All regressions include state and year fixed effects. Student and school controls are listed in Table 2. #The policy controls are based on Table A1 of Kraft et al. (2020). Standard errors in brackets are clustered at the state level. Sample sizes is rounded to the nearest 10 per IES disclosure guidelines. $***p < 0.01, **p < 0.05, *p < 0.1$.

In Figure 3, we plot the event-study dummies to track the dynamics of the edTPA effect on math, reading, and the total of the two, respectively. In all the three samples, we observe a flat pre-treatment trend that validates the assumption of a DID design. For the treatment effects, the first two post-treatment periods are more precise since we have an unbalanced panel. The third period has a slight downward trend driven by WA and NY. We do not over-interpret the negative estimate, nor extrapolate the long-term impacts of edTPA on students. In general, the overall pattern in Figure 3 points to a robust conclusion that there is no evidence about a positive impact of edTPA on student learning during the sample period that we invesetigate.

After assessing the overall impacts on all students, we next turn to students whose teachers are newly-minted and have to go through edTPA. This inquiry measures the direct

Figure 3: Event study figures for student achievement



(a) Math **(b)** Reading **(c)** All

Sources: NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
Notes: The dependent variable is the standardized test score. Event period -1 is normalized to 0. The three subfigures (a), (b), and (c) use assessments in Math, Reading, and both subjects, respectively. The underlying regressions contain contains student, school and policy controls, conditional on year fixed effects and state-grade-subject fixed effects. The figures show the 95% confidence interval with robust standard errors clustered at the state level.

policy question about whether edTPA upgrades teacher effectiveness via the exam process.

In Panel A of Table 10, we limit the sample to new teachers (less than 2 years of experience) and present test scores results of their students. Running the full specification, we observe a slight negative effect of edTPA on student test scores in all subjects. Qualitatively, a negative impact on student learning is possible for new teachers given other studies showing that edTPA exam distracts internship learning of teacher candidates (Greenblatt, 2016; Shin, 2019). In Panel B, we further restrict the sample to students whose teachers are new and hold a license. The null effects of edTPA on test scores in all subjects remain stable.

While we cannot statistically conclude a definite negative effect, the upper bound of the confidence interval is still informative to our main hypothesis – whether edTPA has a positive impact on students. Taking into account the precision, we can rule out a positive effect of 0.025 of a standard deviation increase for total score according to Column 1 of Table 10 in Panel B (0.047 for math and 0.026 for reading). The best available and comparable evidence about edTPA and teaching effectiveness of new teachers so far is from Goldhaber et al. (2017), who find that students whose teachers passing edTPA (compared to whom failing it) score 0.061 of a standard deviation higher in math and 0.247 of a standard deviation higher in reading. Our close-to-zero estimates in Table 10 are very distinct from the existing benchmarks.

Table 10: Impacts of edTPA on student achievement - New teachers

	Std. test score		
	Total	Math	Reading
	(1)	(2)	(3)
Panel A. New teacher			
edTPA	-0.017	-0.004	-0.024
	(0.025)	(0.026)	(0.030)
R-squared	0.360	0.342	0.380
Observations	185,840	63,180	122,660
Panel B. New teacher + with license			
edTPA	-0.024	-0.008	-0.033
	(0.025)	(0.028)	(0.030)
R-squared	0.359	0.340	0.379
Observations	175,080	59,780	115,310
State-Grade-Subject FE	X		
State-Grade FE		X	X
Year FE	X	X	X
Student and school controls	X	X	X
Policy controls #	X	X	X

Sources: NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
Notes: The table shows estimates using student samples with new teachers (less than 2 years of experience). Panel A includes students with a new teacher and Panel B includes students with a new teacher who has a license. Columns (1) pool student samples from the three assessments and estimate the effects with the state-by-subject fixed effects. The dependent variables in following columns (2) and (3) are Math and Reading, respectively. Column (2) only include Grade 4 Math due to the lack of teaching experiences in Grade 8 Math for some survey years. The test scores are standardized to a zero mean and one standard deviation within each assessment sample. 'edTPA' refers to the treatment indicator. All regressions include state and year fixed effects. Student and school controls are listed in Table 2. #The policy controls are based on Table A1 of Kraft et al. (2020). Standard errors in brackets are clustered at the state level. Sample sizes are rounded to the nearest 10 per IES disclosure guidelines. $***p < 0.01, **p < 0.05, *p < 0.1$.

6.2 Balancing Test

Student assignment may change with the advent of edTPA, and a related concern is that the estimation sample changes systematically with the edTPA timing. For example, if new teachers are more/less likely to be assigned to disadvantaged students after edTPA, the change in student characteristics then confound the causal impact of edTPA.

We test if the edTPA treatment is correlated with student characteristics by performing a number of auxiliary models. We regress student characteristics on the edTPA indicator conditional on state-by-grade-subject and year fixed effects to assess the bias in the full sample and the new teacher sample, respectively. As shown in Panel A and Panel B of Table 11, edTPA implementation in general is not related to changes in most of the predetermined student characteristics. The most notable relationship is the percent of Hispanic students in the new teacher sample in Panel B. We believe the correlation does not cause systemic bias since the magnitude is both economically and statistically small, that 0.019 more in Hispanic representation only accounts for less than 5% of the sample standard deviation (as in Table 2). Overall, we do not find evidence that there is a systematic sample selection issue in our estimation.

