HCEO WORKING PAPER SERIES

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


Bobby W. Chung
Jian Zou


Working Paper

**2021-039**


07/2021; Revised 09/2022



HUMAN CAPITAL AND ECONOMIC OPPORTUNITY GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

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

Bobby W. Chung†

Jian Zou‡

October, 2022

**Abstract**

States increasingly require prospective teachers to pass exams for program completion and initial licensure, including the recent controversial roll-out of the educative Teacher Performance Assessment (edTPA). 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 traditional route 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

*We are thankful for the comments from Dan Bernhardt, Peter Blair, Robert Bruno, Dan Goldhaber, Morris Kleiner, Brad Larsen, Benjamin Marx, Mindy Marks, Charles Peck, Elizabeth Powers, and Edward Timmons, as well as seminar participants at BE-Lab, University of Illinois (Urbana-Champaign), West Virginia University, 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.

†School of Business, St Bonaventure University, NY, 14778. bchung@sbu.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 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]

Unlike the traditional one-time written examinations, 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 et al., 2017; Petchauer et al., 2018; Gilbert and Kuo, 2019). It is also an open question as to whether the assessment

[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.

benefits students. 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).[2] The complementarity between the test content and quality of teaching is also a key for the new standard to benefit student learning. The overall impacts of edTPA then connect broadly to classical debates in economics about whether occupational licensing is welfare-improving (Friedman, 1962; Leland, 1979; Shapiro, 1986; Kleiner and Soltas, 2019).

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.[3] 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 efficacy of teacher licensure and occupational licensing in general.

We first examine the number of graduates from teacher preparation programs – an important source of new teachers in the US public schools – documented in the Integrated Postsecondary Education Data (IPEDS).[4] Analyzing graduation years from 2011 to 2019, we find that edTPA reduced teacher graduates by a magnitude between 11.2% to 15.8%, depending on the empirical specifications. We also find that the negative effect primarily occurs in four-year traditional-route 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

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

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

[4] We find a similar result using the initial licensure data in the Title II. In our context, IPEDS has fewer measurement errors to measure the treatment effect. More discussion in the Data section.

supply effect of teacher license exams (Kleiner and Petree, 1988; Larsen et al., 2020).

We then assess the effects of edTPA on student learning. 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 also links students to the years of experience of their corresponding subject teachers. This unique feature allows us to accurately measure the potential benefit of edTPA by focusing on 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.

Our results provide 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, 2020), 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).[5] 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.[6] 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 in 70s of state license requirements 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

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

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

lower-quality teachers. We offer complementary evidence that can be generalized 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 demographics (Law and Marks, 2009; Blair and Chung, 2018, 2020; Xia, 2021).

We also 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).

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.[7] 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 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'.[8] Preparation for edTPA takes place alongside the teaching internship. The entire whole process can take semesters.

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)

[7]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)

[8]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.

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).[9]

By 2018, eight states had implemented edTPA to evaluate teaching effectiveness for prospective public school teachers (see Figure 1).[10] 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.[11] 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).[12]

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.[13]

[9] From the latest official statistics of edTPA (https://secure.aacte.org/apps/rl/res_get.php?fid=3621&ref=rl), the average pass rate is between 75% and 92%. The pass rate for ethnic minorities is significantly lower than for their white counterparts.

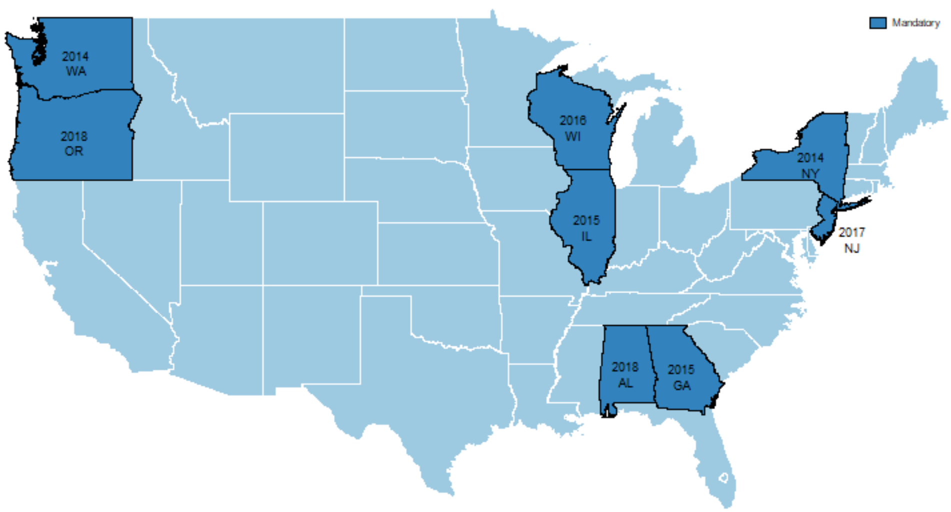
[10] Official document can be found here: https://edtpa.org/resource_item/StatePolicyOverview. We cross-check the mandatory nature in the official websites of state education departments.

[11] 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.

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

[13] The optional states include Arkansas, California, Delaware, Hawaii, Iowa, Maryland, Minnesota, North Carolina, South Carolina, West Virginia. Ohio, and Texas. We do not observe the timing of edTPA in the optional states. In Table B2 of Appendix, we check the sensitivity of this sample criteria by replicating the main table (Table 3). When we compare the eight treated states with the pooled control states (optional plus never treated), we find a slightly larger effect size which implies optional states had a slight increase in teacher supply over the sample period. This also highlights the reason to drop them in the main analysis because edTPA is not the only option in optional states and it may not be binding for program graduation.

**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.

# 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.[14] We exploit the detailed statistics of program completion by majors and identify graduates in teacher preparation programs (both bachelor's and master's degrees) from school year 2010/2011 to 2018/2019.[15] 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'.[16] 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 (either the traditional or alternative route).[17]

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

[14]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 the typical license (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.

[15]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.

[16]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.

[17]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[18]. 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.[19]

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.[20]

In addition to rich student and school characteristics, the NAEP data links students to characteristics of the corresponding subject teacher. This enables us to narrow down the sample to students whose teachers have less than two years since edTPA only applies to new teachers.[21] To ensure that the teachers have gone through the standard license procedure, we drop students whose subject teachers do not have a teacher license. We also restrict our attention to students with traditional route teachers because the parallel trend assumption does not hold for the alternative route sample. We defer the discussion to the result section.

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

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

[20]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.

[21]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.

