education policy analysis archives

A peer-reviewed, independent,
open access, multilingual journal

epaa aape

Arizona State University

Volume 32 Number 51 September 10, 2024 ISSN 1068-2341

Measuring the Value of Teachers from Traditional Certification Pathways in Texas: A Comprehensive Study[1]

Annelies Rhodes

&

Michael Marder

The University of Texas at Austin

United States

Citation: Rhodes, A., & Marder, M. (2024). Measuring the value of teachers from traditional certification pathways in Texas: A comprehensive study. *Education Policy Analysis Archives, 32*(51). https://doi.org/10.14507/epaa.32.8556

Abstract: While teacher preparation in the United States continues a long period of decline, the largest-producing state, Texas, is experiencing substantial changes in how it prepares teachers. The number of teachers prepared by traditional university pathways continues to decline, and the number from alternative pathways is rising. Using extensive data from Texas, we find that traditionally prepared teachers from universities obtain significantly higher student learning gains than alternatives. We use value-added models to estimate changes in student test scores in many grade levels and test subjects as a function of teacher preparation pathway. We compare all Traditional programs to all Alternative programs, and we compare all For-Profit programs to all Not for-Profit programs. For most subjects and grade levels, students learn significantly more from Traditional or Not for-Profit program teachers: 0.02 to 0.05 in standard deviation units. There is not one significant estimate in any model where

[1]Partial funding support for this work was provided by Dean Charles Martinez of the University of Texas at Austin College of Education as part of a study of Educator Pathways in Texas. Additional support came from endowment of the UTeach Science Program at The University of Texas at Austin.

Journal website: http://epaa.asu.edu/ojs/
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 4/2/2024
Revisions received: 24/6/2024
Accepted: 7/8/2924

students learn more from Alternative and For-Profit programs teachers than they do from Traditional and Not For-Profit program teachers.
Keywords: teacher preparation; alternative certification; value-added models

Midiendo el valor de los docentes de las vías de certificación tradicionales en Texas: Un estudio integral
Resumen: Mientras que la preparación de los docentes en los Estados Unidos continúa un largo período de declive, el estado con mayor producción, Texas, está cambiando la forma en que prepara a los docentes. El número de docentes preparados por vías universitarias tradicionales sigue disminuyendo, y el número de los que proceden de vías alternativas aumenta. Utilizando datos extensos de Texas, descubrimos que los docentes preparados en universidades obtienen ganancias de aprendizaje significativamente mayores para los estudiantes que los que proceden de vías alternativas. Utilizamos modelos de valor agregado para estimar los cambios en las calificaciones de los estudiantes en los exámenes de muchos niveles de grado y materias de prueba en función de la vía de preparación de los docentes. Comparamos todos los programas tradicionales con todos los programas alternativos, y comparamos todos los programas con fines de lucro con todos los programas sin fines de lucro. Para la mayoría de las materias y niveles de grado, los estudiantes aprenden significativamente más de los docentes de programas tradicionales o sin fines de lucro: 0,02 a 0,05 en unidades de desviación estándar. No hay una sola estimación significativa en ningún modelo en la que los estudiantes aprendan más de los docentes de programas alternativos y con fines de lucro que de los docentes de programas tradicionales y sin fines de lucro.
Palabras-clave: preparación docente; certificación alternativa; modelos de valor agregado

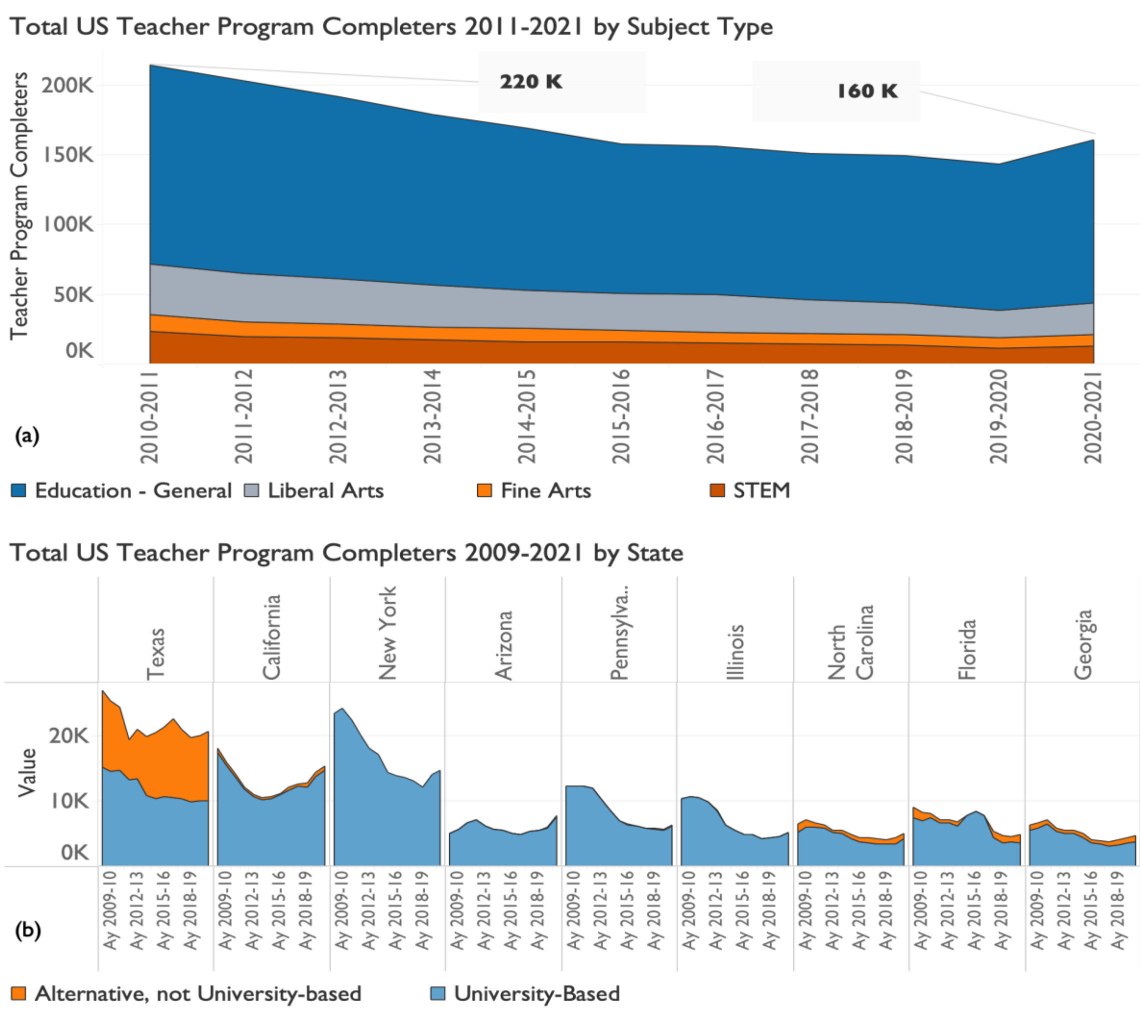
Medindo o valor dos professores dos caminhos de certificação tradicionais no Texas: Um estudo abrangente
Resumo: Enquanto a preparação de professores nos Estados Unidos continua em um longo período de declínio, o maior estado produtor, o Texas, está mudando a forma como prepara os professores. O número de professores preparados por caminhos universitários tradicionais continua a diminuir, e o número de caminhos alternativos aumenta. Usando dados extensivos do Texas, descobrimos que os professores preparados em universidades obtêm ganhos de aprendizagem significativamente maiores do que os alternativos. Usamos modelos de valor agregado para estimar mudanças nas pontuações dos testes dos alunos em muitos níveis de ensino e disciplinas de teste como uma função do caminho de preparação do professor. Comparamos todos os programas tradicionais com todos os programas alternativos e comparamos todos os programas com fins lucrativos com todos os programas sem fins lucrativos. Para a maioria das disciplinas e níveis de ensino, os alunos aprendem significativamente mais com professores de programas tradicionais ou sem fins lucrativos: 0,02 a 0,05 em unidades de desvio padrão. Não há uma estimativa significativa em nenhum modelo em que os alunos aprendam mais com professores de programas alternativos e com fins lucrativos do que com professores de programas tradicionais e sem fins lucrativos.
Palavras-chave: preparação de professores; certificação alternativa; modelos de valor agregado

Measuring the Value of Teachers from Traditional Certification Pathways in Texas: A Comprehensive Study

For decades, the United States has endured teacher shortages for important subjects including Special Education, Computer Science, Physics, Mathematics, and Foreign Languages. Additionally, according to data from the US Department of Education (2022), the US is experiencing a decline in K-12 teacher production at least one decade in duration that threatens to expand the areas of shortage and worsen them. Figure 1a shows the total number of completers of all U.S. teacher preparation programs from 2011 until 2021. During this decade, despite a slight uptick in the latest year, the number of teacher program completers has declined more than 25%. As shown in Figure 1b, also from the U.S. Department of Education (2022), production between 2009 and 2021 dropped in eight of the nine states that produce the most teachers.

Figure 1

Teacher program completers by certificate type (Panel a) and teachers prepared in highest-producing states by program type, 2009-2021 (Panel b)



Source: Data from US Department of Education (2022). *Note:* Teaching area determined from Title 2 Area field except for programs for which this is blank and Subject is used instead.