Table 11: Balancing test - Correlation between edTPA and student characteristics

	White	Black	Hispanic	Female	IEP	Eng learner
Panel A. All teachers	(1)	(2)	(3)	(4)	(5)	(6)
edTPA	-0.008	-0.013	0.014	-0.004	0.001	0.006
	(0.011)	(0.008)	(0.011)	(0.002)	(0.005)	(0.008)
R-squared	0.095	0.122	0.075	0.000	0.006	0.035
Observations	2,794,660	2,794,660	2,794,660	2,794,660	2,794,660	2,794,660
Panel B. New teachers	(7)	(8)	(9)	(10)	(11)	(12)
edTPA	-0.025	0.001	0.019*	0.000	0.000	0.014
	(0.017)	(0.013)	(0.011)	(0.008)	(0.009)	(0.012)
R-squared	0.108	0.147	0.086	0.001	0.008	0.052
Observations	185,840	185,840	185,840	185,840	185,840	185,840

Sources: NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
Notes: Panel A includes all students and Panel B includes students with new teachers. The samples exclude optional states. The dependent variables are students' predetermined characteristics, while the independent variable 'edTPA' is an indicator where its value equals 1 if state s passes compulsory edTPA policy and 0 otherwise. All regressions include state-grade-subject and year fixed effects. Robust standard errors clustered at state level are in brackets. Sample sizes are rounded to the nearest 10 per IES disclosure guidelines. $***p < 0.01, **p < 0.05, *p < 0.1$.

6.3 Alternative Data

To verify the null impact, we explore the Stanford Education Data Archive (SEDA) as an additional data source. Although SEDA does not differentiate teacher types, its merit over NAEP is the annual data structure that expands our post-treatment period. The SEDA provides district-level average Math and Reading test scores for students from Grade 3 to 8 between 2009 and 2018, which were comparable to the NAEP sample (Fahle et al., 2021).[36] The SEDA contains 374,951 and 383,192 district-level observations for Math and Reading over the 3-8 grades during our sample period. To address district confounding factors, we further merge the SEDA with a set of district-level characteristics obtained from the American Community Survey – Education Tabulation (ACS-ED).[37] The district-level time varying controls include enrollment number (in log value), population percentage with a college degree or above, percentage of black population, and household median income (in log value). Consistently, the results in Table C1 show no evidence of edTPA improvement on student achievement, albeit with less precision. In Column (6), the upper bound of our estimate can dismiss a moderate effect size of 0.084 of a standard deviation for overall test scores.

6.4 Heterogeneity

The richness of NAEP allows us to look beyond the average effects on test scores. In this extension, we perform heterogeneity analysis.[38] In all the following analyses, we perform the full specification with student, school, and policy controls.

We utilize the rich detail about school information to explore if the effects of edTPA differ by schools. We run the full specification and interact the edTPA indicator with several

[36] We employ the SEDA version 4.1 that is available at `https://edopportunity.org/get-the-data/seda-archive-downloads/`.

[37] The ACS-ED provides school district characteristics via their five-year estimates, which is available at `https://nces.ed.gov/programs/edge/demographic/acs`. We use the median year of the five-year period as the survey year. For example, the ACS-ED 2007-11 is used for year of 2009.

[38] In our earlier version, we also find no significant heterogeneity by student ability using quantile regressions. Results are available upon request.

hard-to-staff characteristics, namely whether the school has a percentage of black/Hispanic students above the sample median, whether the school is in an urban area, and whether the school participates in the free lunch program.

Results are presented in Table 12. For the full sample in Panel A, we do not find heterogeneity by the four school characteristics. In Panel B, we also do not find significant heterogeneity using the new teacher sample.

Table 12: Heterogeneous Impacts of edTPA reforms: by school characteristics

	Std. test score			
X≡	Black	Hispanic	Urban	Free lunch
Panel A. All teachers	(1)	(2)	(3)	(4)
edTPA*X	0.002	0.000	0.053	-0.008
	(0.027)	(0.036)	(0.034)	(0.033)
edTPA	0.005	0.007	-0.016	0.010
	(0.019)	(0.038)	(0.019)	(0.031)
R-squared	0.362	0.362	0.362	0.362
Observations	2,794,660	2,794,660	2,794,660	2,794,660
Panel B. New teachers	(5)	(6)	(7)	(8)
edTPA*X	0.002	-0.020	0.050	0.005
	(0.040)	(0.029)	(0.036)	(0.034)
edTPA	-0.016	-0.013	-0.038	-0.018
	(0.032)	(0.031)	(0.033)	(0.036)
R-squared	0.358	0.358	0.358	0.358
Observations	185,840	185,840	185,840	185,840
State FE	X	X	X	X
Year FE	X	X	X	X
Student controls	X	X	X	X
School controls	X	X	X	X
Policy controls #	X	X	X	X

Sources: NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
Notes: Panel A includes all students and Panel B includes students with new teachers. The samples exclude optional states. The dependent variable is the standardized test score. 'edTPA' refers to the treatment indicator. All regressions include state and year fixed effects. Student and school controls are listed in Table 2. #The policy controls are based on Table A1 of Kraft et al. (2020). Standard errors in brackets are clustered at the state level. Sample sizes are rounded to the nearest 10 per IES disclosure guidelines. $***p < 0.01, **p < 0.05, *p < 0.1$.