As far as the data provides, we assess student performances in the mathematics score at grade 4, and the reading scores at grades 4 and 8.[22] 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.[23]

**Table 2:** Summary statistics (NAEP) - Estimation sample

| | Grade 4 Math | Grade 4 Reading | Grade 8 Reading |
|---|---|---|---|
| **A. Outcomes:** | | | |
| Assessment score (raw) | 235.76 (28.20) | 216.75 (34.49) | 252.97 (37.37) |
| **B. Student controls:** | | | |
| White | 0.46 (0.50) | 0.46 (0.50) | 0.50 (0.50) |
| Black | 0.14 (0.35) | 0.13 (0.34) | 0.13 (0.34) |
| Hispanic | 0.28 (0.45) | 0.29 (0.45) | 0.24 (0.43) |
| Female | 0.49 (0.50) | 0.50 (0.50) | 0.49 (0.50) |
| Individualized Education Program (IEP) | 0.13 (0.33) | 0.12 (0.33) | 0.11 (0.32) |
| English learner | 0.10 (0.29) | 0.09 (0.29) | 0.06 (0.23) |
| **C. School controls:** | | | |
| Charter school | 0.05 (0.22) | 0.05 (0.22) | 0.04 (0.20) |
| Urban area | 0.77 (0.42) | 0.77 (0.42) | 0.74 (0.44) |
| Share of black student | 16.68 (25.64) | 16.50 (25.32) | 15.73 (25.35) |
| Lunch program | 0.56 (0.50) | 0.56 (0.50) | 0.50 (0.50) |
| Student enrollment ($\geq$ 500) | 0.45 (0.50) | 0.45 (0.50) | 0.50 (0.50) |
| **Number of Student** | 51,460 | 53,530 | 41,260 |

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* This table shows summary statistics of estimation sample (students with new traditional route teachers) for student achievement using NAEP. The mean is shown in the cell and the standard deviation is shown in the parentheses. Each column presents one of the three student assessment samples: Math at Grade 4 , Reading at Grade 4, and Reading at Grade 8. Raw assessment scores are reported in the summary statistics. The number of observations is rounded to the nearest 10 per IES disclosure guidelines. Summary statistics for all states are presented in Table A2 of appendix.

[22] 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.

[23] Table A2 of appendix presents the summary statistics by including also the optional states. Table A3 contrasts the summary statistics between the traditional and alternative route sample.

# 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.[24] $\alpha_u$ and $\alpha_t$ are institution and year fixed effects, respectively. To account for serial correlation within a state, we cluster the standard errors at the state level.

[24]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:

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

where $\mathrm{Y}_{ist}$ 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). $\alpha_s$ and $\alpha_t$ 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.

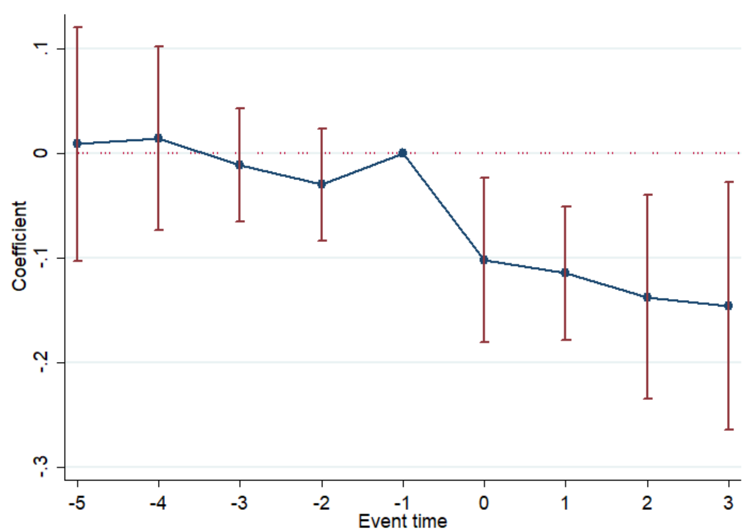
While this specification is almost identical to the one for teacher supply above, there is a difference on the time period because of the data structure. NAEP is a biennial assessment and the NAEP data we obtained is from 2009 to 2019, the time period of this specification ranges from -5 to 2 with each 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 effects show that there is no systematic deviation in pre-trends. This further validates the difference-in-differences model in producing a reliable post-edTPA counterfactual. In Table 3, we present the estimates from the diff-in-diff strategy in various specifications. With the basic time-varying controls and fixed effects, Column 1 shows that edTPA reduced the number of teacher graduates by 15.8%.

**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.

From Column 2 to 5, we discern the edTPA from concurrent policies in the preK-12 public schools. In Column 2, we include the set of policy controls suggested by Kraft et al. (2020) that may influence new teacher supply. The negative effect drops by a standard deviation, but the coefficient remains statistically significant at the 1% level. In Column 3, we also 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 for edTPA does not pick up the influence of the accountability reform. As shown by Kraft et al. (2020) (in their Section 8.1), the high-stake evaluation only affected the one-year alternative-route programs. The reason is that the decision to enter the teacher profession is more responsive than those in a four-year traditional route program. Whereas they show the reform caused a marginally significant reduction in program completion in the one-year programs, we will show that our effects only occurs to four-year traditional-route programs (discussed later in Table 4). Consistent with the fact that the two polices affect different groups of teacher graduates, when we control for the accountability reform in Column 3, our estimate only changes slightly because the traditional-route candidates constitutes the majority of the teacher preparation programs. The negative effect of the edTPA on new teacher supply remains strong when we also control for regional-specific time trends in Column 4.

To further show that the edTPA estimate does not pick up the influence of the teacher accountability reforms, we perform two additional exercises. We first estimate the DID regression using a 'conditional sample'. In the conditional sample, we only compare the edTPA states with control states that have implemented the accountability reforms during our sample period.[25] The idea is to condition on the same policy environment, namely the presence of accountability reforms prior to edTPA, to make the treated states more alike the control states. To achieve the comparison in this additional test, we exclude Washington and Illinois in the treatment because their accountability reforms came after, not prior to, the edTPA. Using the conditional sample in Column 5, the negative effect on new teacher supply remains significant. The larger estimate in Column 5 is due to the exclusion of Washington.

[25]The excluded control states are California, Iowa, Montana, Nebraska, Vermont, and Wyoming. According to Kraft et al. (2020), these states do not have the accountability reforms.

The Education Board in Washington engaged actively in pilot programs since 2009, which helped the preparation of the teacher candidates. It also set a low cut-off score of 35 that boosted the passing rate (Meuwissen and Choppin, 2015; Meuwissen et al., 2016).[26]

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

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| edTPA | -0.158*** | -0.116*** | -0.111*** | -0.112*** | -0.140*** |
| | (0.0446) | (0.0336) | (0.0332) | (0.0362) | (0.0318) |
| R-squared | 0.244 | 0.248 | 0.248 | 0.252 | 0.241 |
| Observations | 7,281 | 7,281 | 7,281 | 7,281 | 6,429 |
| Confounding policies# | | X | X | X | X |
| Reform (Kraft et al., 2020) | | | X | X | X |
| Regional trend | | | | X | X |
| Conditional Sample## | | | | | 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 (by race). 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 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.
## Conditional sample drops control states that did not implement the accountability reforms studied by Kraft et al. (2020). They include California, Iowa, Montana, Nebraska, Vermont, and Wyoming. We also drop Illinois and Washington as the treatment states since they implement the reforms after the edTPA.

In Figure B1 of appendix, we perform a permutation test as the 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 left figure, our DID estimate (-0.112) is distinctively different than the placebo distribution (p-value=0.03). On the right figure, we

[26] When we separate WA from the rest of edTPA states, we find that WA has both economically and statistically insignificant impacts on new teacher supply, in both the IPEDS and the Title II data.

perform a similar exercise with six instead of eight placebo treated states. This time, we only mimic the timing of the six treated states in the 'conditional sample' that had implemented accountability reforms prior to edTPA. The purpose is to compare our DID estimate in Column 5 of Table 3 (-0.14) with the placebo treatments. Again, it is significantly different from the placebo distribution (p-value=0.03).