Texas plays an important role in the U.S. teacher preparation landscape for two reasons. The first, as shown in Figure 1b, is that Texas prepares more new teachers than any other state. Second, also shown in Figure 1b, Texas prepares more than half its teachers through alternative teacher programs not based in institutions of higher education (IHE). The not-IHE-based entities in Texas are mainly for-profit companies with much of their coursework online, including a single large company that prepares more teachers than any other single entity in the United States.

The competing models for teacher certification programs correspond to competing beliefs (Cochran-Smith et al., 2015; Cochran-Smith & Villegas, 2015). To some it seems obvious that teachers whose course of study is largely web-based will not do as well as those carefully mentored in person for months or years before teaching (Kirksey & Gottlieb, 2023; Will, 2022). To others it seems obvious that traditional teacher preparation through universities is expensive and time consuming and can act as a barrier to the teaching profession (Hess, 2002; Maier, 2012; Walsh & Jacobs, 2007). Our aim here is to use the extensive data built up in Texas across the past decade on teachers from traditional and alternative pathways to address these two points of view. We use value-added models to analyze student exam performance as a function of the pathway their teachers took into the profession. Increased numbers of teachers have been retiring and resigning after the pandemic, while decreased teacher production numbers have been exacerbating shortages (Nguyen et al., 2022). As all states consider retaining or opening alternative certification pathways in a similar way, they should be able to learn from the Texas experience about effects on students.

Research Questions

How are elementary and secondary student outcomes in Texas impacted by the policies that regulate the certification routes available to prospective teachers? Specifically, how do value-added estimates for changes in student test scores in various grade levels and test subjects depend upon teacher preparation pathway, including alternative pathways overall, and for-profit alternative pathways? How do value-added estimates vary by student, classroom, and campus demographics, including gender, race/ethnicity, and household income status?

Literature Review

Teacher Shortages and Alternative Certification

When Sutcher et al. (2019) analyzed teacher supply and demand in the United States five years ago, they opened with the observation that "[o]ver the last several years, headlines across the country broadcasted severe teacher shortages." After the pandemic, headlines became even more insistent (Bruno, 2023; Natanson, 2022; Ward, 2023), although not without dissenting voices arguing no shortage exists (Thompson, 2022). It may take years for data to become available that settle the extent of teacher shortages, but the most systematic evaluations indicate they exist at least in specific geographic and subject areas and number in the tens of thousands (Edwards et al., 2022; Nguyen et al., 2022). The overall National Composite Score for perceived teacher demand in 2023 is the highest ever seen since the measure was first defined in 1981 (American Association for Employment in Education, 2023). Thus, it is not surprising to encounter state policies that aim to increase the number of teachers by providing numerous pathways into the profession.

Unfortunately, the goals of quality and quantity in teacher production compete; addressing one problem threatens to make the other worse. Policies aimed at rapidly increasing the quantity of teachers can reduce the quality of the teacher workforce if preparation is inadequate. On the other hand, policies aimed at better teacher preparation to produce higher quality can create barriers that keep candidates from entering the teaching profession. Our purpose in this study is to examine outcomes of a policy that Texas legislators designed to address the problem of quantity—the expansion of alternative (non-university) pathways to teacher certification. Policymakers can balance the tension between quality and quantity through state regulations governing teacher certification (Corcoran, 2007). Traditionally, educator preparation programs (EPPs) at colleges and universities certified most teachers (Boyd et al., 2007), and state level education laws set the policies governing EPPs. During the past several decades, numerous states have created alternative pathways to teacher certification that remove barriers to teaching (Suell & Piotrowski, 2007). Some features of traditional certification that may constitute barriers are admissions requirements, tuition and housing costs, and demands on candidates' time (Carinci et al., 2020). By contrast, alternative certification pathways offer an expedited route to teaching for individuals who already have a bachelor's degree (Grossman & Loeb, 2008).

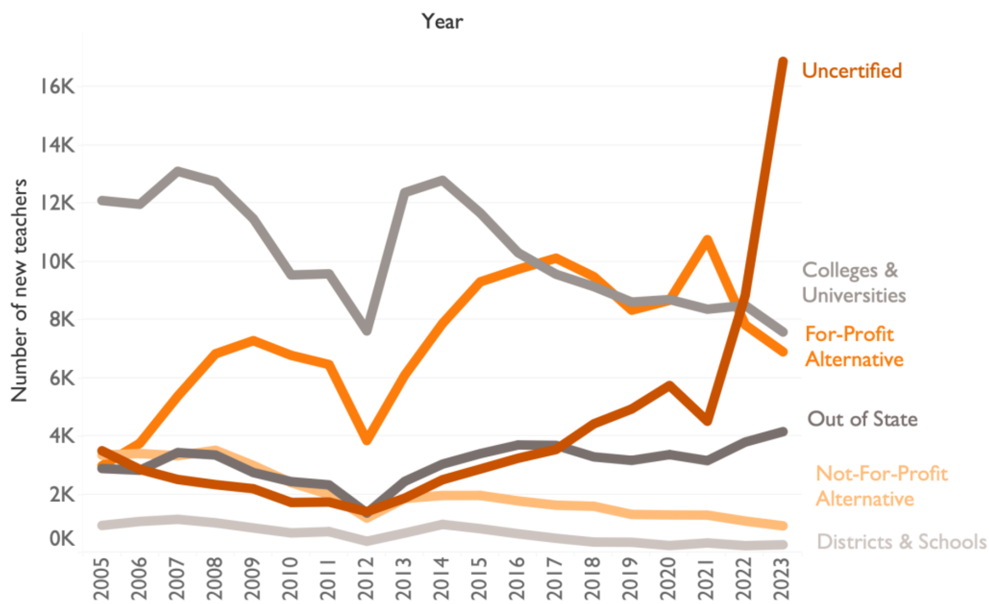
Some advocates of alternative pathways have argued that easing requirements will raise the quality of the teacher workforce by attracting more people to teaching (Boyd et al., 2007). States could put resources into selecting the best teachers based on their performance on the job, rather than preparing them in advance (Gordon et al., 2006; Hess, 2002). Additionally, proponents argue that alternative pathways increase opportunities to recruit teachers of color and teachers in high-need subject areas and schools (Redding, 2022). The quality of EPPs has become the focus of policymakers as a mechanism to improve teacher quality (Cochran-Smith & Villegas, 2015). In principle, traditional and alternative programs are subject to the same state requirements for teacher certification. In practice, alternative programs often avail themselves of provisions traditional programs choose to avoid. This can mean dispensing with components of traditional programs such as a student teaching semester. States can structure alternative certification programs so that teacher preparation is in-service, meaning the teacher candidates teach full time while completing program requirements (Kirksey & Gottlieb, 2023). There is little transparency required of alternative EPPs in terms of pedagogical coursework or consistency in field experience requirements (Day & Mason-Williams, 2023). There are also considerable differences within alternative EPPs in terms of structure (Schmidt et al., 2020). One example of a broad difference is that some alternative EPPs are for-profit companies (Iteach, n.d.; Texas Teachers of Tomorrow, n.d.) while others are non-profits housed in local school districts or other public educational entities, such as regional service centers (Etheredge, 2015).

Certification in Texas

Since states set the guidelines for EPPs and certification, there are differences in teacher pathways in different states. Texas has been the basis for several studies on EPPs due to the fact that it prepares more teachers overall as well as more teachers from non-traditional pathways than any other state (Etheredge, 2015; Marder et al., 2020). Figure 2 shows the number of first-year teachers in Texas according to their preparation route. Texas' State Board for Educator Certification first authorized For-Profit alternative certification in 2002 and it has grown remarkably since then. Texas' for-profit teacher preparation sector is the largest in the US, and the teachers it produces surpassed Texas colleges and universities in 2017. It is noteworthy that the number of new teachers who neither have a teaching certificate nor are enrolled in a program, alternative or otherwise has been growing since 2012, and in 2023 this was by far the dominant path into teaching. In this work we have focused on the contrast between traditional and alternative certification, but the explosive growth of uncertified teachers deserves future study.

Figure 2

First-year teachers in Texas according to their preparation route



Source: Data analysis from SBEC in Texas Educational Research Center (ERC).

Measuring EPP Efficacy with Value-Added Models

States can regulate EPPs through statutes, rules, and funding. Which EPPs they promote depends on beliefs about the effects of different types of teachers on students. Studies of EPPs include analysis of program completion rates, program grades, evaluation ratings of graduates (Bastian et al., 2018), teacher placement, teacher retention, etc. (Cochran-Smith et al., 2015; Cochran-Smith & Villegas, 2015). One common measure of efficacy of EPPs is to analyze teacher quality measured by student learning outcomes. High quality teachers have a large influence on student test scores (Hanushek et al., 2016; Sorensen & Ladd, 2020). One study estimated that a one-standard deviation increase in teacher value-added was larger than the effect of a 10-student reduction in class size (Koedel, Mihaly, et al., 2015). Additionally, teachers with higher value-added scores positively influence later in life outcomes such as wages and college-attendance (Chetty et al., 2014). Another study estimated that in a class of twenty students, a one-standard deviation increase in teacher value-added generates marginal gains of over $400,000 annually in present value of student future earnings and proportionately increases with larger class sizes (Hanushek, 2011).