6.5 Alternative Sample in NAEP

The focus on novice teachers in Table 10 may not provide a full picture about the edTPA effects on teaching skills because new teachers usually struggle when they first teach and need time to adapt. The potential benefits of edTPA on teaching skills, therefore, may take time to fully realize.

In this subsection, we investigate an alternative sample that contains students who have relatively seasoned teachers. To identify the relevant treated population in NAEP, we pair the teachers who are both relatively experienced ("3-5 years of experiences") and are required to pass edTPA for their initial licensure when they graduated three to five years ago. Given these sample requirements, we refine the treatment group to be the four earlier-treated states (WA, NY, IL, and GA). We then run a DID model focusing on students whose teachers have 3-5 years of experiences, with 2009, 2011, and 2013 NAEP as the pre-treatment periods and 2019 NAEP as the post-treatment period. We do not consider the relatively experienced ("3-5 years of experiences") in 2015 and 2017 since edTPA was not in effect yet in their the graduation year.[39] To make a clean comparison, we do not include the later treated states (AL, NJ, OR, and WI) in the estimation sample.

In Panel A of Table 13, we do not observe significant impacts of edTPA on students whose teachers are relatively experienced. Compared to the main effect in Table 9, the confidence levels in this sub-sample are less indicative that can only rule out positive effects of a moderate magnitude.

Since IL and GA adopted edTPA in 2015, teachers with five years of experience in the two states in 2019 NAEP are not required to pass the test. To reduce this measurement error, in Panel B of Table 13, we further drop IL and GA and only evaluate the treatment effect on relatively experienced teachers in WA and NY. While the coefficients turn negative, we do not find significant impacts.

[39]Technically, the relatively experienced in 2015 and 2017 shall be placed in the control group. However, they may not be a clean comparison if there exists indirect spillover discussed in Section 6.1.

Overall, while the estimates are less precise, we do not find evidence that edTPA benefits student learning even for students with relatively experienced teachers. Regarding the quality effect of licensing regulation, the insignificant impact is not uncommon (Bowblis and Smith, 2018; Carpenter and Dick, 2012; Chung, 2022; Farronato et al., 2020; Kleiner and Todd, 2009; Kleiner et al., 2016; Skarbek, 2008). In the teaching profession, factors outside the license requirement can confound the relationship between teacher licensing and student outcomes. For example, Buddin and Zamarro (2009) suggest that input-based (instead of performance-based) reward systems may fail to motivate teachers to fully realize their teaching potential. Another possibility is that the useful content of edTPA was offset by negative learning effects documented by the existing education studies (Bergstrand Othman et al., 2017; Greenblatt, 2016; Shin, 2019).

Table 13: Medium-term Impacts of edTPA on student achievement - Teachers with 3-5 years of teaching experiences

	Std. test score		
	Total (1)	Math (2)	Reading (3)
Panel A. Treated States: WA, NY, IL, GA			
edTPA	0.017	0.026	0.015
	(0.033)	(0.034)	(0.037)
R-squared	0.345	0.336	0.362
Observations	534,920	186,220	348,700
Panel A. Treated States: WA, NY			
edTPA	-0.028	-0.019	-0.034
	(0.060)	(0.040)	(0.072)
R-squared	0.342	0.330	0.362
Observations	492,270	171,140	321,130
State-Grade-Subject FE	X		
State-Grade FE		X	X
Year FE	X	X	X
Student and school controls	X	X	X
Policy controls #	X	X	X

Sources: NAEP 2009, 2011, 2013, and 2019.

Notes: The table shows estimates using student samples who have teachers with 3-5 years of teaching experiences. Survey years 2015 and 2017 are dropped because teachers with 3-5 years teaching experiences were partially treated by edTPA in these years. Panel A assesses students in four early treated states (WA, NY, IL, and GA) and Panel B includes students in the two states that first adopted edTPA (WA and NY). Columns (1) pool student samples from the three assessments and estimate the effects with the state-grade-by-subject fixed effects. The dependent variables in following columns (2) and (3) are Math and Reading, respectively. Column (2) only include Grade 4 Math due to the lack of teaching experiences in Grade 8 Math for some survey years. The test scores are standardized to a zero mean and one standard deviation within each assessment sample. 'edTPA' refers to the treatment indicator. All regressions include state-by-grade and year fixed effects. Student and school controls are listed in Panel B of Table 1. #The policy controls are based on Table A1 of Kraft et al. (2020). Standard errors in brackets are clustered at the state level. Sample sizes are rounded to the nearest 10 per IES disclosure guidelines. $***p < 0.01, **p < 0.05, *p < 0.1$.