In Table 4, we differentiate the effects by program types and the race of students. We run the analyses using the full specification with all institution controls, regional time trend, and policy controls using the full sample. Column 1 to 3 show that edTPA reduced the number of teacher graduates in the traditional route and the magnitudes are similar to the results using the full sample in Table 3. By contrast, we do not observe significant changes in the number of teacher graduates in post-graduate programs — one source of the alternative route to certification - as shown from Column 4 to 6. The bigger effect on the traditional route is consistent with the observation that prospective teachers pursuing the alternative route are generally more committed (Sass, 2015). The null effect for the alternative route is also a useful placebo test since some of the master students might have already had a license. The null impact on graduate 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 alternative-route candidates are more likely than traditional-route candidates to be affected in their case.

**Table 4:** Affecting only the traditional-route programs

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Bachelor's degree | | | Post-graduate degree | | |
| | All | White | Non-white | All | White | Non-white |
| edTPA | -0.149*** | -0.150*** | -0.123*** | -0.00793 | -0.00471 | 0.00747 |
| | (0.0451) | (0.0392) | (0.0450) | (0.0335) | (0.0364) | (0.0251) |
| Observations | 6,680 | 6,680 | 6,680 | 5,756 | 5,756 | 5,756 |
| R-squared | 0.201 | 0.226 | 0.065 | 0.356 | 0.313 | 0.181 |

*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 (by race and program type). All regressions include time-varying controls in Panel A of Table 1, regional time trend, policy controls, year and institution fixed effects. 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.

## 5.2 Robustness

We perform several tests to show that our identified effect on teacher supply does not capture other confounding factors. In Table 5, we find that the placebo treatment has essentially zero effects on the number of non-education graduates. This alleviates the concern that the drop of teacher graduates simply reflects state-specific shocks in tertiary education.

**Table 5:** Placebo test on non-education majors

| | (1)<br>Total | (2)<br>White | (3)<br>Non-white |
|---|---|---|---|
| Placebo treatment | -0.0119 | -0.0322 | -0.0189 |
| | (0.0317) | (0.0382) | (0.0248) |
| Observations | 7,281 | 7,281 | 7,281 |
| R-squared | 0.063 | 0.033 | 0.169 |

*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, policy controls, year and institution fixed effects. 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.

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.[27] In all columns of Table 6, we regress an indicator equals 1 if a state adopted 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

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

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.

**Table 6:** 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.

Next, in Table 7, 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 resource, and the financial background of students. In Column 7 of Table 7, we also find that there is no significant changes in teacher demand measured by public school enrollments.[28]

**Table 7:** 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.

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 ATTs have a negative weight. 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).[29] 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 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. Table 8 shows similar reductions in new teachers with slightly bigger

[28] We pool the state level statistics (2011-2019) from the National Center for Education Statistics (NCES).
[29] A recent application is by Cengiz et al. (2019), who analyze the effect of minimum wage laws.

magnitudes compared to the main analysis in Table 3.

**Table 8:** Robust to Addressing Staggered DID Issue - Stacked DID approach

| | (1)<br>Total | (2)<br>White | (3)<br>Non-white |
|---|---|---|---|
| edTPA | -0.162*** | -0.168*** | -0.114*** |
| | (0.0318) | (0.0353) | (0.0316) |
| Observations | 25,705 | 25,705 | 25,705 |
| R-squared | 0.179 | 0.194 | 0.084 |

*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, policy controls, a regional trend, 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.

## 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 from two aspects. First, we utilize the university-wide the $25^{th}$ SAT and ACT 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 top 50% indicator with the treatment dummy to see if there exists heterogeneity by institution selectivity.[30]

In the main table below, we present the results using the traditional route sample and attach the supplement result using both traditional and alternative route in the appendix. In Panel A and B of Table 9, we present the heterogeneity result by ranking institutions based on their $25^{th}$ SAT and ACT percentile scores. In Column 1, both the results using the SAT and ACT rank show that more selective universities have smaller negative impacts as suggested by the interaction terms. When we differentiate the effects by race, the difference in the magnitude between less and more selective universities is more apparent for non-white prospective teachers as in Column 3. From Column 4 to 6, we see a similar pattern when we further break down non-white students into Black, Hispanic, and Other races. In appendix, using the full sample in Table B3, the differential impacts are less significant largely diluted by the null impacts we find for alternative route programs in Table 4.

We then look at the heterogeneity by the racial composition of a university. In Panel C, we categorise universities based on the percent of non-white graduates in non-education majors. When we concentrate on non-white candidates in Column 3, the negative effect is significantly bigger for minority-concentrated institutions. We also observe the significant differential effects for black in Column 4 and Hispanics in Column 5 in minority-concentrated universities.

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

**Table 9:** Heterogeneity by the type of university - Traditional route

| | (1) Total | (2) White | (3) Non-white | (4) Black | (5) Hispanic | (6) Other race |
|---|---|---|---|---|---|---|
| *Panel A: X = University ranks at top 50% (SAT score $25^{th}$percentile)* | | | | | | |
| edTPA | -0.162** | -0.161** | -0.172*** | -0.133* | -0.0971** | -0.149*** |
| | (0.0653) | (0.0611) | (0.0634) | (0.0660) | (0.0359) | (0.0524) |
| edTPA*X | 0.0460*** | 0.0200 | 0.130*** | 0.0849* | 0.0676*** | 0.0997** |
| | (0.0112) | (0.0266) | (0.0338) | (0.0479) | (0.0183) | (0.0425) |
| R-squared | 0.200 | 0.225 | 0.059 | 0.051 | 0.040 | 0.083 |
| *Panel B: X = University ranks at top 50% (ACT score $25^{th}$percentile)* | | | | | | |
| edTPA | -0.184*** | -0.183*** | -0.188*** | -0.130* | -0.117*** | -0.163*** |
| | (0.0590) | (0.0582) | (0.0623) | (0.0658) | (0.0407) | (0.0486) |
| edTPA*X | 0.0935*** | 0.0670** | 0.169*** | 0.0825* | 0.112** | 0.135** |
| | (0.0308) | (0.0323) | (0.0623) | (0.0485) | (0.0424) | (0.0537) |
| R-squared | 0.201 | 0.225 | 0.060 | 0.051 | 0.041 | 0.084 |
| *Panel C: X = Minority students exceed 50% in non-education majors* | | | | | | |
| edTPA | -0.108* | -0.121** | -0.0798 | -0.0638 | -0.0166 | -0.122** |
| | (0.0622) | (0.0540) | (0.0599) | (0.0504) | (0.0432) | (0.0544) |
| edTPA*X | -0.164*** | -0.141* | -0.176*** | -0.104*** | -0.199** | 0.0109 |
| | (0.0498) | (0.0734) | (0.0420) | (0.0279) | (0.0803) | (0.0668) |
| R-squared | 0.198 | 0.239 | 0.095 | 0.044 | 0.031 | 0.103 |
| Observations | 6,575 | 6,575 | 6,575 | 6,575 | 6,575 | 6,575 |

*Sources:* IPEDS, 2011-2019.

*Notes:* Sample is restricted to traditional route 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 top 50% according to their pre-2014 characteristics, namely the $25^{th}$ percentile SAT in Panel A, ACT scores in Panel B, and minority (non-white students) concentration in non-education majors in Panel C. 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.

