

education policy analysis archives

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



Arizona State University

Volume 28 Number 27

February 24, 2020

ISSN 1068-2341

Math and Science Outcomes for Students of Teachers from Standard and Alternative Pathways in Texas¹

Michael Marder

Bernard David

University of Texas at Austin



Caitlin Hamrock

E3 Alliance, Austin, Texas

United States

Citation: 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>

Abstract: Texas provides a unique opportunity to examine teachers without standard university preparation, for it prepares more teachers through alternative pathways than any other state. We find two advantages for mathematics and science teachers prepared in the standard way. First, since 2008 they have been staying in the classroom longer than those who pursued alternative routes. Second, we analyze student performance on Algebra 1 and Biology exams over the period 2012-2018. Algebra I students with experienced teachers from standard programs gain .03 to .05 in standard deviation units compared to students whose teachers were alternatively prepared. For Biology students there are fewer statistically significant differences, although when differences exist they almost all favor standard programs. These effects are difficult to measure in part because teachers are not assigned to teach courses with high-stakes exams at random. Nevertheless, we find strong

¹ Funding in partial support of this work from the National Math and Science Initiative and the ExxonMobil Foundation is gratefully acknowledged.

evidence in Algebra I that students learn more when their teachers have standard preparation. In Biology there is also evidence but less compelling. Thus, we recommend that all states bolster traditional university-based teacher certification, that Texas not take drastic action to curtail alternative certification, and that other states not allow it to grow too quickly.

Keywords: teacher preparation; alternative certification; hierarchical linear models; STEM education

Resultados de matemáticas y ciencias para estudiantes de maestros de caminos estándar y alternativos en Texas

Resumen: Texas ofrece una oportunidad única para examinar a los maestros sin una preparación universitaria estándar, ya que prepara a más maestros a través de vías alternativas que cualquier otro estado de los EE. UU. Encontramos dos ventajas para los maestros de matemáticas y ciencias preparados de la manera estándar. Primero, desde 2008 han estado en el aula más tiempo que aquellos que buscaron rutas alternativas. En segundo lugar, analizamos el rendimiento de los alumnos en los exámenes de Álgebra 1 y Biología durante el período 2012-2018. Encontramos evidencia sólida en Álgebra I de que los estudiantes aprenden más cuando sus maestros tienen una preparación estándar. En biología también hay evidencia pero menos convincente. Por lo tanto, recomendamos que todos los estados refuercen la certificación tradicional de maestros basada en la universidad, que Texas no tome medidas drásticas para reducir la certificación alternativa, y que otros estados no permitan que crezca demasiado rápido.

Palabras-clave: preparación docente; certificación alternativa; modelos lineales jerárquicos; Educación STEM

Resultados de matemática e ciências para estudantes de professores de caminhos padrão e alternativos no Texas

Resumo: O Texas oferece uma oportunidade única de examinar professores sem preparação padrão para a universidade, pois prepara mais professores por caminhos alternativos do que qualquer outro estado dos EUA. Encontramos duas vantagens para professores de matemática e ciências preparados da maneira padrão. Primeiro, desde 2008 eles ficam na sala de aula há mais tempo do que aqueles que buscam rotas alternativas. Segundo, analisamos o desempenho dos alunos nos exames de Álgebra 1 e Biologia durante o período 2012-2018. Encontramos na Álgebra I fortes evidências de que os alunos aprendem mais quando seus professores têm uma preparação padrão. Em Biologia, há também evidências, mas menos convincentes. Portanto, recomendamos que todos os estados reforcem a certificação tradicional de professores com base na universidade, que o Texas não tome medidas drásticas para reduzir a certificação alternativa e que outros estados não permitam que ele cresça muito rapidamente.

Palavras-chave: preparação de professores; certificação alternativa; modelos lineares hierárquicos; Educação STEM

Introduction

The United States faces a growing shortage of teachers, with recent estimates of 100,000 more new teachers needed each year than are available (Sutcher, Darling-Hammond, & Carver-Thomas, 2019). Conversations about the complex issue of teacher shortages often turn to the certification pathways that are available to prospective teachers, and, complicating the issue, how pathways that increase the quantity of teachers may not provide the quality of training required for them to be successful. Alternative certification programs grew out of the desire to develop market-based (Hess, 2002) disruptive (Christensen, 1997) competitors to traditional university programs, affording individuals who have already completed a bachelor's degree an expedited route into the

classroom. Given the complementary features and goals of these certification programs, it is clear that one cannot consider the issues of teacher shortages and preparation programs independently. Texas provides a unique opportunity to explore these related issues since it prepares more teachers and has also proceeded further towards obtaining teachers from non-standard pathways than any other state.