7 Other Outcomes

In addition to student test scores, we explore alternative measurements about the potential impacts on schools and students as supplementary exercises. In this final section, we analyze the restricted-use teacher data in the Schools and Staffing Survey (SASS) in 2011-2012 and the National Teacher and Principal Survey (NTPS) in 2015-16 and 2017-2018. SASS is the former version of NTPS conducted by the U.S. Department of Education to survey U.S. elementary and secondary schools about staffing issues. Although there are only three waves of surveys, we construct an annually repeated cross-section of data of new public school teachers using their years of graduation (Larsen et al., 2020). In the 2011-2012 SASS, we pool teachers who graduated in 2009, 2010, or 2011; in the 2015-2016 NTPS, we pool teachers who graduated in 2012, 2013, 2014, or 2015; and in the 2017-2018 NTPS, we pool teachers who graduated after 2015. Combining the teacher sample in three waves, we resemble a yearly repeated cross-section panel for newly-minted public school teachers from 2009 to 2017.

Based on the resembled panel, we first investigate the license type of a teacher. As suggested by the immediate drop in new teacher graduates identified in the main analysis, the teacher composition in the licensure background of schools should change because fewer teachers are qualified for a standard license. Some school districts are allowed to fill vacancies with teachers with temporary licenses which typically do not require edTPA. Applying the generalized DID model to teacher-level data, we analyze an indicator that equals 1 if the new teacher holds a standard full license as the outcome. In Column 1 of Table C2 in the appendix, we find that edTPA reduced the likelihood of having a standard license by 13.8 percentage points. To the extent that job security is crucial for teacher retention, the decrease in teachers who hold a standard license is concerning to addressing teacher shortage.

We next investigate a set of unique subjective measures of teaching readiness. Teachers are asked to rate their first-year readiness, from 1 (not at all) to 4 (very well prepared), in six aspects class management: discipline, methods, subject matter, computer, assess students, differentiate instruction. Based on the six aspects, we construct a single measurement of 'subjective readiness' using factor analysis to avoid multiple hypotheses. In this exercise, we limit our sample to teachers who hold a regular license that requires passing edTPA. Table C3 of the appendix first provides summary statistics of the six raw measures and the factor variable.[40] In Column 2 of Table C2 in the appendix, we find that new teachers in the edTPA states have lower subjective readiness. Referencing the statistics in Table C3 of the appendix, the effect size is about 40% of a standard deviation of the factor variable. The lower subjective readiness coincides with the qualitative evidence that the edTPA exam distracted learning during internship (Greenblatt, 2016; Shin, 2019).

[40]Since the questions are asked only in 2011-2012 SASS and 2015-2016 NTPS, the resembled panel has a shorter sample period from 2009 to 2015 and the treatment effect consists of the first four states.

8 Conclusion

This paper makes the first attempt to provide causal evidence about the effect of edTPA on teacher supply and student performance, leveraging the quasi-random setting where states integrated edTPA into their licensure systems in different years.

For teacher supply, we analyze university-level graduation data from IPEDS which captures the major source of new teachers in the US. We find that edTPA reduced the number of teacher graduates and disproportionately hurt minority candidates in less-selective or minority-concentrated universities. Our results speak to the potential consequence on the existing shortage and diversity issue in the US public schools. The loss of minority teachers is also worrying given many researchers have found that teachers of the same race bring about a role-modeling effect for minority students (Dee, 2004; Gershenson et al., 2018).

Whereas licensure in general is a regulation on inputs, its quality influence on outputs is often obscure. While the student test score is not the only quality aspect, it concerns consumers (i.e. parents and education stakeholders) the most. Cross-checking different data sources, specifications, sample criterion, and heterogeneity, we do not see significant overall impacts on student test scores. We neither find evidence that edTPA improved student test scores of new teachers – the population supposed to be influenced the most. At the same time, we find suggestive evidence that edTPA reduced subjective readiness of new teachers.

A limitation we face is the lack of post-treatment years for the student outcomes. Besides that COVID severely confounds post-2019 program evaluations on students, some early-treated states had also revoked edTPA. Both factors put a ceiling on the long-term conclusion of edTPA. New teachers usually struggle in the first few years, and the skills they learn in the exams may take time to fully realize. While our discussion is constrained by the adoption timing and sample, this research serves as a starting point. The explanations behind the statistical patterns we document will be an important future agenda for both qualitative and quantitative research. An important note is that there may be positive changes in teaching methods that secondary data could not reflect. Our aim is not to provide an exhaustive list of explanations, but to document important cause-and-effect patterns. The short-term conclusion is nonetheless informative for policymakers to be aware of the immediate challenges.

A final note is that our results do not cast a veto against the entire teacher licensure system. Rather, we focus on a particular component of the licensure system that is frequently debated in the current education community. Our discussion is widely applicable to the educational policymakers nationwide, especially in the states which had integrated or are planning to integrate edTPA as a necessary component for initial teacher licensure. As of the time we prepare the manuscript, Georgia, Washington, and Wisconsin had removed the edTPA requirements, while Texas is trying a pilot program. The heterogeneity patterns we identify provide policymakers the areas to improve the assessment, if they add or retain the mandatory nature of edTPA. For example, a middle-ground solution is to provide more supports (financially and mentally) and guidelines to help prospective teachers get through the hurdle, which is found to have improved the experience of teacher candidates (Lachuk and Koellner, 2015; Muth et al., 2018).