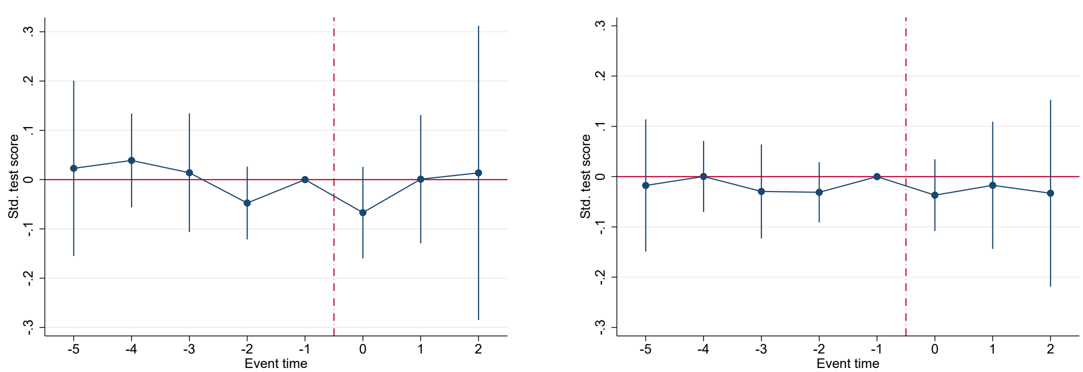
# 6 Results - Student Achievement

## 6.1 Main Pattern

In this section, we test if edTPA benefits student learning using the student sample whose teachers obtained the standard license through the traditional-route program.

Figure 3 plots the estimates of event study dummies for standardized test scores to visualize the trend. We combine the three subjects - namely mathematics in grade 4, reading in grades 4 and 8 - to show an aggregate picture. On the left, we do not include any control variables to demonstrate the pattern in the raw data, conditional on subject, year, and state fixed effects. On the right, we present the trend using the full specification. In both figures, there is no significant deviation in pre-treatment trend that validates the assumption of the difference-in-differences approach. Under the full specification, the pre-treatment trend is more stable and flat, demonstrating the importance of controlling time-varying factors and concurrent education policies. In both specifications, we do not see edTPA imposed any significant effect on student test scores.

**Figure 3:** Event study figure for student test score




**(a)** No controls **(b)** With controls

*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 underlying regression of subfigure (a) contains no controls to show raw data patterns, while subfigure (b) contains student, school and policy controls, conditional on state and year fixed effects, as well as assessment fixed effects. The figures show the 95% confidence interval with robust standard errors clustered at the state level.

We then present the difference-in-differences estimates by subject - namely Grade 4 reading, Grade 4 mathematics, and Grade 8 reading – in Table 10. In Column 1, 4, and 7, where the regressions only include state and year fixed effects, edTPA has a negative coefficient in all the three samples but the estimates are all statistically insignificant.

In the full specifications in Column 3, 6, and 9, adding student, school, and policy controls do not alter the null effects. In fact, the effect size also economically insignificant. For example, in Column 3, compared to the students in non-edTPA states, students in the edTPA states only have a lower score by 0.5% of a standard deviation.[31]

We also check the need to use alternative DID methods by assessing the 'bad comparison' problem. Consistent with the null effect findings, the De Chaisemartin and d'Haultfoeuille (2020) decomposition shows that none of the weights in our estimation samples are negative, and the sum of the negative weights is also zero. The analysis indicates the 'bad comparison' problem is not a concern when interpreting estimates in Table 10.

As stated in the data section, the reason we focus on students with traditional route teachers is that the parallel trend assumption does not hold in the alternative route sample. As we demonstrate in Figure C1 and C2, the pre-treatment trends are largely bumpy and have significant deviation in most of the cases. Nonetheless, for completeness, we also present the main pattern using the alternative-route sample in the appendix.[32]

[31]In Table C1 of appendix, when we include the optional states in the control group, the negative effect in Grade 4 reading is slightly stronger than the main result but the effect size is only marginally significant.

[32]We include the DID estimates in the Panel C of Table C2 of Appendix. Because of the pre-trend deviation, we do not over-interpret the negative effects in the alternative-route sample.

**Table 10:** Impacts of edTPA on students' achievement

| | Std. test score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Grade 4 Math | | | Grade 4 Reading | | | Grade 8 Reading | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| edTPA | -0.030 | -0.023 | -0.005 | -0.048 | -0.044 | -0.054 | -0.059 | 0.004 | 0.010 |
| | (0.051) | (0.032) | (0.033) | (0.038) | (0.026) | (0.034) | (0.051) | (0.039) | (0.044) |
| R-squared | 0.029 | 0.331 | 0.332 | 0.025 | 0.309 | 0.370 | 0.036 | 0.371 | 0.372 |
| Observations | 51,460 | 51,460 | 51,460 | 53,530 | 53,530 | 53,530 | 41,260 | 41,260 | 41,260 |
| State FE | X | X | X | X | X | X | X | X | X |
| Year FE | X | X | X | X | X | X | X | X | X |
| Student controls | | X | X | | X | X | | X | X |
| School controls | | X | X | | X | 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 teachers obtained the license through a traditional teacher preparation program. The dependent variables in column (1) to (3), (4) to (6), and (7) to (9) are Grade 4 Math, Grade 4 Reading, and Grade 8 Reading, respectively. The test scores are standardized to a zero mean and one standard deviation in the sample. 'edTPA' refers to the treatment indicator. All regressions include state 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. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

## 6.2 Further Analysis

### 6.2.1 Balancing Test

Our student sample is based on the years of experience of teachers. One concern of this sample selection criteria 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, our negative estimates would falsely be attributed to 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 and year fixed effects. As shown in the Table 11, edTPA in general is not related to changes in most of the predetermined student characteristics (except for Hispanic in Grade 8 reading). Overall, we do not find evidence that there is a systematic sample selection issue in our estimation.[33]

### 6.2.2 Heterogeneity

The richness of NAEP allows us to look beyond the effects on average test scores. In this extension, we perform two extra exercises using the traditional route sample. First, we explore the heterogeneity by school characteristic. Second, we perform a quantile regression to look at the distributional effects by student ability. In all the following analyses, we perform the full specification with student, school, and policy controls.

[33] Table C3 of appendix contains the corresponding robustness result using the full sample (both traditional and alternative-route).

**Table 11:** edTPA is not correlated with changes in student characteristics

| | White | Black | Hispanic | Female | IEP | Eng learner |
|---|---|---|---|---|---|---|
| *Panel A. Grade4 Math* | (1) | (2) | (3) | (4) | (5) | (6) |
| edTPA | 0.006 | -0.010 | 0.007 | -0.013 | -0.005 | 0.021 |
| | (0.017) | (0.014) | (0.015) | (0.011) | (0.009) | (0.014) |
| R-squared | 0.088 | 0.112 | 0.067 | 0.001 | 0.007 | 0.047 |
| Observations | 51,460 | 51,460 | 51,460 | 51,460 | 51,460 | 51,460 |
| *Panel B. Grade4 Reading* | (7) | (8) | (9) | (10) | (11) | (12) |
| edTPA | -0.019 | -0.009 | 0.022 | 0.005 | -0.003 | 0.002 |
| | (0.018) | (0.019) | (0.017) | (0.013) | (0.013) | (0.018) |
| R-squared | 0.083 | 0.098 | 0.068 | 0.001 | 0.009 | 0.048 |
| Observations | 53,530 | 53,530 | 53,530 | 53,530 | 53,530 | 53,530 |
| *Panel C. Grade8 Reading* | (13) | (14) | (15) | (16) | (17) | (18) |
| edTPA | -0.044 | -0.001 | 0.036*** | -0.012 | 0.014 | 0.027 |
| | (0.027) | (0.019) | (0.013) | (0.015) | (0.015) | (0.017) |
| R-squared | 0.111 | 0.132 | 0.103 | 0.001 | 0.011 | 0.037 |
| Observations | 41,260 | 41,260 | 41,260 | 41,260 | 41,260 | 41,260 |
| State FE | X | X | X | X | X | X |
| Year FE | X | X | X | X | X | X |

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The traditional route samples in panel A, B, and C are from three NAEP assessments: Grade 4 Math, Grade 4 Reading, and Grade 8 Reading, respectively. 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 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. $***p < 0.01, **p < 0.05, *p < 0.1$.