To isolate the contribution of an EPP to teacher quality (measured by student learning), existing research utilizes value-added models (VAM; Carinci et al., 2020). VAMs estimate teacher value-added by regressing student test scores on their previous year score with an indicator for EPP and controlling for demographic characteristics. Large-scale VAM studies can be useful for evaluating factors impacting test scores and measuring effects of programs and interventions (Darling-Hammond et al., 2012). In value-added studies, standardized test score changes are used to measure student learning, which in turn becomes a measure of the quality of teaching, and that correlates with the teacher's preparation. Many factors contribute to student test score outcomes, some of which are variables available for use in regression models.

The accuracy and potential biases of value-added models are the subject of a growing literature base, which we will discuss in the next section. We believe that VAM estimates can have a predictive value and identify teacher effects on student learning, but there is no consensus in the literature on bias or guidelines for "best practice" (Angrist et al., 2017). VAMs include a mix of student prior test scores, student demographic characteristics, and often control for teacher demographics and classroom and/or campus demographics (Aaronson et al., 2007; Backes et al., 2018; Henry et al., 2014; von Hippel et al., 2016). Most VAM studies use lagged-score modeling, which takes the cumulative effect of prior school or non-school inputs into account to compensate for the systematic bias derived from the non-random assignment of students to teachers from different EPPs (Chetty et al., 2014; Koedel, Parsons, et al., 2015; von Hippel et al., 2016).

There is not complete consensus in the literature on whether to include teacher experience in VAMs as part of the estimate of student learning gains. When looking at EPPs, the desire to include this factor stems from the finding that teachers tend to get better as they gain more experience (Podolsky et al., 2019). However, there are questions about how long a teacher's quality can be assigned to their preparation rather than their experience (Bastian et al., 2018). This is confounded by the fact that teachers from traditional certification pathways typically stay in teaching longer (Marder et al., 2022) and that lower performing teachers are more likely to leave, which can bias the estimates if the cohort is limited to only beginning teachers (Aaronson et al., 2007).

Limitations of VAMs

While VAMs are a common tool in the literature, numerous authors have criticized them for being prone to systemic bias, where small changes to the model based on various assumptions can result in large shifts in the estimates and differences are largely due to sampling error or unmeasured causal factors (Koedel, Mihaly, et al., 2015; Marder et al., 2020; von Hippel et al., 2016). Some articles employing value-added studies that find that certification type and certification programs have no significant effect on teacher quality (Gordon et al., 2006), or that there are no significant differences in educator preparation programs (von Hippel et al., 2016). A broad critique of VAMs is that they primarily rely on standardized test scores as a means to measure change in student learning, when it is debatable that standardized tests legitimately measure what a student knows and is able to do (Amrein-Beardsley, 2009). Another critique of value-added models in relation to teacher preparation pathways is that they do not say anything about the features of the programs that make them successful or ineffective. Rather, they suggest that some programs are better than others but cannot elaborate why (Carinci et al., 2020; von Hippel et al., 2016). These concerns informed the models we constructed. We ran at least four versions of each model to check robustness, compared groups of programs with each other to obtain large sample sizes for which effects would be visible, and set up comparisons between groups of programs in cases where we understood the programmatic differences between the groups.

This Study

Studies using VAMs to compare the value added to student learning from teachers in relation to teacher preparation fall typically into three broad categories: 1) Compare specific EPPs to other EPPs, 2) Compare EPPs across a region or state, or 3) Compare pathways to certification. The central aim of this study, which is to understand how different certification pathways connect to student learning outcomes, is in the third of those categories. The starting point was work of Marder et al. (2020) that developed value-added models for Algebra I and Biology teachers from alternative and traditional certification pathways in Texas. Those results were expanded as part of an Educator Pathways study (Marder et al., 2022). The work done here builds on previous studies by broadening the scope. The Educator Pathways study provided an initial release of the analyses reported in this paper. However, for results reported here, we rewrote all the data cleaning routines and all VAM code, as well as added new studies. The overarching goal for these studies is to examine how the enormous growth of alternative certification pathways in Texas is impacting student learning.

This study is one of the largest studies to use value-added models to estimate a teacher's contribution to student learning outcomes across multiple years, grade levels, and subject areas. The data encompasses students and teachers across Texas from 2012 to 2019 in Grades 3 through 11 from a range of subject areas including reading, mathematics, science, and history. Additionally, we used multiple value-added models that address both potential bias in the VAMs as well as different causal assumptions for what variables contribute most to student learning.

Data and Background

Source

This study uses longitudinal data available in the Texas Education Research Center (ERC; https://texaserc.utexas.edu/), which is a repository from several Texas state agencies that provide student and teacher level data. We included data from the Texas Education Agency (TEA), the Public Education Information Management System (PEIMS), as well as the State Board for Educator Certification (SBEC) in this study. Within the data available in ERC, the ability to link students to teachers starts with the data from 2012, which is why the current study analyzes students and teachers in the time period from 2012 through 2019. We built the models using student standardized test scores from the State of Texas Assessments of Academic Readiness (STAAR), which is an annual assessment for Texas students. STAAR tests are typically administered in the Spring for Grades 3-8 in Mathematics and Reading Language Arts, in Science Grades 5 and 8, Social Studies Grade 8, and End-of-Course (EOC) for Algebra I, English I, English II, Biology, and U.S. History. The English I, English II, and U.S. History tests began in the current format in 2014, so student cohorts in these subjects were limited to 2014 through 2019 in this study. The cutoff was 2019 because the pandemic made testing in 2020 and 2021 difficult to interpret.

Teacher Population

This study does not include all teachers employed in Texas from 2012 to 2019. Given reliance on STAAR scores, only teachers from the tested subjects named above are included. Additionally, the focus of this study is on the two most common certification pathways of teachers in Texas. Therefore, only teachers that were certified through a Traditional, Postbaccalaureate, or Alternative pathway are included. We assigned teachers to a certification pathway based on a combination of variables obtained from the data, such as "certification program," "certification type," and "first certification" variables from the SBEC data. We compared teacher certification routes in two ways. The first grouped Traditional and Post-Baccalaureate together and compared to teachers from Alternative certification pathways. The assignment of certification pathway to "Traditional or PostBacc" is based on the inclusion of a student teaching experience as part of the EPP, compared to "Alternative" pathways, where there is not a student teaching component. Alternative certification programs can have a student teaching (or residency) component, but only around 1% of new teachers in Texas each year have followed such a pathway (Marder et al., 2020). Teachers from a traditional pathway are also called "University-Prepared" teachers, given that these EPPs are typically at postsecondary institutions. The second set of analyses grouped this same population of teachers by whether their certification program was "For-Profit" or "Not For-Profit." Not For-Profit EPPs are defined as universities, colleges, school districts, schools, and educational service centers, while For-Profit are the rest of the EPPs. Although teachers often earn multiple certifications during their tenure, given that the focus of this study is on teacher preparation programs, we assigned teachers based on their earliest teacher certification. Since the models are centered on student growth spanning a school year, teachers were only included if they were the teacher of record for the class for an entire year. We excluded teachers that were not assigned to a class from October 1 through May 1 in a given academic year.

The number of unique teachers we analyzed in each tested subject range from 12,117 to 63,813. Apart from the Grade 4 tests, more than 50% of the teachers in each test cohort are alternatively certified. Grade 9 Biology has the highest percentage of alternatively certified teachers at 73%. These counts are based on unique teachers each year, so a teacher is counted for each year they appear in the cohort. Counts are not broken down by year because the differences in counts between years is minimal. Figure S1(a)[1] shows the teacher populations and percentages in each grade/test by Alternative versus Traditional or PostBacc, and S1(b) shows For-Profit or Not For-Profit EPPs. More than 58% of the teachers in each test cohort are from Not For-Profit EPPs.

We excluded teachers certified prior to 2003 from this study. The teacher preparation environment shifted in the past two decades in Texas to allow for many teachers to enter the workforce from alternate pathways, so comparing teachers within this time frame is most relevant to our research questions. In addition, information about preparation pathway prior to 2003 is highly unreliable. Figure S2(a) provides a breakdown of teacher's years of experience in each test cohort for Alternative and Traditional certification pathways. Given the restriction in this study to teachers certified in 2003 or after, the maximum years of experience in each cohort is 16. The differences in years of experience does not vary considerably based on certification pathway, where typically 40-50% of each cohort has 0-3 years of experience, 38-45% have 4-9 years of experience, and 10-20% have 10+ years of experience. Figure S2(b) provides a breakdown of teacher's years of experience in each test cohort for For-Profit and Not For-Profit EPPs. The differences in years of experience do vary based on being a For-Profit EPP, where the majority (46-58%) of those teachers have 0-3 years of experience. Teachers from Not For-Profit EPPs typically have more years of experience than teachers from For-Profit EPPs when compared at each cohort. For example, in the Math-4 cohort, 46% of teachers from Not For-Profit EPPs have 4-9 years of experience and 18% have 10 or more years of experience compared to 41% and 12% for teachers from For-Profit EPPs respectively.

Student Population and Student-Teacher Linking

We made separate student cohorts for each STAAR test that include each available year of student-to-teacher links. In the models, we only included the students linked to teachers based on the criteria described in the previous section. Additionally, we only included students if they had a valid raw STAAR score greater than 0 for two consecutive grade levels in a subject as a pre- and a post-test score. If students had more than one score for the same test due to multiple attempts, we kept the highest score for each student, and we combined all student scores across all the years for which linked student and teacher data were available. For example, the 5th Grade Reading data set includes all 5th grade students in Texas from 2012 through 2019 that took the reading test in both 4th and 5th grade. In this case, 4th grade reading score is the student's "pre-test" and their 5th grade reading score is the student's "post-test." Table S1 shows the pre-test used for each of the 18 post-tests. We used differences in pre- and post-tests to measure student growth, and standardized raw test scores by dividing the student's individual raw score by the maximum score on that test that year.