References

Anderson, D Mark, Ryan Brown, Kerwin Kofi Charles, and Daniel I Rees, "Occupational licensing and maternal health: Evidence from early midwifery laws," *Journal of Political Economy*, 2020, *128* (11), 4337–4383.

Angrist, Joshua D and Jonathan Guryan, "Does teacher testing raise teacher quality? Evidence from state certification requirements," *Economics of Education Review*, 2008, *27* (5), 483–503.

Baker, Andrew, David F Larcker, and Charles CY Wang, "How much should we trust staggered difference-in-differences estimates?," *Available at SSRN 3794018*, 2021.

Blair, Peter Q and Bobby W Chung, "How much of barrier to entry is occupational licensing?," *British Journal of Industrial Relations*, 2019, *57* (4), 919–943.

_ **and** _ , "A model of occupational licensing and statistical discrimination," in "AEA Papers and Proceedings," Vol. 111 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2021, pp. 201–205.

_ **and** _ , "Job market signaling through occupational licensing," *Review of Economics and Statistics*, 2022, pp. 1–45.

Bowblis, John R and Austin C Smith, "Occupational licensing of social services and nursing home quality: A regression discontinuity approach," *Industrial and Labor Relations Review*, 2018.

Buddin, Richard and Gema Zamarro, "Teacher qualifications and student achievement in urban elementary schools," *Journal of Urban Economics*, 2009, *66* (2), 103–115.

Carpenter, II and M Dick, "Testing the utility of licensing: Evidence from a field experiment on occupational regulation," *Journal of Applied Business and Economics*, 2012, *13* (2), 28–41.

Carroll, Stephen, Robert Reichardt, Cassandra Guarino, and Andrea Mejia, "The distribution of teachers among California's school districts and schools," Technical Report, RAND CORP SANTA MONICA CA 2000.

Carter, MS and CM Carter, "How principals can attract teachers to the middle grades," *Schools in the Middle*, 2000, *9* (8), 22–25.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The effect of minimum wages on low-wage jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Chung, Bobby W, "The costs and potential benefits of occupational licensing: A case of real estate license reform," *Labour Economics*, 2022, *76*, 102172.

Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor, "Teacher-student matching and the assessment of teacher effectiveness," *Journal of human Resources*, 2006, *41* (4), 778–820.

_ , _ , and _ , "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects," *Economics of Education Review*, 2007, *26* (6), 673–682.

_ , _ , and _ , "Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects," *Journal of Human Resources*, 2010, *45* (3), 655–681.

Cullen, Julie Berry, Cory Koedel, and Eric Parsons, "The compositional effect of rigorous teacher evaluation on workforce quality," *Education Finance and Policy*, 2021, *16* (1), 7–41.

Darling-Hammond, Linda and Maria E Hyler, "The role of performance assessment in developing teaching as a profession.," *Rethinking schools*, 2013, *27* (4), 10–15.

Dee, Thomas S, "Teachers, race, and student achievement in a randomized experiment," *Review of economics and statistics*, 2004, *86* (1), 195–210.

_ and James Wyckoff, "Incentives, selection, and teacher performance: Evidence from IMPACT," *Journal of Policy Analysis and Management*, 2015, *34* (2), 267–297.

Fahle, Erin M, Belen Chavez, Demetra Kalogrides, Benjamin R Shear, Sean F Reardon, and Andrew D Ho, "Stanford Education Data Archive technical documentation version 4.1 June 2021," 2021.

Farronato, Chiara, Andrey Fradkin, Bradley Larsen, and Erik Brynjolfsson, "Consumer protection in an online world: An analysis of occupational licensing," Technical Report, National Bureau of Economic Research 2020.

Feng, Li and Tim R Sass, "The impact of incentives to recruit and retain teachers in "hard-to-staff" subjects," *Journal of Policy Analysis and Management*, 2018, *37* (1), 112–135.

Friedman, Milton, *Capitalism and freedom*, University of Chicago press, 1962.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge, "The long-run impacts of same-race teachers," Technical Report, National Bureau of Economic Research 2018.

Gilbert, Kristen A and Nai-Cheng Kuo, "Ethical, legal, and pedagogical issues in edTPA," *Critical Education*, 2019, *10* (8), 1–17.

Gitomer, Drew H, José Felipe Martínez, Dan Battey, and Nora E Hyland, "Assessing the assessment: Evidence of reliability and validity in the edTPA," *American Educational Research Journal*, 2019, p. 0002831219890608.

Gittleman, Maury, Mark A Klee, and Morris M Kleiner, "Analyzing the labor market outcomes of occupational licensing," *Industrial Relations: A Journal of Economy and Society*, 2018, *57* (1), 57–100.

Goldhaber, Dan, "Everyone's doing it, but what does teacher testing tell us about teacher effectiveness?," *Journal of human Resources*, 2007, *42* (4), 765–794.