***Heterogeneity by School Type -*** 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 hard-to-staff characteristics, namely the percent of black students, the percent of Hispanic students, 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 Panel A and B, we do not find heterogeneity by the four school characteristics in Grade 4 reading and Grade 4 mathematics. In Panel C, we find a strong differential effect if the school is majority-black at the 1% significance level. While the coefficient is positive, the magnitude is economically small such that combining the interaction with the base term still cannot reject the null effect hypothesis.

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

| | **Student outcome: Std. test score** | | | |
|---|---|---|---|---|
| X≡ | % Black | % Hispanic | Urban | Free lunch |
| *Panel A. Grade 4 Math* | (1) | (2) | (3) | (4) |
| edTPA*X | 0.001 | 0.001 | 0.068 | -0.023 |
| | (0.001) | (0.001) | (0.067) | (0.050) |
| edTPA | -0.023 | -0.019 | -0.030 | 0.009 |
| | (0.044) | (0.028) | (0.041) | (0.040) |
| R-squared | 0.332 | 0.334 | 0.332 | 0.332 |
| Observations | 51,460 | 51,460 | 51,460 | 51,460 |
| *Panel B. Grade 4 Reading* | (5) | (6) | (7) | (8) |
| edTPA*X | 0.000 | 0.000 | 0.011 | 0.003 |
| | (0.000) | (0.001) | (0.044) | (0.039) |
| edTPA | -0.055* | -0.057* | -0.058* | -0.056 |
| | (0.030) | (0.032) | (0.033) | (0.044) |
| R-squared | 0.372 | 0.373 | 0.372 | 0.372 |
| Observations | 53,530 | 53,530 | 53,530 | 53,530 |
| *Panel C. Grade 8 Reading* | (9) | (10) | (11) | (12) |
| edTPA*X | 0.002*** | -0.003 | 0.092 | 0.042 |
| | (0.001) | (0.043) | (0.055) | (0.060) |
| edTPA | -0.038 | 0.025 | -0.033 | -0.016 |
| | (0.050) | (0.175) | (0.058) | (0.067) |
| R-squared | 0.372 | 0.373 | 0.372 | 0.372 |
| Observations | 41,260 | 41,260 | 41,260 | 41,260 |
| 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 |

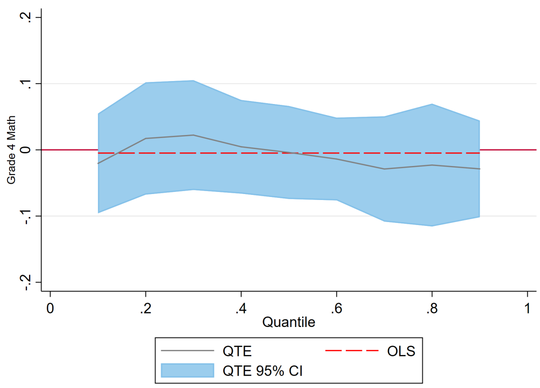
*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The table uses student samples with teachers obtained the license through a traditional teacher preparation program. The sample in Panel A, B, and C is from Grade 4 Math, Grade 4 Reading, and Grade 8 Reading assessment, respectively. The 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. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

***Distributional Effects -*** We now investigate the heterogeneous effects by student's ability through a quantile regression model (Koenker and Hallock, 2001). The idea is to check if the effect on test scores falls disproportionate on certain types of students, and the average effect from the linear model may mask important differential effects.

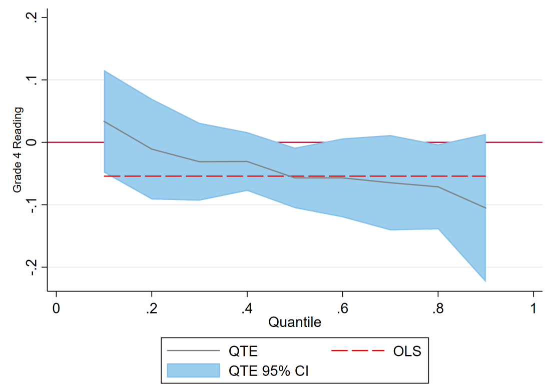
In each sub-figure of Figure 4, the horizontal axis tracks the percentiles and the vertical axis indicates the estimate of the edTPA impacts on achievement at a given percentile. All regressions include the same student, school, and policy controls. The gray dashed line shows the estimate of quantile treatment effects, while the red solid line shows the OLS estimate for comparison. The blue shaded region indicates the corresponding 95% confidence intervals for the quantile regression estimates.

The impacts of edTPA are close to zero for students of all types. The first sub-figure of Figure 4 shows no heterogeneous result for the mathematics score of grade 4 students. For Grade 4 reading in the second sub-figure, we observe that negative effect emerges from the students at the median and concentrates at the higher score percentiles. However, the magnitudes are not statistically significant. For grade 8 reading scores in the third sub figure, we again do not see significant heterogeneity by student ability.

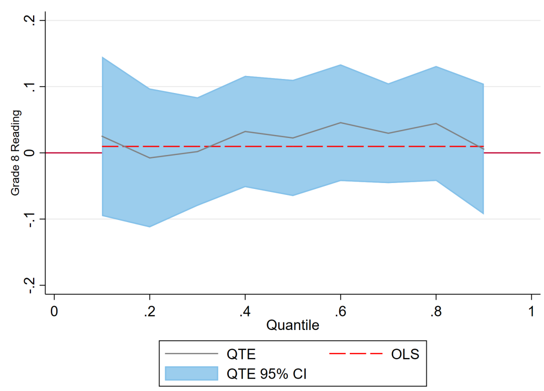
**Figure 4:** Heterogeneous effects of edTPA reforms on student achievement




**(a)** Grade 4: Math




**(b)** Grade 4: Reading




**(c)** Grade 8: Reading

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The figure shows estimates using student samples with teachers obtained the license through a traditional teacher preparation program. The dependent variables of subfigure (a), (b), and (c) are the standardized grade 4 math score, grade 4 reading score, and grade 8 reading score, respectively. All regressions include state fixed effect, year fixed effects, student and school controls, and policy controls. The gray dashed line shows the estimate of quantile treatment effects, while the red solid line shows the OLS estimate. The blue shaded region shows the 95% confidence interval with robust standard errors clustered at the state level.

# 7 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 are alarming to 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).

The challenge to evaluate the benefits of edTPA is to find an accurate measure for the quality effect. While the student test score is not the only quality aspect, it concerns consumers (i.e. parents and education stakeholders) the most. NAEP provides us a unique test score data to identify students with new teachers among a nationally representative sample of students in the US. Testing different specifications, sample criterion, and heterogeneity, we do not find evidence that edTPA improved student test scores of the new teachers. 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. Therefore, 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.

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.

**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**, "Job market signaling through occupational licensing," Technical Report, National Bureau of Economic Research 2018.

_ **and** _ , "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," Technical Report, National Bureau of Economic Research 2020.

**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**, "Trade-Offs of Occupational Licensing: Understanding the Costs and Potential Benefits," *Available at SSRN 3707990*, 2020.