The central aim in this study is to inform decisions on whether to allow and how to regulate alternative certification programs leading to rapid certification. Underlying the issue of teacher quality and teacher shortages is a complex system of interconnected costs and benefits. Regulation of any certification pathway should be fully informed by an understanding of the implications for both the number and adequate preparation of teachers. We focus on STEM teachers, examining teacher retention and student test scores in Algebra I and Biology. In this area of teaching, because of severe shortages we document, any effort to increase the supply of teachers must not be dismissed casually. We will show, particularly in Algebra 1, that students learn more in classes of teachers from standard programs, and that since 2008, alternatively certified teachers have not persisted in teaching as long as those from standard programs. Yet the sizes of these effects are not large enough to justify precipitous actions, and thus our conclusions and policy recommendations focus on how alternative certification pathways might fit into the ecosystem of STEM teaching in a way that neither disrupts this complex system, nor students' opportunities to learn.

Some of the most acute shortages of secondary teachers arise in the STEM fields, as indicated by surveys of employers (American Association for Employment in Education, 2016) or by the data in Table 1. Nearly 40% of mathematics teachers either lack full teaching certification or lack a major or minor in mathematics. In the physical sciences, over 60% of teachers lack one or the other of these qualifications. If one uses Advanced Placement Computer Science (CS) exams (College Board, 2016) as the basis for an estimate, less than 20% of U.S. high schools even offer computer science. Such shortages are poised to become greater because the number of teachers prepared in the highest-producing states has been falling (Figure 1).

Table 1
STEM teachers out of field in main assignment or not certified

Subject	Number of Teachers	Percentage with no major in main assignment or not certified
Mathematics	144,800	38%
Science	126,300	27%
Biology	51,900	35%
Physical Science	64,600	62%
Chemistry	24,300	66%
Earth Sciences	12,400	68%
Physics	13,300	63%

Source: Schools and Staffing Survey (2012).