Goldhaber, Dan D and Dominic J Brewer, "Teacher licensing and student achievement," *Better teachers, better schools*, 1999, pp. 83–102.

_ **and** _ , "Does teacher certification matter? High school teacher certification status and student achievement," *Educational evaluation and policy analysis*, 2000, *22* (2), 129–145.

Goldhaber, Dan, James Cowan, and Roddy Theobald, "Evaluating prospective teachers: Testing the predictive validity of the edTPA," *Journal of Teacher Education*, 2017, *68* (4), 377–393.

_ **, John Krieg, Roddy Theobald, and Nate Brown**, "Refueling the STEM and special education teacher pipelines," *Phi Delta Kappan*, 2015, *97* (4), 56–62.

Greenblatt, Deborah, "The consequences of edTPA," *Educational Leadership*, 2016, *73* (8), 51–54.

_ **and Kate E O'Hara**, "Buyer Beware: Lessons Learned from EdTPA Implementation in New York State.," *Teacher Education Quarterly*, 2015, *42* (2), 57–67.

Guarino, Cassandra M, Lucrecia Santibanez, and Glenn A Daley, "Teacher recruitment and retention: A review of the recent empirical literature," *Review of educational research*, 2006, *76* (2), 173–208.

Hébert, Cristyne, "Assessment in the clinical experience: student teaching and the edTPA," *Teaching Education*, 2019, *30* (4), 415–436.

Kane, Thomas J, Jonah E Rockoff, and Douglas O Staiger, "What does certification tell us about teacher effectiveness? Evidence from New York City," *Economics of Education review*, 2008, *27* (6), 615–631.

King, Jacqueline E. and Weade James, "Colleges of Education: A National Portrait, Second Edition," Technical Report, American Association of Colleges for Teacher Education 2022.

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

_ , "Enhancing quality or restricting competition: the case of licensing public school teachers," *U. St. Thomas JL & Pub. Pol'y*, 2010, *5*, 1.

_ **, Allison Marier, Kyoung Won Park, and Coady Wing**, "Relaxing occupational licensing requirements: Analyzing wages and prices for a medical service," *The Journal of Law and Economics*, 2016, *59* (2), 261–291.

_ **and Alan B Krueger**, “Analyzing the extent and influence of occupational licensing on the labor market,” *Journal of Labor Economics*, 2013, *31* (S1), S173–S202.

_ **and Daniel L. Petree**, “Unionism and Licensing of Public School Teachers: Impact on Wages and Educational Output,” in Richard B. Freeman and Casey Ichniowski, eds., *When Public Sector Workers Unionize*, University of Chicago Press, 1988.

_ **and Evan J Soltas**, “A welfare analysis of occupational licensing in US states,” Technical Report, National Bureau of Economic Research 2019.

_ **and Richard M Todd**, “Mortgage broker regulations that matter: Analyzing earnings, employment, and outcomes for consumers,” in “Studies of labor market intermediation,” University of Chicago Press, 2009, pp. 183–231.

Kraft, Matthew A, “Interpreting effect sizes of education interventions,” *Educational Researcher*, 2020, *49* (4), 241–253.

_ **, Eric J Brunner, Shaun M Dougherty, and David J Schwegman**, “Teacher accountability reforms and the supply and quality of new teachers,” *Journal of Public Economics*, 2020, *188*, 104212.

Kugler, Adriana D and Robert M Sauer, “Doctors without borders? Relicensing requirements and negative selection in the market for physicians,” *Journal of Labor Economics*, 2005, *23* (3), 437–465.

Lachuk, Amy Johnson and Karen Koellner, “Performance-based assessment for certification: Insights from edTPA implementation,” *Language Arts*, 2015, *93* (2), 84–95.

Larsen, Bradley, Ziao Ju, Adam Kapor, and Chuan Yu, “The Effect of Occupational Licensing Stringency on the Teacher Quality Distribution,” *NBER Working Paper*, 2020, (w28158).

Law, Marc T and Mindy S Marks, “Effects of occupational licensing laws on minorities: Evidence from the progressive era,” *The Journal of Law and Economics*, 2009, *52* (2), 351–366.

Leland, Hayne E, “Quacks, lemons, and licensing: A theory of minimum quality standards,” *Journal of political economy*, 1979, *87* (6), 1328–1346.

Liu, Edward, Susan Moore Johnson, and Heather G Peske, “New teachers and the Massachusetts signing bonus: The limits of inducements,” *Educational Evaluation and Policy Analysis*, 2004, *26* (3), 217–236.

Muth, Nicole, Kathleen Kremer, Val Keiper, Richard Schnake, and Renae MacCudden, “Implementation of edTPA Completion Prior to Student Teaching.,” *Mid-Western Educational Researcher*, 2018, *30* (3), 71–92.

National Center for Education Statistics, “Characteristics of Public School Teachers Who Completed Alternative Route to Certification Programs,” 2022. Retrieved 2022-09-20, from https://nces.ed.gov/programs/coe/indicator/tlc.

Othman, Lama Bergstrand, Rowand Robinson, and Nancy F Molfenter, "Emerging issues with consequential use of the edTPA: Overall and through a special education lens," *Teacher Education and Special Education*, 2017, *40* (4), 269–277.