**Clotfelter, Charles T, Helen F Ladd, and Jacob L Vigdor**, "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.

**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**, "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.

**Koenker, Roger and Kevin F Hallock**, "Quantile regression," *Journal of economic perspectives*, 2001, *15* (4), 143–156.

**Kraft, Matthew A, 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.

**Meuwissen, Kevin, Jeffrey Choppin, Kathryn Cloonan, and Hairong Shang-Butler**, "Teaching candidates' experiences with the edTPA as an initial certification test in New York and Washington States: Survey results from the second year of a two-year study," *Retrieved on August*, 2016, *12*, 2016.

**Meuwissen, Kevin W and Jeffrey M Choppin**, "Preservice teachers' adaptations to tensions associated with the edTPA during its early implementation in New York and Washington states.," *Education Policy Analysis Archives*, 2015, *23* (103), n103.

**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.

**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**

**October, 2022**

# 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.

**Table A2:** Summary statistics (NAEP) - Include optional states

| | Grade 4 Math | Grade 4 Reading | Grade 8 Reading |
|---|---|---|---|
| **A. Outcomes:** | | | |
| Assessment score | 235.44 (28.22) | 216.08 (34.62) | 251.99 (37.37) |
| **B. Student controls:** | | | |
| White | 0.43 (0.49) | 0.43 (0.50) | 0.48 (0.50) |
| Black | 0.15 (0.36) | 0.15 (0.36) | 0.14 (0.35) |
| Hispanic | 0.29 (0.45) | 0.29 (0.45) | 0.25 (0.43) |
| Female | 0.49 (0.50) | 0.50 (0.50) | 0.49 (0.50) |
| Individualized Education Program (IEP) | 0.13 (0.33) | 0.12 (0.32) | 0.11 (0.31) |
| English learner | 0.10 (0.31) | 0.10 (0.30) | 0.06 (0.24) |
| **C. School controls:** | | | |
| Charter school | 0.05 (0.22) | 0.06 (0.23) | 0.05 (0.22) |
| Urban area | 0.77 (0.42) | 0.77 (0.42) | 0.74 (0.44) |
| Share of black student | 18.38 (26.09) | 18.21 (25.75) | 16.76 (25.07) |
| Lunch program | 0.58 (0.49) | 0.57 (0.49) | 0.52 (0.50) |
| Student enrollment ($\geq$ 500) | 0.49 (0.50) | 0.49 (0.50) | 0.52 (0.50) |
| **Number of Student** | 70,390 | 72,970 | 56,940 |

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* This table shows summary statistics of the sample (students with new teachers) for student achievement using NAEP, including optional states. The mean is shown in the cell while the standard deviation is shown in the parentheses. Each column presents one of the three student assessment samples: Math at Grade 4, Reading at Grade 4, and Reading at Grade 8. Raw assessment scores are reported in the summary statistics. The number of observations is rounded to the nearest 10 per IES disclosure guidelines.

**Table A3:** Summary statistics (NAEP) - Estimation sample, by traditional/alternative route

| | Grade4 Math | | Grade4 Reading | | Grade8 Reading | |
|---|---|---|---|---|---|---|
| | Traditional | Alternative | Traditional | Alternative | Traditional | Alternative |
| **A. Outcomes:** | | | | | | |
| Assessment score | 235.76 | 227.99 | 216.75 | 208.02 | 252.97 | 247.66 |
| | (28.20) | (29.49) | (34.39) | (36.09) | (37.37) | (36.14) |
| **B. Student controls:** | | | | | | |
| White | 0.46 | 0.28 | 0.46 | 0.29 | 0.50 | 0.33 |
| | (0.50) | (0.45) | (0.50) | (0.45) | (0.50) | (0.47) |
| Black | 0.14 | 0.27 | 0.13 | 0.26 | 0.13 | 0.27 |
| | (0.35) | (0.44) | (0.34) | (0.44) | (0.34) | (0.44) |
| Hispanic | 0.28 | 0.34 | 0.29 | 0.33 | 0.24 | 0.28 |
| | (0.45) | (0.48) | (0.45) | (0.47) | (0.43) | (0.45) |
| Female | 0.49 | 0.48 | 0.50 | 0.49 | 0.49 | 0.49 |
| | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) |
| IEP | 0.13 | 0.13 | 0.12 | 0.13 | 0.11 | 0.13 |
| | (0.33) | (0.34) | (0.33) | (0.34) | (0.32) | (0.34) |
| English learner | 0.10 | 0.13 | 0.09 | 0.12 | 0.06 | 0.07 |
| | (0.29) | (0.33) | (0.29) | (0.32) | (0.23) | (0.26) |
| **C. School controls:** | | | | | | |
| Charter school | 0.05 | 0.10 | 0.05 | 0.11 | 0.04 | 0.09 |
| | (0.22) | (0.30) | (0.22) | (0.31) | (0.20) | (0.29) |
| Urban area | 0.77 | 0.84 | 0.77 | 0.83 | 0.74 | 0.80 |
| | (0.42) | (0.37) | (0.42) | (0.38) | (0.44) | (0.40) |
| Share of black student | 16.68 | 34.06 | 16.50 | 33.34 | 15.73 | 31.54 |
| | (25.64) | (36.32) | (25.32) | (35.93) | (25.35) | (34.33) |
| Lunch program | 0.56 | 0.70 | 0.56 | 0.69 | 0.50 | 0.63 |
| | (0.50) | (0.46) | (0.50) | (0.46) | (0.50) | (0.48) |
| Student enrollment ($\geq$ 500) | 0.45 | 0.48 | 0.45 | 0.49 | 0.50 | 0.51 |
| | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) |
| **Observations** | 51,460 | 9,040 | 53,530 | 9,470 | 41,260 | 12,680 |

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* This table shows summary statistics of estimation sample (students with new teachers) for student achievement using NAEP, by traditional and alternative route of teachers. The mean is shown in the cell while the standard deviation is shown in the parentheses. Each column presents one of the three student assessment samples: Math at Grade 4 , Reading at Grade 4, and Reading at Grade 8. Raw assessment scores are reported in the summary statistics. The number of observations is rounded to the nearest 10 per IES disclosure guidelines.

# 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 |
| Reform (Kraft et al., 2020) | | | X | X |
| Regional trends | | | | X |

*Source:* Title II, 2011-2019

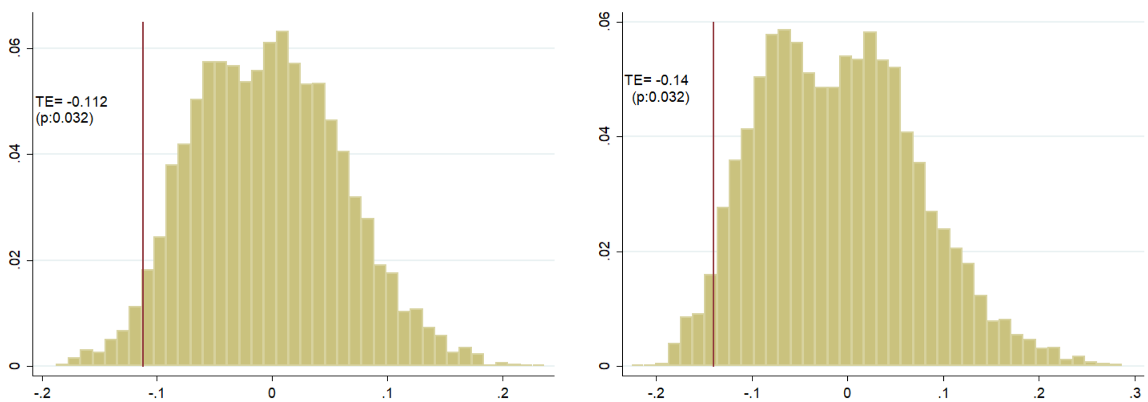
*Notes:* Washington is dropped due to the measurement errors in the data. Dependent variable in all regressions is the log of the number of initial teacher licensure issued in a state. All regressions include year and state fixed effects, and state-level time-varying controls (unemployment rate, percent 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.