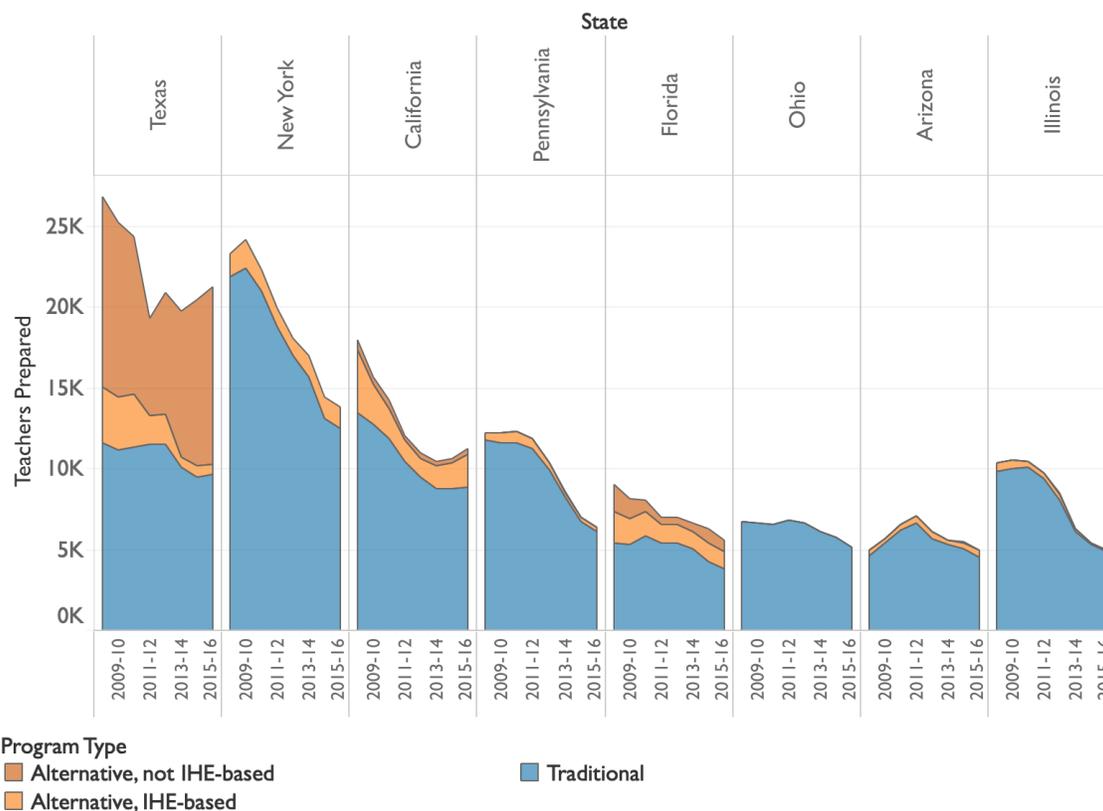


Figure 1. Teachers prepared in the highest-producing states by program type, 2010-2016. IHE refers to Institution of Higher Education
 Data from US Department of Education (2018b, Completers 2008-2016).

The decline in production of teachers is particularly striking in New York and California. Texas, which now produces far more teachers than any other state despite having a smaller population than California, has also seen a drop, but it is less severe overall, and the trend in production has been upward since 2011-2012. The main reason for this is that Texas now obtains most of its teachers from alternative certification programs, and these are dominated by a few companies that specialize in rapid online teacher certification. Thus it is of interest to determine if and in what way preparation pathways impact teacher quality because of the increasing quantity of alternatively certified teachers, and the use of alternative certification programs to ameliorate teacher shortages.

We focus on STEM teachers in particular, for which Figure 2 shows production has been declining as well. The difficulty of recruiting STEM teachers has been a subject of concern at least since the *Gathering Storm* report (Augustine, 2006). The largest effort at the Federal level to increase the number of STEM teachers is NSF’s Noyce Scholarship program (National Science Foundation, 2016). In 2012, the most recent year for which data are available, around 1250 unique individuals obtained a first year of scholarship or stipend support through Noyce (Bobronnikov et al., 2014). This is less than 1% of the national need. Thus the institutions that have traditionally supplied the United States with teachers are not supplying enough in STEM shortage areas, and programs intended to rectify these problems are doing so at a smaller scale than would be needed to solve the problem.