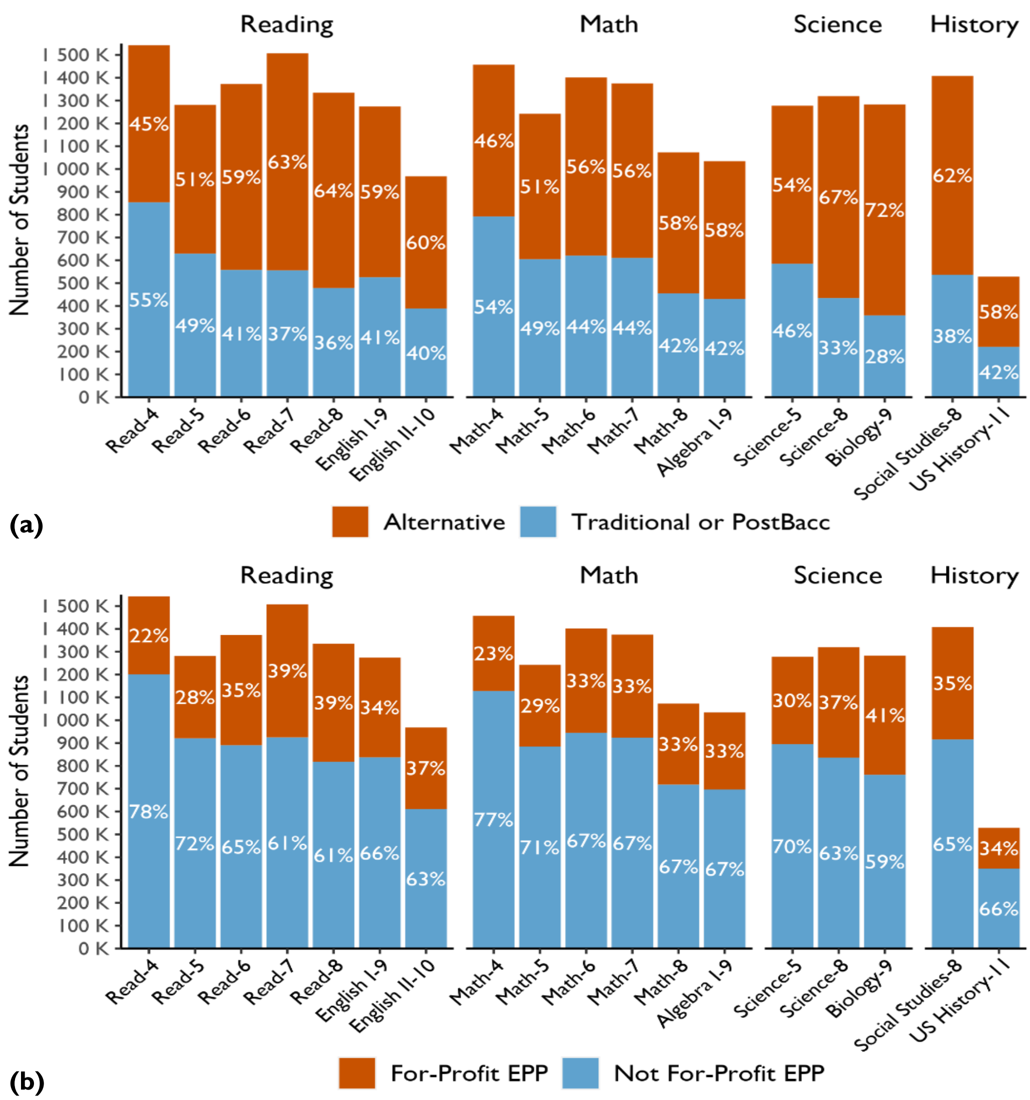
Teachers assigned to each student are the teachers of record for each student's class in the post-test year. For example, the 5th Grade Reading data set described above would include data on

[1] Figures and Tables with labels such as S1 are in supplemental data, available at https://doi.org/10.18738/T8/OMR23L

the student's 5th grade reading teacher. We linked students and teachers through course identification data sets provided by TEA. Given the importance of linking pre- and post-tests with only one year between, we limited each cohort to students in one grade level, which was determined by the grade level most students are in for that STAAR test. For example, we excluded any student that was not in 4th Grade but took the 4th Grade STAAR. Figure 3(a) shows the number of unique students in each test cohort and the percentage of students with teachers from each pathway. Students can be included in the data set more than one time if they have more than one teacher of record in the same year. We assigned each student score a weight in the models based on the number of teachers assigned to the student.

Figure 3

Alternative or Traditional certification pathways (Panel a) and certification through For-Profit or Not For-Profit EPPs (Panel b)



Note: Bars indicate the number of students in each test cohort. Colors indicate the percentage of students in each cohort that were taught by a teacher certified through pathway indicated.

The total number of students in each test cohort is typically between 1 and 1.5 million, except for U.S. History, which had fewer students due to fewer years included in the cohort. Figure 3(b) shows the number of students in each cohort and the percentage of students with teachers from each EPP type when analyzing For-Profit and Not For-Profit EPPs. The number of students in each of these splits follows a similar pattern to that of the number of teachers. More students have teachers from Alternative pathways and more students have teachers from Not For-Profit EPPs.

Student demographic data is based on the student data provided by TEA data sets, including STAAR and PEIMS data. The demographic data comes from student information in the post-test year. Demographic information includes race/ethnicity, household economic status, gifted status, special education status, and emergent bilingual status. Tables S2 and S3 show the cohort averages for these variables, for both Traditional compared to Alternative pathways, as well as For-Profit compared to Not For-Profit EPPs. On average, teachers from Alternative and For-Profit certification pathways have more emergent bilingual students and students from low-income households compared to teachers from Traditional and Not-For-Profit certification pathways. There are not large differences between the pathways for the percentage of gifted and special education students in each Grade/Test cohort. When comparing the race/ethnicity of students in each cohort, a larger percentage of Black and Hispanic students have Alternatively certified teachers, while a larger percentage of White students have teachers with a Traditional or PostBacc certification. With a few exceptions in the percentage of Hispanic students (Science-5 and U.S. History-11), these trends are similar when the cohorts are split by For-Profit or Not for-Profit EPPs.

Methodology

The data preparation and methodology in this study largely align with the original smaller study from (Marder et al., 2020). We measured student growth by analyzing changes in student test scores from one year to the next with four value-added models and built a model for each of the eighteen Grade/Test cohorts separately. Value-added models estimate teacher value added to student learning by regressing student test scores on their previous year score with an indicator for EPP type and controlling for demographic characteristics. The estimates in the outputs of these models predict the value-added to student achievement beyond "expected learning," which would have an estimate of 0. Therefore, positive estimates indicate more than expected learning and negative estimates indicate less than expected learning. These multilevel models are designed to compensate for demographic differences within schools and classrooms. Students may appear multiple times in each cohort due to a variety of factors: multiple teachers assigned to class, moving to a new school, courses split into two or more semesters with unique class identifiers, etc. To account for this in the models, we weighted each student record inversely with the number of times the student appeared in the cohort.

Model Specifications

We built four different multi-level value-added models using lmer in R (Bates et al., 2015). Each model has a similar baseline with variable specifications to address different causal assumptions with regard to student learning predictors. We specify each model below to show the variation. All four models include the outcome variable of S_{iY}, which is the test score of student i in year Y. The model estimates this from $S_{\mathrm{i,Y-1}}$, which is the student's pre-test score from the previous year, as well as this score in a quadratic polynomial. It is common to include prior student test scores to quadratic or cubic order to account for potential bias for teachers with scores at the high and low tails of score ranges (Backes et al., 2018; Marder et al., 2020). Each model includes a random intercept for teacher *j* ($T_{j[i]}$) and classroom *n* ($Class_{n[[i]]}$), where we nest students within teacher and class. We included student level demographics for household economic status, race/ethnicity, gifted status, special education status, and emergent bilingual status in each model as coefficients $\Sigma_X X_{g[i]}$, where *g* is the group affiliation for each student. We specify the variable of interest, certification pathway or EPP type with $StdCert_{m[j[i]]}$ for teacher *j* from program *m* as a fixed effect. The second level of each multi-level model specifies the random intercepts for teacher ($T_j \sim N(\mu_T; \sigma_T^2)$) and class ($Class_n \sim N(\mu_L; \sigma_L^2)$) for each model (Gelman & Hill, 2007, Chapter 12.5). We specify each model below (Equations 1 – 4), with an explanation of the causal assumptions that correspond to the model structure. Figure S3 illustrates these causal assumptions further in a Directed Acyclic Graph (DAG).

Model 1 ("No Campus Intercept") is the baseline model, which includes only the covariates described above that are included in all models and does not include a campus intercept. This model assumes that high quality teachers are the primary cause of higher performing students, rather than high-quality campuses playing a bigger role in higher student performance as Models 2 and 3 assume.