Petchauer, Emery, Anica G Bowe, and Julene Wilson, "Winter is coming: Forecasting the impact of edTPA on Black teachers and teachers of color," *The Urban Review*, 2018, *50* (2), 323–343.

Reback, Randall, Jonah Rockoff, and Heather L Schwartz, "Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind," *American Economic Journal: Economic Policy*, 2014, *6* (3), 207–41.

Rudner, Lawrence M and Nancy Elizabeth Adelman, *What's happening in teacher testing: An analysis of state teacher testing practices*, Office of Educational Research and Improvement, US Department of Education, 1987.

Sartain, Lauren and Matthew P Steinberg, "Teachers' labor market responses to performance evaluation reform: Experimental evidence from Chicago public schools," *Journal of Human Resources*, 2016, *51* (3), 615–655.

Sass, Tim R, "Licensure and worker quality: A comparison of alternative routes to teaching," *The Journal of Law and Economics*, 2015, *58* (1), 1–35.

Sato, Mistilina, "What is the underlying conception of teaching of the edTPA?," *Journal of Teacher Education*, 2014, *65* (5), 421–434.

Shapiro, Carl, "Investment, moral hazard, and occupational licensing," *The Review of Economic Studies*, 1986, *53* (5), 843–862.

Shin, Minsun, "“The edTPA took away from my student teaching experience”: The impact of the edTPA on student teaching experiences," *Contemporary Issues in Early Childhood*, 2019, *20* (3), 309–312.

Skarbek, David, "Occupational licensing and asymmetric information: Post-hurricane evidence from Florida," *Cato J.*, 2008, *28*, 73.

Thornton, Robert J and Edward J Timmons, "Licensing one of the world's oldest professions: Massage," *The Journal of Law and Economics*, 2013, *56* (2), 371–388.

Xia, Xing, "Barrier to Entry or Signal of Quality? The Effects of Occupational Licensing on Minority Dental Assistants," *Labour Economics*, 2021, p. 102027.

Appendix to: "Teacher Licensing, Teacher Supply, and Student Achievement: Nationwide Implementation of edTPA"

Bobby W. Chung & Jian Zou

June, 2023

A Additional Summary Statistics

Table A1: Summary statistics (IPEDS) - Include optional states

	Mean	SD	Min	Max
A. Outcomes:				
Education graduates	138.25	184.57	0.00	3496.00
Education graduates (white)	100.85	135.32	0.00	1763.00
Education graduates (non-white)	37.40	71.67	0.00	1968.00
B. Time-varying controls:				
Graduates (non-education majors)	1623.25	2190.37	1.00	16364.00
Minority graduates (% of non-education majors)	18.17	18.72	0.00	100.00
SAT submission rate	51.70	33.25	0.00	100.00
ACT submission rate	54.24	30.44	0.00	100.00
SAT 25 percentile score	474.28	65.21	215.00	745.00
SAT 75 percentile score	581.62	64.95	349.00	800.00
ACT 25 percentile score	20.24	3.33	3.00	33.00
ACT 75 percentile score	25.44	3.27	8.00	35.00
First-year FT enrollment	1101.15	1370.80	6.00	10099.00
Part-time/full-time faculty ratio	0.03	0.11	0.00	2.60
Grant (% student)	76.63	16.46	16.00	100.00
Grant (dollar amount, thousands)	46209.60	51275.09	198.32	488027.59
Loan (% student)	58.71	16.50	0.00	100.00
Loan (dollar amount, thousands)	21562.41	24967.36	0.00	406393.00

Sources: IPEDS 2011-2019.
Notes: This table shows summary statistics of sample for teacher supply using IPEDS, including optional states.

B Additional Results on Teacher Supply

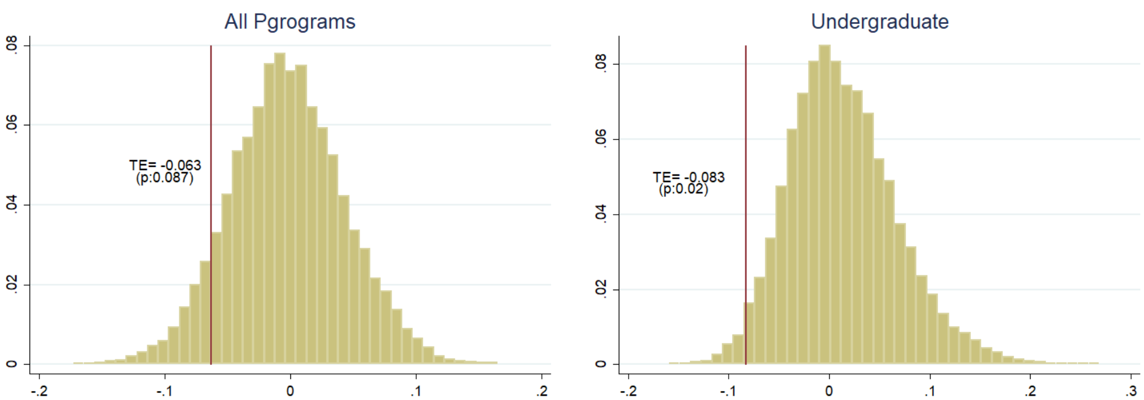
Table B1: Alternative data - State-level initial licensure in Title II