**Table B2:** Robustness - Including optional states in the control group

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| edTPA | -0.232*** | -0.147*** | -0.141*** | -0.130*** | -0.131*** |
| | (0.0669) | (0.0378) | (0.0336) | (0.0278) | (0.0342) |
| R-squared | 0.176 | 0.186 | 0.188 | 0.204 | 0.218 |
| Observations | 10,598 | 10,598 | 10,598 | 10,598 | 8,918 |
| Confounding policies# | | X | X | X | X |
| Reform (Kraft et al., 2020) | | | X | X | X |
| Regional trend | | | | X | X |
| Conditional Sample## | | | | | X |

*Source:* IPEDS, 2011-2019.
*Notes:* The sample in all regressions includes the optional states (Arkansas, California, Delaware, Hawaii, Iowa, Maryland, Minnesota, North Carolina, South Carolina, West Virginia. Ohio, and Texas) in the control group. 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 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.
## Conditional sample drops control states that did not implement the accountability reforms studied by Kraft et al. (2020). They include California, Iowa, Montana, Nebraska, Vermont, and Wyoming. We also drop Illinois and Washington as the treatment states since they implement the reforms after the edTPA.

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




**Note:** The permutation tests in this figure construct the distribution of placebo effects (10,000 rounds of permutation) using the non-edTPA states in the 'conditional' sample that implemented accountability reforms (Kraft et al., 2020). The first figure compares our treatment effect in Column 4 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 the treatment effect in Column 5 of Table 3 with another set of placebo treatments. The placebo treatments mimic the implementation timing of edTPA relative to the teacher accountability reform in the six edTPA states: two implemented edTPA 1 year after; two implemented edTPA 2 years after; and the remaining two implemented edTPA 4 and 5 years after.

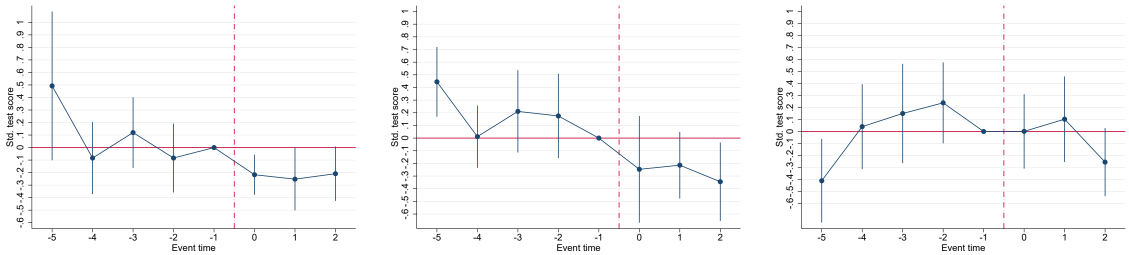
**Table B3:** Heterogeneity by the type of university - Traditional and alternative route

| | (1) Total | (2) White | (3) Non-white | (4) Black | (5) Hispanic | (6) Other race |
|---|---|---|---|---|---|---|
| *Panel A: X = University ranks at top 50% (SAT score $25^{th}$percentile)* | | | | | | |
| edTPA | -0.120*** | -0.0874*** | -0.151*** | -0.109 | -0.0761 | -0.129*** |
| | (0.0316) | (0.0300) | (0.0481) | (0.0939) | (0.0472) | (0.0473) |
| edTPA*X | 0.0208 | -0.0271 | 0.112* | 0.0950 | 0.0886 | 0.0524 |
| | (0.0398) | (0.0486) | (0.0629) | (0.109) | (0.0599) | (0.0392) |
| R-squared | 0.253 | 0.261 | 0.097 | 0.076 | 0.032 | 0.100 |
| *Panel B: X = University ranks at top 50% (ACT score $25^{th}$percentile)* | | | | | | |
| edTPA | -0.144*** | -0.112*** | -0.161*** | -0.0993 | -0.0867 | -0.142*** |
| | (0.0370) | (0.0409) | (0.0524) | (0.0899) | (0.0652) | (0.0489) |
| edTPA*X | 0.0705 | 0.0234 | 0.138 | 0.0793 | 0.114 | 0.0825 |
| | (0.0750) | (0.0856) | (0.0893) | (0.104) | (0.102) | (0.0586) |
| R-squared | 0.253 | 0.261 | 0.098 | 0.075 | 0.032 | 0.101 |
| *Panel C: X = Minority students exceed 50% (non-education majors)* | | | | | | |
| edTPA | -0.0728** | -0.0959*** | 0.00163 | 0.00419 | -0.0448 | 0.0130 |
| | (0.0328) | (0.0313) | (0.0387) | (0.0556) | (0.0267) | (0.0314) |
| edTPA*X | -0.0793* | -0.00209 | -0.224*** | -0.130*** | 0.0331 | -0.278*** |
| | (0.0463) | (0.0413) | (0.0475) | (0.0462) | (0.0290) | (0.0712) |
| R-squared | 0.254 | 0.270 | 0.165 | 0.067 | 0.027 | 0.155 |
| Observations | 7,204 | 7,204 | 7,204 | 7,204 | 7,204 | 7,204 |

*Sources:* IPEDS, 2011-2019.
*Notes:* 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 top 50% according to their pre-2014 characteristics, namely the $25^{th}$ percentile SAT in Panel A, ACT scores in Panel B, and minority (non-white students) concentration in non-education majors in Panel C. 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.

# C Additional Results on Student Outcomes

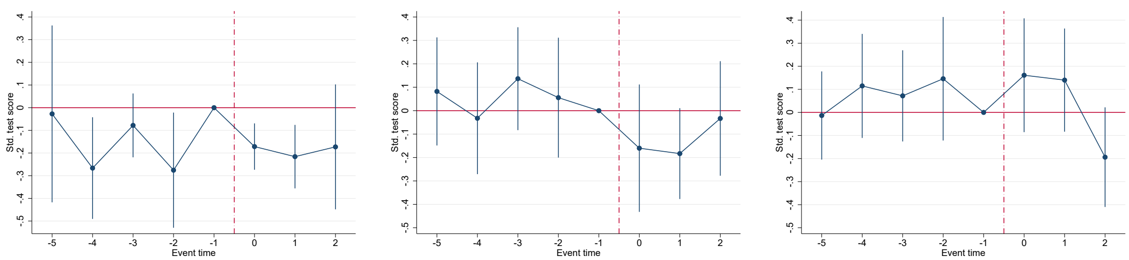
**Figure C1:** Event study figures for the alternative route sample: No controls




**(a)** Grade4 Math  **(b)** Grade4 Reading  **(c)** Grade8 Reading

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The figure shows estimates using student samples with teachers obtained the license through alternative routes. The dependent variable is the standardized test score for Grade 4 Math (subfigure a), Grade 4 Reading (subfigure b), and Grade 8 Reading (subfigure c). Event period -1 is normalized to 0. The underlying regressions contain no controls to show raw data patterns, conditional on state, and year fixed effects. The figures show the 95% confidence interval with robust standard errors clustered at the state level.