(M1 – No Campus Intercept)

$$S_i = \sum_{Y=2012}^{2019} \sum_{\beta=1}^{2} \lambda_{\beta Y} S_{i,Y-1}^{\beta} + T_{j[i]} + Class_{n[[i]]} + StdCert_{m[j[i]]} + \Sigma_X X_{g[i]} + \epsilon_i$$

$$T_j \sim N(\mu_T; \sigma_T^2);\ Class_n \sim N(\mu_L; \sigma_L^2)$$

In addition to the baseline described above, the Model 2 ("Random Campus Intercept") also nests student *i* within campus *k* ($C_{k[i]}$). This model assumes a causal link between campus quality and student learning, rather than linking student scores directly to the teacher, which supposes that high quality campuses improve student scores separately or in addition to teacher quality.

(M2 – Random Campus Intercept)

$$S_i = \sum_{Y=2012}^{2019} \sum_{\beta=1}^{2} \lambda_{\beta Y} S_{i,Y-1}^{\beta} + T_{j[i]} + Class_{n[[i]]} + C_{k[i]} + StdCert_{m[j[i]]} + \Sigma_X X_{g[i]} + \epsilon_i$$

$$T_j \sim N(\mu_T; \sigma_T^2);\ Class_n \sim N(\mu_L; \sigma_L^2);\ C_k \sim N(\mu_C; \sigma_C^2)$$

Model 3 ("Random Campus Intercept & Teacher Experience") assumes the same link between campus and student learning as Model 2 by nesting students and teachers within campuses. This model also controls for years of teaching experience. The years of experience variable has bins of 0-3 years, 4-9 years, and 10 or more years. This model assumes that higher student learning gains are linked to increased teaching experience (e.g., Bastian et al., 2018). However, this causality is confounded by evidence that teachers from a Traditional pathway remain in teaching for longer than teachers from Alternative programs, which is why we do not control for it separately from certification pathway in other models (Marder et al., 2020). In Models 1, 2, and 4 the certification pathway is associated with years of teaching experience so it is not an independent causal factor, compared to Model 3 where "Years of Experience" is expected to have a separate and independent association to "Teacher Quality."

(M3 – Random Campus Intercept & Teacher Experience)

$$S_i = \sum_{Y=2012}^{2019} \sum_{\beta=1}^{2} \lambda_{\beta Y} S_{i,Y-1}^{\beta} + T_{j[i]} + E_{j[i]} + C_{k[i]} + Class_{n[[i]]} + StdCert_{m[j[i]]} +$$

$$\Sigma_X X_{g[i]} + \epsilon_i$$

$$T_j \sim N(\mu_T; \sigma_T^2);\; C_k \sim N(\mu_C; \sigma_C^2);\; Class_n \sim N(\mu_L; \sigma_L^2)$$

Model 4 ("Average Classroom Demographics, No Campus Intercept") also has no campus intercept but adds in classroom averages of student demographics as covariates. This model assumes a direct link between classroom demographics and student learning beyond nesting students within classrooms in the model.

(M4 – Average Classroom Demographics, No Campus Intercept)

$$S_i = \sum_{Y=2012}^{2019} \sum_{\beta=1}^{2} \lambda_{\beta Y} S_{i,Y-1}^{\beta} + T_{j[i]} + Class_{n[[i]]} + StdCert_{m[j[i]]} + \Sigma_X X_{g[i]} + \Sigma x \gamma_x \overline{X}_{n[i]} + \epsilon_i$$

$$T_j \sim N(\mu_T; \sigma_T^2);\; Class_n \sim N(\mu_L; \sigma_L^2)$$

We provide estimates from these models in the next section in Table 1. In addition, we repeated the analyses by using only a subset of students for each of the following subgroups: students from low-income households, students with status of Gifted, students receiving Special Education services, or students of Emergent Bilingual status, as well as Black students, Hispanic students, and White students. These models allowed us to find the effects of teacher certification pathways on specific groups of students. Due to the very large number of models involved, we only used the specifications from Model 2 for this portion of the analysis. We provide these estimates in the next section in Table 2.

Results

We created model outputs for each Grade/Test for the four models based on Traditional or Alternative certification pathways, as well as For-Profit or Not For-Profit EPPs. Additionally, we ran models for these groups across student demographic subgroups using the Random Campus Intercept model, which is explained in the next section. We provide full results with all coefficients for each of these models in the Supplementary Document (Model Outputs) for this paper.

Full Cohort Models

Table 1 describes the estimate from each model and Grade/Test for certification pathway or EPP type, which is the variable of interest in the research questions of this study. The estimates are significant in all 18 Grade/Tests for Models 1, 2, and 3. Model 4, where we added classroom demographic averages as covariates, has less significant results. In all models for all cohorts, each estimate is positive for teachers from Traditional certification pathways compared to those that were prepared through an Alternative pathway, meaning student scores for teachers from Traditional pathways increased for every grade and in every subject area. The estimates from the VAM analyzing teachers prepared by Not For-Profit EPPs were even larger, meaning student scores for teachers from Not For-Profit EPPs increased even more for every grade and every subject area.

Columns 2 to 5 of Table 1 shows the estimates of additional student learning for teachers from Traditional certification pathways. Although all estimates are positive for Traditional certified teachers, there is a range in the estimates across Grade/Tests as well as across models. The smallest increases in student learning for teachers from Traditional pathways were in the elementary and middle grades for Reading (0.008-0.022 *SD* higher) as well as Grade 4 Math (0.018-0.024 *SD* higher). Estimates were typically higher in Math than in Reading subject areas when comparing the same grade level. For example, Grade 8 Reading had an increase of 0.016-0.020 standard deviation units for Traditional certified teachers, while Grade 8 Math had an increase of 0.029-0.044 standard deviation units. This pattern continued when looking at the End-of-Course tests for High School students, where Algebra I had the highest additional learning for teachers from Traditional pathways (0.041-0.055 *SD*). Grade/Tests that had the highest estimates for additional months of learning for teachers from Traditional pathways were English I (0.018-0.047 *SD*) and Grade 8 Social Studies (0.035-0.045 *SD*). In Model 1 ("No Campus Intercept"), where the assumption is that the teacher is primarily responsible for student learning, gains of learning were robust across grades and subject areas (0.017-0.055 *SD*).

Table 1

Value-added estimates from Models M1-M4 for certification variable for teachers from Traditional Certification Path compared with Alternative and Not For-Profit paths compared with For-Profit

Test	M1 ***Traditional−Alt***	M2 ***Traditional−Alt***	M3 ***Traditional−Alt***	M4 ***Traditional−Alt***	M1 ***NotForProfit−ForProfit***	M2 ***NotForProfit−ForProfit***	M3 ***NotForProfit−ForProfit***	M4 ***NotForProfit−ForProfit***
Math-4	0.024 (***)	0.018 (***)	0.021 (***)	0.018 (***)	0.037 (***)	0.030 (***)	0.024 (***)	0.029 (***)
Math-5	0.039 (***)	0.022 (***)	0.024 (***)	0.030 (***)	0.050 (***)	0.033 (***)	0.030 (***)	0.042 (***)
Math-6	0.044 (***)	0.035 (***)	0.037 (***)	0.016 (*)	0.046 (***)	0.037 (***)	0.034 (***)	0.023 (***)
Math-7	0.042 (***)	0.036 (***)	0.035 (***)	0.022 (***)	0.052 (***)	0.046 (***)	0.048 (***)	0.036 (***)
Math-8	0.043 (***)	0.043 (***)	0.044 (***)	0.029 (***)	0.053 (***)	0.047 (***)	0.041 (***)	0.042 (***)
Algebra I-9	0.055 (***)	0.041 (***)	0.042 (***)	0.047 (***)	0.065 (***)	0.049 (***)	0.049 (***)	0.057 (***)
Read-4	0.022 (***)	0.012 (***)	0.014 (***)	0.010 (***)	0.028 (***)	0.018 (***)	0.012 (***)	0.018 (***)
Read-5	0.018 (***)	0.008 (***)	0.009 (***)	0.008 (**)	0.017 (***)	0.008 (**)	0.007 (**)	0.010 (***)
Read-6	0.022 (***)	0.011 (***)	0.011 (***)	−0.001	0.019 (***)	0.006 (*)	0.007 (*)	0.002
Read-7	0.017 (***)	0.010 (***)	0.010 (***)	−0.002	0.017 (***)	0.012 (***)	0.012 (***)	0.004
Read-8	0.020 (***)	0.016 (***)	0.016 (***)	−0.002	0.022 (***)	0.018 (***)	0.018 (***)	0.004
English I-9	0.047 (***)	0.035 (***)	0.035 (***)	0.018 (***)	0.055 (***)	0.035 (***)	0.034 (***)	0.030 (***)
English II-10	0.018 (***)	0.012 (***)	0.010 (**)	0.002	0.026 (***)	0.016 (***)	0.015 (***)	0.011 (***)
Science-5	0.036 (***)	0.013 (**)	0.015 (**)	0.026 (***)	0.039 (***)	0.021 (***)	0.019 (***)	0.031 (***)
Science-8	0.020 (**)	0.015 (**)	0.016 (**)	0.002	0.033 (***)	0.026 (***)	0.021 (***)	0.019 (**)
Biology-9	0.031 (***)	0.013 (*)	0.012 (*)	0.013	0.048 (***)	0.035 (***)	0.038 (***)	0.034 (***)
Social Studies-8	0.045 (***)	0.035 (***)	0.036 (***)	0.012	0.056 (***)	0.046 (***)	0.044 (***)	0.034 (***)
US History-11	0.031 (***)	0.027 (***)	0.027 (***)	0.025 (***)	0.029 (***)	0.028 (***)	0.029 (***)	0.024 (**)