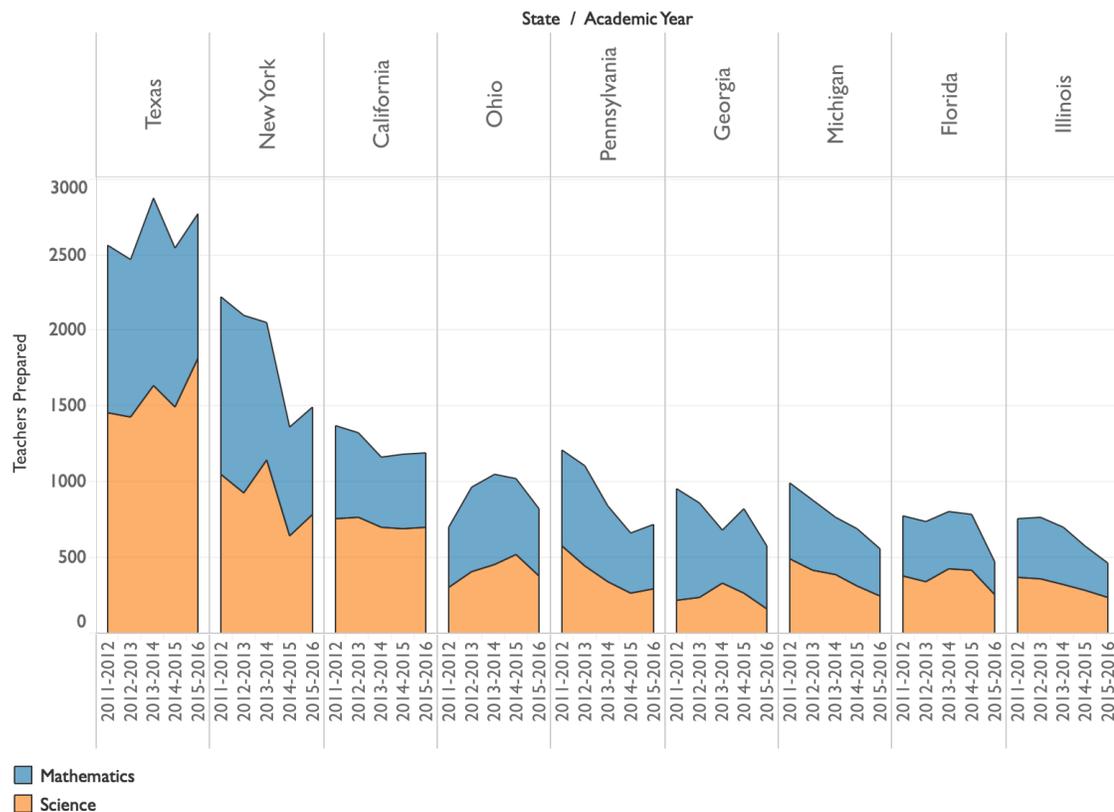


Figure 2. Mathematics and science teachers prepared in the highest-producing states, 2012-2016
Data from US Department of Education (2018a).

Not only have teacher shortages persisted for decades (National Commission on Excellence in Education, 1983), the quality of preparation programs in colleges of education has been subject to criticism. Standard preparation programs stand accused on the one hand of being an “industry of mediocrity” (Greenberg, McKee, & Walsh, 2013) and on the other hand of surviving despite low quality because they are a “cash cow” for universities (Duncan, 2010). Some reformers have advocated allowing almost anyone with a content degree to enter teaching, and sorting out who is qualified to continue by assessing their performance on the job (Hess, 2002). Value-added models have specifically been recommended as the way to identify effective teachers (Gordon, Kane, & Staiger, 2006; Kane & Staiger, 2008). The combination of skepticism about colleges of education and concern over teacher shortages has motivated policy-makers across the country to allow alternative pathways to teaching. Alternative certification programs now exist in all states, but there is great variation in the regulations that control what they are able or not able to do. Some alternative certification programs live within universities and differ only slightly from the standard programs at the same institutions. Others are web-based and come close to implementing the recommendation that almost anyone with a content degree should be allowed into teaching as rapidly as possible.

In the general context of market-based education reform, national policies affecting educator preparation are developing to incorporate two new ingredients: accountability from value-added models, and parallel preparation systems.

Accountability systems can use student test scores not only to evaluate the effectiveness of educators, but to evaluate the effectiveness of educator preparation programs. This nearly became mandatory across the country in 2016. In the fall of that year, the US Department of Education released guidance that directed every state to develop ratings of each Teacher Preparation Program (TPP). The guidance required states to “make meaningful differentiations in teacher preparation program performance” (US Department of Education, 2016, p. 670). To accomplish this goal, “For each year and each teacher preparation program in the State, a State must calculate the aggregate student learning outcomes of all students taught by novice teachers,” where a novice teacher is a “teacher of record in the first three years of teaching” (US Department of Education, 2016, p. 656). Poor performance according to these measures could lead to consequences as severe as program closure. The rules spurred considerable opposition (Tatto et al., 2016), and they were eventually rescinded by Congress in the spring of 2017. Even so, this idea has not gone away at the state level. For example, current rule holds Texas TPPs accountable for “achievement, including improvement in achievement, of students taught by beginning teachers for the first three years following certification, to the extent practicable” (Texas Education Agency, 2019). Technical challenges have so far prevented implementation.

The second development is expansion of parallel teacher preparation systems that transform teacher development from a preservice to an in-service activity. Parallel systems have mainly drawn attention in connection with Teach for America and New York City Teaching Fellows (Shiva Mungal, 2015; Shiva Mungal, Trujillo, & Scott, 2016). We will report here on the intersection of parallel preparation systems with for-profit certification, which provide the possibility of rapid growth to large scale.

Much of the research addressing these reforms has focused on the use of student test scores to hold teacher preparation programs accountable. The main finding, as we will review in the next section, is that it is very difficult to discern differences between programs based on the learning gains of the students of their graduates. These research findings may have been published with the expectation of warding off use of value-added models to evaluate educator preparation programs. However, a consequence, intentional or not, has been to provide a research base that supports expansion of for-profit alternative teacher preparation pathways. For if there is no measurable difference in student test scores between teachers from standard and alternative programs, then there is no reason states should not change their rules to enable alternative pathways.