**Figure C2:** Event study figures for the alternative route sample: With controls




**(a)** Grade4 Math  **(b)** Grade4 Reading  **(c)** Grade8 Reading

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The figure shows estimates using student samples with teachers obtained the license through alternative routes. The dependent variable is the standardized test score for Grade 4 Math (subfigure a), Grade 4 Reading (subfigure b), and Grade 8 Reading (subfigure c). Event period -1 is normalized to 0. The underlying regressions contain student and school controls listed in Table 2, conditional on state, and year fixed effects. The figures show the 95% confidence interval with robust standard errors clustered at the state level.

**Table C1:** Robustness check: Impacts of edTPA on students' achievement - Include optional states

| | Std. test score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Grade 4 Math | | | Grade 4 Reading | | | Grade 8 Reading | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| edTPA | -0.030 | -0.019 | -0.004 | -0.044 | -0.044* | -0.063* | -0.065 | 0.001 | 0.011 |
| | (0.047) | (0.029) | (0.029) | (0.036) | (0.025) | (0.033) | (0.051) | (0.037) | (0.040) |
| State FE | X | X | X | X | X | X | X | X | X |
| Year FE | X | X | X | X | X | X | X | X | X |
| Student controls | | X | X | | X | X | | X | X |
| School controls | | X | X | | X | X | | X | X |
| Policy controls # | | | X | | | X | | | X |
| R-squared | 0.029 | 0.328 | 0.329 | 0.023 | 0.309 | 0.369 | 0.37 | 0.329 | 0.372 |
| Observations | 70,390 | 70,390 | 70,390 | 72,970 | 72,970 | 72,970 | 56,940 | 56,940 | 56,940 |

*Data Source:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The estimations include students with teachers obtained the license through a traditional teacher preparation program, with states where edTPA is optional for teacher licensure included in the control group. The dependent variables in column (1) to (3), (4) to (6), and (7) to (9) are Grade 4 Math, Grade 4 Reading, and Grade 8 Reading, respectively. The test scores are standardized to a zero mean and one standard deviation in the 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. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

**Table C2:** Impacts of edTPA reforms on students' achievement

| | Std. test score | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Panel A** | Grade 4 Math | | | Grade 4 Reading | | | Grade 8 Reading | | |
| *Full sample* | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| edTPA | -0.038 | -0.039 | -0.005 | -0.080* | -0.080*** | -0.074** | -0.058 | 0.010 | 0.019 |
| | (0.050) | (0.030) | (0.027) | (0.042) | (0.029) | (0.029) | (0.040) | (0.039) | (0.044) |
| R-squared | 0.033 | 0.339 | 0.341 | 0.027 | 0.379 | 0.380 | 0.040 | 0.379 | 0.379 |
| Observations | 60,500 | 60,500 | 60,500 | 63,000 | 63,000 | 63,000 | 53,940 | 53,940 | 53,940 |
| **Panel B** | | | | | | | | | |
| *Traditional route* | (10) | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
| edTPA | -0.030 | -0.023 | -0.005 | -0.048 | -0.044 | -0.054 | -0.059 | 0.004 | 0.010 |
| | (0.051) | (0.032) | (0.033) | (0.038) | (0.026) | (0.034) | (0.051) | (0.039) | (0.044) |
| R-squared | 0.029 | 0.331 | 0.332 | 0.025 | 0.309 | 0.370 | 0.036 | 0.371 | 0.372 |
| Observations | 51,460 | 51,460 | 51,460 | 53,530 | 53,530 | 53,530 | 41,260 | 41,260 | 41,260 |
| **Panel C** | | | | | | | | | |
| *Alternative routes* | (19) | (20) | (21) | (22) | (23) | (24) | (25) | (26) | (27) |
| edTPA | -0.254** | -0.156* | -0.053 | -0.380*** | -0.303*** | -0.218*** | -0.070 | 0.037 | 0.031 |
| | (0.104) | (0.083) | (0.067) | (0.124) | (0.089) | (0.061) | (0.081) | (0.054) | (0.060) |
| R-squared | 0.066 | 0.357 | 0.363 | 0.061 | 0.337 | 0.405 | 0.055 | 0.381 | 0.382 |
| Observations | 9,040 | 9,040 | 9,040 | 9,470 | 9,470 | 9,470 | 12,680 | 12,680 | 12,680 |
| State FE | X | X | X | X | X | X | X | X | X |
| Year FE | X | X | X | X | X | X | X | X | X |
| Student controls | | X | X | | X | X | | X | X |
| School controls | | X | X | | X | X | | X | X |
| Policy controls # | | | X | | | X | | | X |

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* Panel A uses full student sample, Panel B uses student samples with teachers obtained the license through a traditional teacher preparation program, and Panel C uses students with teachers obtained the license through alternative routes. Samples in all panels exclude students in optional states. The dependent variables in column (1) to (3), (4) to (6), and (7) to (9) are Grade 4 Math, Grade 4 Reading, and Grade 8 Reading, respectively. The test scores are standardized to a zero mean and one standard deviation in the 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. ***, **, and * represent 1%, 5%, and 10% significant level, respectively.

**Table C3:** Balancing test: edTPA is not correlated with changes in student characteristics - Full sample

| | White | Black | Hispanic | Female | IEP | Eng learner |
|---|---|---|---|---|---|---|
| *Panel A. Grade4 Math* | (1) | (2) | (3) | (4) | (5) | (6) |
| edTPA | -0.008 | -0.012 | 0.011 | -0.012 | -0.007 | 0.014 |
| | (0.018) | (0.015) | (0.016) | (0.010) | (0.010) | (0.014) |
| R-squared | 0.097 | 0.135 | 0.071 | 0.001 | 0.007 | 0.049 |
| Observations | 60,500 | 60,500 | 60,500 | 60,500 | 60,500 | 60,500 |
| *Panel B. Grade4 Reading* | (7) | (8) | (9) | (10) | (11) | (12) |
| edTPA | -0.012 | -0.013 | 0.020 | 0.011 | 0.002 | 0.005 |
| | (0.021) | (0.021) | (0.018) | (0.013) | (0.014) | (0.018) |
| R-squared | 0.092 | 0.125 | 0.072 | 0.001 | 0.008 | 0.052 |
| Observations | 63,000 | 63,000 | 63,000 | 63,000 | 63,000 | 63,000 |
| *Panel C. Grade8 Reading* | (13) | (14) | (15) | (16) | (17) | (18) |
| edTPA | -0.045* | 0.006 | 0.036** | -0.014 | 0.011 | 0.023 |
| | (0.026) | (0.017) | (0.014) | (0.014) | (0.016) | (0.014) |
| R-squared | 0.126 | 0.162 | 0.114 | 0.001 | 0.011 | 0.039 |
| Observations | 53,940 | 53,940 | 53,940 | 53,940 | 53,940 | 53,940 |
| State FE | X | X | X | X | X | X |
| Year FE | X | X | X | X | X | X |

*Sources:* NAEP 2009, 2011, 2013, 2015, 2017, and 2019.
*Notes:* The samples in panel A, B, and C are students with teachers from both traditional and alternative routes, excluding those in optional states, from three NAEP assessments: Grade 4 Math, Grade 4 Reading, and Grade 8 Reading, respectively. 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 fixed effects 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. $***p < 0.01, **p < 0.05, *p < 0.1$.