Note: M1=No Campus Intercept, M2=Random Campus Intercept, M3=Random Campus Intercept & Teacher Experience, M4= Avg. Classroom Demographics & No Campus Intercept. Significance – * $|t| > 1.96$, ** $|t| > 2.58$, *** $|t| > 3.29$

Columns 6 to 9 of Table 1 show the estimates of additional student learning for teachers from Not For-Profit EPPs. The estimates for these groups follow a similar pattern to those described above when analyzing certification pathway. However, estimates for additional student learning for teachers from Not For-Profit EPPs were typically higher than for Traditional certification pathways. The smallest increases in student learning for teachers from Not For-Profit EPPs were in the elementary and middle grades for Reading (0.006-0.028 *SD* higher). Estimates were typically higher in Math than in Reading subject areas when comparing the same grade level. For example, Grade 8 Reading had an increase of 0.018-0.022 standard deviation units for teachers from Not For-Profit EPPs, while Grade 8 Math had an increase of 0.041-0.053 standard deviation units. This pattern continued when looking at the End-of-Course tests for High School students, where Algebra I had the highest additional learning for teachers from Not For-Profit EPPs (0.049-0.065 *SD*). The other Grade 9 EOC tests also had some of the highest learning gains, with English I at 0.030-0.055 higher and Biology at 0.034-0.048 higher. Grade 8 Social Studies also had some of the biggest gains with an increase of 0.034-0.056 standard deviation units for teachers from Not For-Profit EPPs. Similar to the other analyses, Model 1 ("No Campus Intercept") typically had the most robust gains of student learning across grades and subject areas (0.017-0.065 *SD*).

Demographic Subgroup Models

Table 2 shows the estimates for each subgroup in the student demographics category for teachers from a Traditional and Postbacc versus Alternative certification pathways and for teachers from Not For-Profit EPPs versus For-Profit EPPs. We split students into subgroups by demographic categories: students from low-income households, gifted students, special education students, emergent bilingual students, Black students, Hispanic students, and White students. This means we filtered the full cohort of students to only students in that subgroup and then estimated expected learning of that subgroup of students using Model 2 as described in the previous section.

Table 2

Student Subgroup Estimates for M2-Random Campus Intercept Model (SD)

Compare Traditional and Post-Bacc Teachers to Alternative Certified								
Test	Full	Eco	Gift	Sped	EmBi	Bl	Hi	Wh
Math								
Math-4	0.018 (***)	0.017 (***)	0.012 (*)	0.035 (***)	0.019 (**)	0.020 (**)	0.016 (***)	0.023 (***)
Math-5	0.022 (***)	0.019 (***)	0.006	0.042 (***)	0.021 (***)	0.026 (***)	0.017 (***)	0.032 (***)
Math-6	0.035 (***)	0.032 (***)	0.025 (**)	0.033 (***)	0.033 (***)	0.033 (***)	0.031 (***)	0.034 (***)
Math-7	0.036 (***)	0.037 (***)	0.017	0.048 (***)	0.029 (***)	0.038 (***)	0.033 (***)	0.040 (***)
Math-8	0.043 (***)	0.039 (***)	0.031 (**)	0.057 (***)	0.040 (***)	0.038 (***)	0.038 (***)	0.039 (***)
Algebra I-9	0.041 (***)	0.041 (***)	0.043 (***)	0.044 (***)	0.033 (***)	0.032 (***)	0.044 (***)	0.029 (***)
Reading								
Read-4	0.012 (***)	0.011 (***)	0.011 (**)	0.024 (***)	0.015 (**)	0.008	0.012 (***)	0.005
Read-5	0.008 (***)	0.007 (**)	0.003	0.025 (***)	0.015 (***)	0.007	0.008 (**)	0.008 (*)
Read-6	0.011 (***)	0.011 (***)	0.016 (***)	0.030 (***)	0.003	0.010	0.010 (**)	0.016 (***)
Read-7	0.010 (***)	0.007 (*)	0.010 (**)	0.018 (**)	0.008	0.003	0.010 (**)	0.011 (**)
Read-8	0.016 (***)	0.011 (***)	0.004	0.026 (***)	0.006	0.022 (***)	0.009 (**)	0.010 (*)
English I-9	0.035 (***)	0.028 (***)	0.030 (***)	0.033 (***)	0.010	0.034 (***)	0.029 (***)	0.036 (***)
English II-10	0.012 (***)	0.009 (**)	0.008	0.014 (*)	0.001	0.020 (***)	0.006	0.015 (***)
Science								
Science-5	0.013 (**)	0.014 (**)	0.009	0.036 (***)	0.017 (*)	0.029 (***)	0.011 (*)	0.007
Science-8	0.015 (**)	0.015 (*)	0.014	0.015	0.018 (*)	0.020 (*)	0.013 (*)	0.018 (*)
Biology-9	0.013 (*)	0.010	0.004	0.023 (**)	0.014	0.020 (*)	0.010	0.012
Social Studies								
Soc Studies-8	0.035 (***)	0.032 (***)	0.030 (**)	0.034 (***)	0.019 (*)	0.044 (***)	0.026 (***)	0.044 (***)
History-11	0.027 (***)	0.025 (***)	0.038 (***)	0.024 (*)	0.018	0.022 (*)	0.024 (***)	0.018 (*)
Compare Not-For-Profit to For-Profit Teachers								
Math								
Math-4	0.030 (***)	0.032 (***)	0.018 (**)	0.038 (***)	0.030 (***)	0.035 (***)	0.027 (***)	0.029 (***)
Math-5	0.033 (***)	0.030 (***)	0.025 (***)	0.044 (***)	0.030 (***)	0.038 (***)	0.029 (***)	0.042 (***)
Math-6	0.037 (***)	0.034 (***)	0.038 (***)	0.033 (***)	0.035 (***)	0.025 (***)	0.036 (***)	0.040 (***)
Math-7	0.046 (***)	0.042 (***)	0.024	0.047 (***)	0.028 (***)	0.042 (***)	0.039 (***)	0.059 (***)
Math-8	0.047 (***)	0.047 (***)	0.033 (**)	0.066 (***)	0.045 (***)	0.045 (***)	0.041 (***)	0.040 (***)
Algebra I-9	0.049 (***)	0.050 (***)	0.061 (***)	0.048 (***)	0.036 (***)	0.049 (***)	0.052 (***)	0.035 (***)
Reading								
Read-4	0.018 (***)	0.018 (***)	0.008	0.025 (***)	0.015 (**)	0.027 (***)	0.018 (***)	0.017 (***)

Test	Full	Eco	Gift	Sped	EmBi	Bl	Hi	Wh
Read-5	0.008 (**)	0.007 (*)	0.004	0.015 (*)	0.011 (*)	0.012 (*)	0.009 (**)	0.006
Read-6	0.006 (*)	0.009 (**)	0.010 (*)	0.022 (***)	0.001	0.005	0.008 (*)	0.015 (***)
Read-7	0.012 (***)	0.011 (***)	0.010 (**)	0.017 (**)	0.006	0.015 (***)	0.011 (***)	0.012 (**)
Read-8	0.018 (***)	0.011 (***)	0.003	0.021 (***)	0.004	0.015 (**)	0.012 (***)	0.013 (***)
English I-9	0.035 (***)	0.030 (***)	0.037 (***)	0.028 (***)	0.011	0.033 (***)	0.032 (***)	0.040 (***)
English II-10	0.016 (***)	0.015 (***)	0.022 (***)	0.016 (*)	0.004	0.023 (***)	0.011 (**)	0.025 (***)
Science								
Science-5	0.021 (***)	0.021 (***)	0.012 (*)	0.032 (***)	0.017 (*)	0.034 (***)	0.019 (***)	0.018 (**)
Science-8	0.026 (***)	0.026 (***)	0.027 (***)	0.027 (**)	0.030 (***)	0.027 (***)	0.024 (***)	0.027 (***)
Biology-9	0.035 (***)	0.036 (***)	0.027 (***)	0.034 (***)	0.034 (***)	0.036 (***)	0.033 (***)	0.031 (***)
Social Studies								
Soc Studies-8	0.046 (***)	0.044 (***)	0.049 (***)	0.044 (***)	0.034 (***)	0.052 (***)	0.037 (***)	0.057 (***)
History-11	0.028 (***)	0.028 (***)	0.026 (*)	0.034 (**)	0.029 (*)	0.025 (**)	0.026 (***)	0.025 (**)

Note: **Eco=Economically Disadvantaged, Gift=Gifted, SpEd=Special Education, EmBi=Emergent Bilingual, Bl=Black, Hi=Hispanic, Wh=White. Significance – * $|t| > 1.96$, ** $|t| > 2.58$, *** $|t| > 3.29$**

Overall, all subgroups of students benefit from having teachers from Traditional certification pathways and Not For-Profit EPPs. Table 2 demonstrates that no subgroup has a negative estimate for Traditional and Not For-Profit certifications. For example, if one looks at the column with estimates for only students from low-income households, they range from 0.007 to 0.041 standard deviations for students from low-income households when they have teachers from Traditional pathways. When students from low-income households have teachers from Not For-Profit EPPs they gain an additional 0.007 to 0.050 standard deviations. These gains are repeated to varying degrees across each subgroup shown in columns 3-9 in Table 2.