Background

The study is set in Texas. This is for the simple reason that, as shown in Figure 1, Texas stands out from the rest of the country in the number and fraction of teachers coming from alternative certification programs. More than two decades ago Texas created a parallel certification structure that operated without much attention and now provides more than half of the state’s new teachers each year. The Texas experience should be of interest to the rest of the country because the Texas experience may be coming to the rest of the country. The two largest Texas companies providing alternative certification are expanding to other states. In the period between 2016 and the end of 2018 these companies secured permission to operate in Florida, Nevada, Utah, Indiana, South Carolina, North Carolina, Hawaii, Arizona, Michigan, Louisiana, and the District of Columbia (iTeach, 2019; Teachers of Tomorrow, 2018). Additional states are likely to follow suit. Thus there

should be interested in acquiring evidence about the effects on students of this growing force in education.

Alternative Certification in Texas

Alternative certification of teachers was first permitted in Virginia in 1982, soon followed by California, Texas, and New Jersey (Suell & Piotrowski, 2007). Alternative certification is difficult to define precisely and can encompass a wide range of programs. In broad terms, alternative certification refers to “pathways designed to attract a wider range of candidates into teaching generally by reducing or eliminating pre-service education coursework and speeding paid entry into the classroom” (Grimmett & Young, 2012, p. 34).

The first alternative certification program in Texas was offered by the Houston Independent School District in 1985, with other school districts following suit. In 2000, the first community college offered an alternative certification program, and the first for-profit company began certifying teachers in 2002 (Etheredge, 2015). Figure 3 shows the numbers of mathematics and science

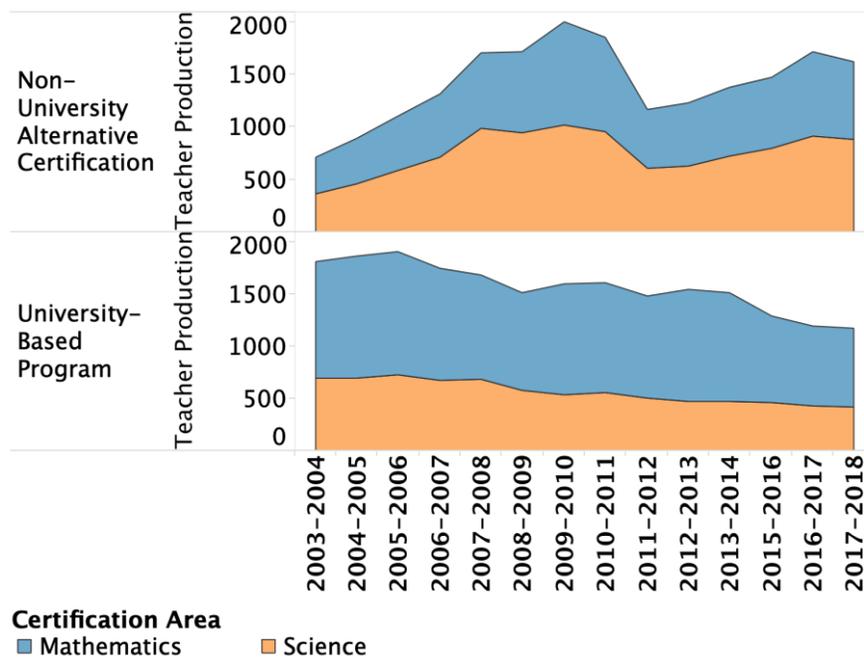


Figure 3. STEM teacher production in Texas from 2004 until 2017, comparing production from standard and alternative certification pathways

Data from Texas Education Agency.

teachers prepared through university-based programs and non-university alternative certification programs in Texas since 2004. From 2004 until 2011, the numbers of alternatively certified teachers rose steadily. In the fall of 2011, production from alternative certification programs suddenly dropped by half, due to widely publicized budget cuts that led tens of thousands of teachers to face dismissal. In such an environment, a drop in the number of people seeking to change careers and enter teaching through an alternative certification program is to be expected. Production eventually recovered, and by 2017 over half of Texas’ new STEM teachers were certified in alternative programs. Figure 4 shows the total number of STEM teachers prepared since 2004. As this figure shows, the number of STEM teachers prepared in 2016-2017 was essentially unchanged from a decade before. Therefore, introducing alternative certification programs in Texas has offset the

decline in the production of STEM teachers from standard certification pathways illustrated in Figure 3, but has not led even to increases one might expect just on the basis of population growth.

The National Research Council describes one particular difficulty in assessing the effectiveness of teacher preparation programs (NRC, 2010): “there is more variation within categories such as ‘traditional’ and ‘alternative’ — and even within the category of master’s degree programs — than there is between the categories” (p. 2). This worry is less applicable to Texas than it may be to other jurisdictions. In practice, the standard and alternative pathways are so different as to form parallel certification systems (Shiva Mungal, 2015).