	(1)	(2)	(3)	(4)
edTPA	-0.295*	-0.125*	-0.122*	-0.0990*
	(0.148)	(0.0686)	(0.0685)	(0.0556)
Observations	449	449	449	449
R-squared	0.192	0.335	0.335	0.382
State control		X	X	X
Confounding policies#		X	X	X
Accountability Reform			X	X
Regional trends				X

Source: Title II, 2011-2019

Notes: Washington is dropped due to the measurement errors in the data. Dependent variable in all regressions is the log of the number of initial teacher licensure issued in a state. All regressions include year and state fixed effects, and state-level time-varying controls (unemployment rate, percent of college-educated population, black population). #Confounding policies are based on Table A1 of Kraft et al. (2020). All regressions are weighted by the 18-65 state population (per 10,000). Standard errors in the parenthesis are clustered at the state level. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

Figure B1: Permutation tests: Placebo treatments in non-edTPA states



Note: The permutation test in this figure constructs the distribution of placebo effects (10,000 rounds of permutation) using the non-edTPA states that implemented accountability reforms (Kraft et al., 2020). The first figure compares our treatment effect on all programs in Column 3 of Table 3 with the empirical placebo effects. The placebo treatments mimic the implementation timing of edTPA relative to the teacher accountability reform in the eight edTPA states: two of them implemented edTPA 1 year prior; two implemented edTPA 1 year after; two implemented edTPA 2 years after; and the remaining two implemented edTPA 4 and 5 years after. The second figure compares our treatment effect on undergraduate programs in Column 4 of Table 3 with the empirical placebo effects. Our estimates do not overlap with the placebo distribution, implying that the identified treatment effects less likely to pick up residual influences of the competing policy.

C Additional Results on Student Outcomes

Table C1: Robustness: Impacts of edTPA on student achievement using SEDA

	Std. test score					
	Math		Reading		Total	
	(1)	(2)	(3)	(4)	(5)	(6)
edTPA	-0.029	-0.004	0.011	0.033	-0.009	0.015
	(0.038)	(0.039)	(0.028)	(0.036)	(0.030)	(0.035)
R-squared	0.226	0.528	0.216	0.559	0.222	0.543
Observations	374,951	374,951	383,192	383,192	758,143	758,143
State-grade FE	X	X	X	X		
State-grade-subject FE					X	X
Year FE	X	X	X	X	X	X
Controls		X		X		X
Policy controls #		X		X		X

Sources: SEDA 2009-2018; ACS-ED 2007-11 to 2016-2020.
Notes: The table shows estimates using the district-level average of test scores from grades 3 to 8 in the SEDA, containing eight states that adopted edTPA as the compulsory option and the control states that do not introduce edTPA. The test scores are standardized to a zero mean and one standard deviation. 'edTPA' refers to the treatment indicator. All regressions include state and year fixed effects. Controls include time-varying district characteristics from NCES's ACS Education Tabulations (i.e., log value of enrollment number, population percentage with a college degree or above, percentage of black population, and log value of household median income). #The policy controls are based on Table A1 of Kraft et al. (2020). Standard errors in brackets are clustered at the state level.

Table C2: Other measures using NTPS/SASS

	Regular license	Subjective readiness
	(1)	(2)
edTPA	-0.138**	-0.328**
	(0.0536)	(0.125)
Observations	4,050	2,300
R-squared	0.113	0.091

Sources: SASS 2011-2012; NTPS 2015-16 and 2017-2018.
Notes: We resemble an annual repeated cross-section panel from 2009 to 2017 (Column 1) and from 2009 to 2015 (Column 2) using the graduation year of a teacher (Larsen et al., 2020). All regressions include the policy controls in the main analysis, a regional time trend, and state and year-fixed effects. In Column 2, the model also controls for teacher (female, black, Hispanic, other races, union, age, and the 2001 standardized average SAT of the graduating college pooled from College Scorecard) and school (elementary/secondary school, city, teacher-student ratio, percent of LEP, percent of IEP) characteristics. Sample weight applies in all regressions. Standard errors are clustered at the state level. Sample sizes are rounded to the nearest 10 per IES disclosure guidelines. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table C3: Summary Statistics of Teacher's Subjective Readiness

	Mean	SD	Min	Max
First year preparation - discipline	2.70	0.79	1.00	4.00
First year preparation - methods	2.98	0.75	1.00	4.00
First year preparation - subject matter	3.26	0.73	1.00	4.00
First year preparation - computers	2.96	0.85	1.00	4.00
First year preparation - Assess students	2.92	0.74	1.00	4.00
First year preparation - differentiate instruction	2.80	0.81	1.00	4.00
Factor variable	0.01	0.88	-2.96	1.66
Observations	2300			

Sources: SASS 2011-2012; NTPS 2015-16.

Notes: The table shows the summary statistics of teacher's subject readiness across six categories, as well as the factor variable. Sample sizes are rounded to the nearest 10 per IES disclosure guidelines.