Results from these subgroups were largely consistent with results when compared to the full cohort of students in terms of standard deviation. The differences can be seen in Table 2 with the estimates from "Full Cohort" compared to estimates from each subgroup. Estimates changed for some grade/tests in significance or size of estimate based on the subgroup compared to the full cohort. For example, in math, students' learning for the full cohort increased by 0.018-0.043 standard deviations, while special education students increased by 0.033-0.057 when they had teachers from Traditional pathways. With the exception of Grade 6 Math, students that receive special education services gained more learning from teachers from a Traditional pathway than the full cohort of students. This is not the case for gifted students in math, whose gains from Traditional teachers, while still positive, were smaller than those of the full cohort. For other demographic subgroups our summary is that their improved learning from Traditional teachers was closest to that of the full cohort in math. Similar patterns were seen in math for teachers from Not For-Profit EPPs.

Discussion and Conclusion

Contesting a Narrative

The results from this study are clear and consistent. For every grade, for every test, for every group of students and for every set of model assumptions, there is no case where students learn more from Alternatively prepared teachers than they do from teachers from a Traditional pathway. Similarly, there is no case where students learn more from teachers from For-Profit pathways than they do from teachers who went through Not For-Profit programs. In most cases, students whose teachers came from Traditional or Not For-Profit programs learn significantly more.

Our findings contest a narrative that dominates the public debate: that traditionally prepared teachers are of low quality and do not improve student learning. This narrative has gained strength after the pandemic, with claims such as "Emergency-Hired Teachers Do Just as Well as Those Who

Go Through Normal Training" (Aldeman, 2024), based on VAMs using recent data from New Jersey (Backes & Goldhaber, 2024). Perhaps the lack of attention to the explosive growth of uncertified teachers in Figure 2 comes from the belief that it really does not matter, for the uncertified "do just as well" as those who prepared.

It is true that many studies have concluded that value-added models find no differences between educator preparation programs (Aaronson et al., 2007; Harris & Sass, 2011; Staiger & Rockoff, 2010; von Hippel & Bellows, 2018). Our response is that the system is naturally noisy, the high-stakes tests were not designed to optimize the problem of distinguishing between teachers, and large sample sizes are needed to reject the null hypothesis. With the advantage of many more years of data than previous studies, and by grouping together EPPs of similar type, we arrive at significant conclusions. Independent of grade or subject or model or subgroup, teachers from Traditional and Not For-Profit EPPs have higher teacher quality as measured by value-added models of student test scores.

Scale of Results

Estimates across all grades and tests in Math ranged from 0.016 to 0.055 additional standard deviations for teachers from Traditional pathways and from 0.023 to 0.065 additional standard deviations for teachers from Not For-Profit EPPs. Estimates in Science are not as large, with ranges of 0.012 to 0.036 for Traditional pathways and 0.019 to 0.048 for Not For-Profit EPPs. Most but not all of these estimates are significant. These estimates confirm results from other value-added studies on high school STEM teachers (Backes et al., 2018; Boyd et al., 2009; Marder et al., 2020). The results from this study demonstrate that these learning gains begin, especially in math, in elementary grades. Estimates in English Language Arts were also not as large as Math estimates and in many cases not as large as Science estimates. The exception to this was English I in Grade 9, where the estimates ranged from 0.018 to 0.047 in standard deviations units for students with teachers from Traditional pathways and from 0.030 to 0.055 for students from Not For-Profit EPPs. Although teacher shortages are particularly acute in STEM, we find that students benefit when their teachers come from Traditional and Not For-Profit pathways across all grade levels and disciplines.

Effect Sizes and Insensitivity to Instruction

These effect sizes are small, but this is not unexpected. Kraft (2020) notes that 'larger studies with broad achievement measures have systematically smaller effect sizes" (p. 247). Kraft's Figure 1 shows that for Grades 4 and above the mean effect size for hundreds of interventions in both math and reading is around .05. The small effect size here is partly due to the way the high-stakes tests are constructed. Students can get around 20% on the exams just by guessing. Even if the effect within one year is small, one must ask if the effects are cumulative. If they are, then across 10 years the effect size for students with only Traditional teachers versus students with only Alternatively certified teachers would be of order 0.5, which is large.

The question of whether teacher effects cumulate is unsettled. At one extreme are Sanders & Rivers (1996) who concluded that "teacher effects are both additive and cumulative…" (p. 6). At the other is Rothstein (2010) who resolved that "conventional measures of individual teachers' value-added fade out very quickly and are at best weakly related to long-run effects" (p. 175). While we cannot resolve this matter here, we note that Marder et al. (2020) showed the effect of having a Traditional teacher is one-half to one-third the effect of eligibility for free and reduced lunch, and that the effects of free and reduced lunch eligibility cumulate for around three years and then saturate. Since growing up in poverty is widely accepted as providing a powerful negative impact on education (R. Rothstein, 2004), the small effect sizes may mainly reflect insensitivity of the tests to instruction (Popham, 2007)

Effects on Subgroups

Some of the most acute teacher shortages of all are in Special Education (Day et al., 2023). Therefore, it is notable that when we examine the effects of teachers from different pathways on student subgroups, the single subgroup that benefits the most in our analysis when they have teachers from Traditional or Not For-Profit pathways is students flagged for Special Education.

In the cohorts for this study, teachers from Alternative and For-Profit certification pathways have more Emergent Bilingual students, more students from low-income households, and larger percentages of Black and Hispanic students compared to teachers from Traditional and Not For-Profit certification pathways. Beyond this study, significantly higher percentages of these historically marginalized students are in classes taught by teachers not trained in traditional teacher certification programs (Dee & Goldhaber, 2017). This means these subgroups would potentially be more affected by the long-term impacts on student learning from having teachers from these pathways. We found additional support for this claim by looking in detail at the learning gains of the different student subgroups.

Mechanism for Effects

A critique of this study and similar studies is that they do not provide insight into how and why different preparation pathways raise student scores, just that they do (Goldhaber & Ronfeldt, 2020). In order to inform policy and policymakers, this point needs to be addressed. This study was designed to define Traditional certification so that it corresponds with a particular curricular choice of preparation programs—the inclusion of student teaching. Furthermore, most For-Profit preparation programs are web-based and asynchronous. Thus, our results support policies that steer prospective teachers towards university-based programs with student teaching, and away from asynchronous web-based programs that lack student teaching.

We acknowledge the challenging tension between teacher quantity and quality that schools face when they hire to fill vacancies. The fact that there are more Alternative and For-Profit teachers teaching Black, Hispanic, and economically disadvantaged students confirms that policies that allow for alternative pathways help place teachers in areas of need. However, our results also indicate that these students would be learning more if their teachers were certified through Traditional pathways.

Implications

Alternative certification is far too important in the Texas teacher preparation ecosystem to think of eliminating it. Nevertheless, our work has clear policy implications. For every grade level, every subject, and every group of students, Traditional teacher preparation *as it is* benefits students. Therefore, every policy that affects teacher preparation should be analyzed in light of simple questions. Does the policy make it more or less time-consuming for candidates to obtain Traditional teaching certification? Does the policy make it more or less expensive for candidates to obtain Traditional teaching certification? Does the policy make it more or less onerous for IHE's to offer teacher preparation programs? Does the policy make it more or less onerous for schools to provide support for new teachers? Viewed in this way, a policy decision to provide financial support to candidates during the student teaching semester would be favorable. A policy decision to increase the number of required field hours for Traditional pathways would not.

As illustrated in Figure 1, we cannot take for granted that even the current unsatisfactory number of well-prepared teachers will be available in the future. Decisions within Texas and across the nation to address teacher shortages by opening up new alternative preparation pathways should be made with an understanding that there is a cost to forgoing tradition, and the neediest students will pay it.

Acknowledgements

We thank Charles Martinez, John Fitzpatrick, Pedro Reyes, Celeste Alexander, Laura Torres, Ben David, Jill Marshall, and other participants in the Texas Educator Preparation Pathways Study for many thoughtful comments on early versions of this work.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago Public High Schools. *Journal of Labor Economics*, *25*(1), 95–135. https://doi.org/10.1086/508733

Aldeman, C. (2024, January 10). *Emergency-hired teachers do just as well as those who go through normal training*. https://www.the74million.org/article/emergency-hired-teachers-do-just-as-well-as-those-who-go-through-normal-training/

American Association for Employment in Education. (2023). *Educator supply and demand report 2022-2023*. https://issuu.com/home/read/78puxujzq4i

Amrein-Beardsley, A. (2009). Value-added tests: Buyer, be aware. *Educational Leadership*, *67*(3), 38–42. Professional Development Collection.

Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation*. *The Quarterly Journal of Economics*, *132*(2), 871–919. https://doi.org/10.1093/qje/qjx001

Backes, B., & Goldhaber, D. (2024). *The relationship between pandemic-era teacher licensure waivers and teacher demographics, retention, and effectiveness in New Jersey* [Working Paper No. 286-0623]. CALDER. https://caldercenter.org/sites/default/files/CALDER%20WP%20286-0623.pdf

Backes, B., Goldhaber, D., Cade, W., Sullivan, K., & Dodson, M. (2018). Can UTeach? Assessing the relative effectiveness of STEM teachers. *Economics of Education Review*, *64*, 184–198. https://doi.org/10.1016/j.econedurev.2018.05.002

Bastian, K. C., Patterson, K. M., & Pan, Y. (2018). Evaluating teacher preparation programs with teacher evaluation ratings: Implications for program accountability and improvement. *Journal of Teacher Education*, *69*(5), 429–447. https://doi.org/10.1177/0022487117718182

Boyd, D., Goldhaber, D. D., Lankford, Hamilton., & Wyckoff, J. Humphrey. (2007). The effect of certification and preparation on teacher quality. *The Future of Children*, *17*(1), 45–68. https://doi.org/10.1353/foc.2007.0000

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 416–440. https://doi.org/10.3102/0162373709353129

Bruno, P. (2023). *Pandemic-era school staff shortages: Evidence from unfilled position data in Illinois* (SSRN Scholarly Paper 4306263). https://doi.org/10.2139/ssrn.4306263

Carinci, J. E., Meyer, S. J., & Jackson, C. (2020). *Linking teacher preparation program design and implementation to outcomes for teachers and students*. Information Age Publishing.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Cochran-Smith, M., & Villegas, A. M. (2015). Framing teacher preparation research: An overview of the field, Part 1. *Journal of Teacher Education*, *66*(1), 7–20. https://doi.org/10.1177/0022487114549072

Cochran-Smith, M., Villegas, A. M., Abrams, L., Chavez-Moreno, L., Mills, T., & Stern, R. (2015). Critiquing teacher preparation research: An overview of the field, Part II. *Journal of Teacher Education*, *66*(2), 109–121. https://doi.org/10.1177/0022487114558268

Corcoran, T. (2007). *Teaching matters: How state and local policymakers can improve the quality of teachers and teaching* [Policy Brief]. CPRE. https://doi.org/10.12698/cpre.2007.rb48

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8–15. https://doi.org/10.1177/003172171209300603

Day, J., Nagro, S. A., & Mason-Williams, L. (2023). The nationwide trends and preparation requirements of alternative route programs in special education. *Teacher Education and Special Education*, 08884064231172525. https://doi.org/10.1177/08884064231172525

Dee, T. S., & Goldhaber, D. (2017). *Understanding and addressing teacher shortages in the United States.* [Policy Brief 2017-05]. The Hamilton Project. https://www.hamiltonproject.org/assets/files/understanding_and_addressing_teacher_shortages_in_us_pb.pdf.

Edwards, D. S., Kraft, M. A., Christian, A., & Candelaria, C. A. (2022). *Teacher shortages: A unifying framework for understanding and predicting vacancies* [EdWorkingPaper No. 22-684]. Annenberg Institute for School Reform at Brown University. https://eric.ed.gov/?id=ED625903

Etheredge, D. K. (2015). *Alternative certification teaching programs in Texas: A historical analysis* [Doctoral Dissertation]. University of North Texas. https://digital.library.unt.edu/ark:/67531/metadc799511/

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). Identifying effective teachers using performance on the job. *Policy Brief Series (Hamilton Project)*, *1*, 1–7.

Grossman, P., & Loeb, S. (2008). *Alternative routes to teaching: Mapping the new landscape of teacher education.* Harvard Education Press.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466–479. https://doi.org/10.1016/j.econedurev.2010.12.006

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7), 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Henry, G. T., Bastian, K. C., Fortner, C. K., Kershaw, D. C., Purtell, K. M., Thompson, C. L., & Zulli, R. A. (2014). Teacher preparation policies and their effects on student achievement. *Education Finance and Policy*, *9*(3), 264–303.

Hess, F. M. (2002). Tear down this wall: The case for a radical overhaul of teacher certification. *Educational Horizons, 80*, 169–183.

Iteach. (n.d.). *Alternative Teacher Certification Texas - Iteach.* https://web.archive.org/web/20240426123739/https://iteach.net/alternative-teacher-certification/texas-alternative-teacher-certification/

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* [Research Paper]. MET Project, Bill & Melinda Gates Foundation. https://eric.ed.gov/?id=ED540960

Kirksey, J. J., & Gottlieb, J. J. (2023). *Teacher preparation in the wild west: The impact of fully-online teacher preparation and uncertified teachers in Texas* [Policy Brief No. 4]. The Center for Innovative Research in Change, Leadership, and Education (CIRCLE), Texas Tech University. https://hdl.handle.net/2346/96067

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, *47*, 180–195. https://doi.org/10.1016/j.econedurev.2015.01.006

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs? *Education Finance and Policy*, *10*(4), 508–534.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Maier, A. (2012). Doing good and doing well: Credentialism and Teach for America. *Journal of Teacher Education*, *63*(1), 10–22. https://doi.org/10.1177/0022487111422071

Marder, M., David, B., & Hamrock, C. (2020). Math and science outcomes for students of teachers from standard and alternative pathways in Texas. *Education Policy Analysis Archives*, *28*(27). https://doi.org/10.14507/epaa.28.4863

Marder, M., Reyes, P., Marshall, J., C., A., Maloch, B., & Martinez, C. R. (2022). *Texas Educator Preparation Pathways Study*. https://issuu.com/texaseducation/docs/texas_educator_prep_pathways_study_issuu

Natanson, H. (2022, August 10). 'Never seen it this bad': America faces catastrophic teacher shortage. *Washington Post*. https://www.washingtonpost.com/education/2022/08/03/school-teacher-shortage/

Nguyen, T., Lam, C., & Bruno, P. (2022). *Is there a national teacher shortage? A systematic examination of reports of teacher shortages in the United States*. https://doi.org/10.26300/76eq-hj32

Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community*, *4*(4), 286–308. https://doi.org/10.1108/JPCC-12-2018-0032

Popham, W. J. (2007). Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan*, *89*(2), 146–155. https://doi.org/10.1177/003172170708900211

Redding, C. (2022). Changing the composition of beginning teachers: The role of state alternative certification policies. *Educational Policy*, *36*(7), 1791–1820. https://doi.org/10.1177/08959048211015612

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, *125*(1), 175–214. https://doi.org/10.1162/qjec.2010.125.1.175

Rothstein, R. (2004). *Class and schools*. Economic Policy Institute.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. University of Tennessee Value-Added Research and Assessment Center.

Schmidt, W. H., Burroughs, N. A., Houang, R. T., & Cogan, L. S. (2020). The role of content knowledge in mathematics teacher preparation: A study of traditional and alternative teacher preparation in Texas. *Journal of Teacher Education*, *71*(2), 233–246. https://doi.org/10.1177/0022487118805989

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives*, *24*(3), 97–117.

Suell, J. L., & Piotrowski, C. (2007). Alternative teacher education programs: A review of the literature and outcome studies. *Journal of Instructional Psychology*, *34*(1), 54–58.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2019). Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Education Policy Analysis Archives*, *27*(35). https://eric.ed.gov/?id=EJ1213618

Texas Teachers of Tomorrow. (n.d.). *Earn your Texas teaching certificate*. https://www.teachersoftomorrow.org/texas/

Thompson, D. (2022, August 24). There is no national teacher shortage. *The Atlantic*. https://www.theatlantic.com/newsletters/archive/2022/08/national-teacher-shortage-turnover-student-enrollment/671214/

US Department of Education. (2022). *Title II data collection* [Dataset]. https://title2.ed.gov/Public/DataTools/Tables.aspx
von Hippel, P. T., & Bellows, L. (2018). How much does teacher quality vary across teacher preparation programs? Reanalyses from six states. *Economics of Education Review*, *64*, 298–312. https://doi.org/10.1016/j.econedurev.2018.01.005
von Hippel, P. T., Bellows, L., Osborne, C., Lincove, J. A., & Mills, N. (2016). Teacher quality differences between teacher preparation programs: How big? How reliable? Which programs are different? *Economics of Education Review*, *53*, 31–45. https://doi.org/10.1016/j.econedurev.2016.05.002
Walsh, K., & Jacobs, S. (2007). *Alternative certification isn't alternative.* https://www.nctq.org/nctq/images/Alternative_Certification_Isnt_Alternative.pdf
Ward, M. (2023, November 3). *The teacher shortage problem is bad. Really bad.* POLITICO. https://www.politico.com/newsletters/politico-nightly/2022/08/15/the-teacher-shortage-problem-is-bad-really-bad-00052053
Will, M. (2022, June 8). Alternative certification programs are booming. But candidates aren't finishing. *Education Week*. https://www.edweek.org/teaching-learning/alternative-certification-programs-are-booming-but-candidates-arent-finishing/2022/06

About the Authors

Annelies Rhodes
The University of Texas at Austin, and E3 Alliance
arhodes@e3alliance.org
https://orcid.org/0000-0003-2888-0698
Annelies Rhodes obtained her PhD in STEM Education from the University of Texas at Austin. Her interests lie in the use of educational data to inform policy. She is Senior Director of Research for E3 Alliance.

Michael Marder
The University of Texas at Austin
marder@chaos.utexas.edu
https://orcid.org/0000-0003-0197-4459
Michael Marder is Professor of Physics and Executive Director of UTeach at The University of Texas at Austin. His interest in studying educational data stemmed from over 20 years of work preparing STEM teachers at UT Austin and dozens of other universities.

education policy analysis archives

Volume 32 Number 51 September 10, 2024 ISSN 1068-2341

Teachers College at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A1 (Brazil), SCImago Journal Rank, SCOPUS, SOCOLAR (China).

About the Editorial Team: https://epaa.asu.edu/ojs/index.php/epaa/about/editorialTeam

Please send errata notes to Jeanne M. Powers at jeanne.powers@asu.edu