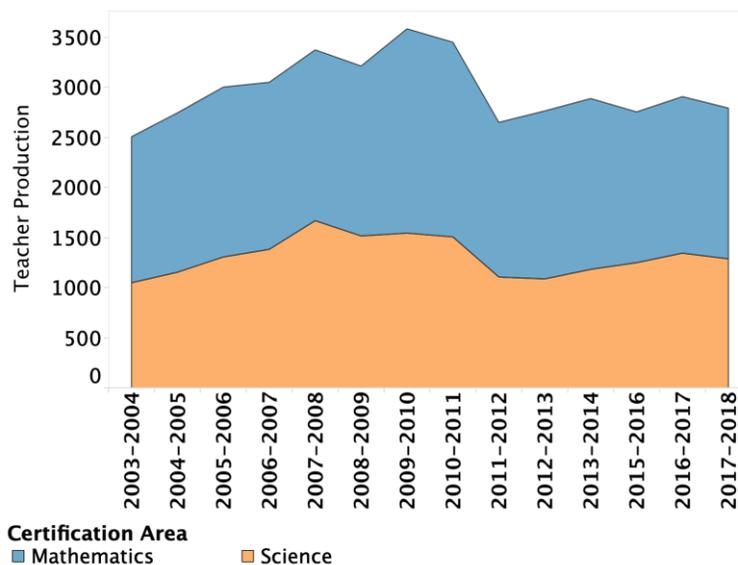


Figure 4. STEM teacher production in Texas from 2004 until 2017, summing all pathways
Data from Texas Education Agency.

Alternative certification emerged from the philosophy that barriers to teaching should be removed. The candidates, all of whom already have finished a first Bachelor’s degree, usually have a few weeks of instruction and observation after which they enter the classroom working full time, completing their pedagogical coursework during an internship year, under a Provisional (or Internship) certificate. Thus, for alternative certification candidates, teacher preparation is an in-service program.

By contrast, standard university programs provide coursework, often but not necessarily as part of a degree, as well as fieldwork prior to a student teaching semester. The candidates then begin teaching full time with a standard certificate. Thus, in the standard model, teacher preparation is a preservice program. This distinction is not apparent from the state rules, since universities are permitted to place students in schools with Provisional certificates, while alternative certification programs are permitted to offer their candidates a student teaching experience. Nevertheless, when we describe our study sample, we will show that the standard and alternate programs are largely distinct when it comes to specific preparation practices.

There is an additional respect in which one might wonder if standard and alternative pathways differ, and that is the undergraduate major of the students. However, from 1987 until 2019, secondary teachers were forbidden to major in education in Texas. Therefore, for all pathways, the teachers have majored in one of the subjects they will teach, typically mathematics for mathematics teachers, and biology for biology teachers.

The Use of Value-Added Modeling to Evaluate Teachers and their Preparation Programs

Value-added modeling arose from work of Hanushek (1971) and Sanders & Rivers (1996). Value-added models use multilevel linear regressions to estimate the expected scores of students in a classroom, based on prior test scores and demographic characteristics. Deviations between this estimate and actual classroom performance can be attributed to the skill of the teacher.

Value-added models are particularly controversial when they are used to make high-stakes decisions about individual teachers (Amrein-Beardsley, 2009; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). One of the largest problems is that the estimates are noisy. It violates basic fairness to make promotion and dismissal decisions about individual teachers based on a volatile stochastic process, even if the estimates used for the decisions are shown to be right on average.

In addition to large margins of uncertainty, value-added models are prone to systematic bias, and small technical changes can have the effect of raising or lowering the apparent value added by whole groups of teachers. This problem should be possible to address with technical improvements, and the conventions that guide practitioners are evolving (Koedel, Mihaly, & Rockoff, 2015; Rivkin, Hanushek, & Kain, 2005). For example, it was once common to see models where students' prior test scores entered to linear order, and this created a systematic bias in favor of teachers whose students' average prior scores were at the high or the low end of the possible range (Marder, 2012). In the last five years it has become common (e.g. Backes, Goldhaber, Cade, Sullivan, & Dodson, 2018) to include students' prior scores to quadratic or cubic order, as a result of which this particular source of bias goes away.

Value-added models can be used in three separate ways in relation to teacher preparation. First, they can be used to evaluate student learning gains due to teachers from specific programs. Examples here are evaluations of Teach for America (Clark et al., 2013; Decker, Mayer, & Glazerman, 2004; Turner, Goodman, Adachi, Brite, & Decker, 2012) and of UTeach (Backes et al., 2018). According to Clark et al. (2013) the difference in value-added effectiveness between TFA graduates and those of comparison programs is .06 standard deviations, and the value added by UTeach graduates is on the same order. These findings set the scale for the largest value-added differences one might expect to find for individual programs.

Second, value-added models can be used to compare a variety of teacher preparation programs in a region or state. Well-studied regions include Florida (Harris & Sass, 2011; Sass, 2011), North Carolina (Henry, Bastian, et al., 2014; Henry, Purtell, et al., 2014), Washington State (Cowan & Goldhaber, 2016; Goldhaber, Liddle, & Theobald, 2013), Missouri (Koedel, Parsons, Podgursky, & Ehlert, 2015), New York City (Boyd, Goldhaber, Lankford, & Wyckoff, 2007; Boyd et al., 2012; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Kane, Rockoff, & Staiger, 2008), California (Kane & Staiger, 2008) and Texas (Mellor, Lummus-Robinson, Brinson, & Dougherty, 2008; von Hippel, Osborne, Lincove, Bellow, & Mills, 2016).

The previous Texas results are of particular relevance for us. Mellor et al. (2008) studied student test scores in classrooms of novice teacher graduates from University of Texas System campuses, with data from 2003 through 2007. Their primary goal was "to determine how student achievement in the classroom might be used as an indicator of the success of teacher preparation programs" (p. 8). At the time there was no state-wide data system in place and they spent years obtaining data from over 400 districts. They carried out a variety of comparisons with multilevel models, but almost none of the effects they found was large enough to rule out having been caused by sampling uncertainty. They sum up by saying, "Our most significant finding was that limitations of most state data and assessment systems, including the one in Texas where our study was conducted, make this kind of research difficult" (p. 24). Six years later, the problem of evaluating

learning gains due to Teacher Preparation Programs (TPPs) was revisited by von Hippel et al. (2016), now with the advantage of a statewide data set. They could not detect program differences and conclude, “The potential benefits of TPP accountability may be too small to balance the risk that noisy TPP estimates will encourage needless, disruptive, and ineffective policy actions” (p. 2). These conclusions are similar to the findings of (Koedel, Parsons, et al., 2015) in Missouri, and have since been extended to several other states (von Hippel & Bellows, 2018).

Third, value-added models can be used to examine the efficacy of broad classes or types of preparation pathways (Harris & Sass, 2011). The present study of teacher preparation pathways in Texas falls into this third class. Some studies (Gordon et al., 2006; Kane et al., 2008) conclude that factors such as preparation routes and advanced degrees have almost no measurable effects on student outcomes. Boyd et al. (2012), analyzing some of the same data from New York City as Kane et al. (2008), conclude that differences in teacher background can be detected; the difference between the studies lies in how the models were constructed. The models of Boyd et al. (2012) pay more attention to grouping teachers with similar characteristics and from similar programs. The largest single effect in the base model of Boyd et al. (2012) is that a teacher have five years of experience, which corresponds to a value-added gain of 0.1 standard deviations in student test scores for middle school mathematics. The largest program differences, which are for Teach for America corps members, are around 0.05 standard deviations, while for College Recommended teachers the effect is around 0.02 standard deviations.

The net result of studies comparing different types of teacher preparation programs has been sustained uncertainty. The National Research Council determined that “Because the information about teacher preparation and its effectiveness is so limited, high-stakes policy debates about the most effective ways to recruit, train, and retain a high-quality teacher workforce remain muddled” (NRC, 2010, p. 5). Grossman & Loeb (2008, p. 185) similarly conclude that “[t]he available research does not paint a complete picture of either optimal recruiting and selection criteria nor optimal preparation opportunities.” In the absence of convincing results about pathways, it is not surprising that the US Department of Education decided that “effectiveness of graduates is not associated with any particular type of preparation program, [so] the only way to determine which programs are producing more effective teachers is to link information on the performance of teachers in the classroom back to their teacher preparation programs” (US Department of Education, 2016, p. 566).

Although some scholars argue that noisy value-added model estimates make it difficult to reliably determine the effectiveness of individual TPP’s (von Hippel & Bellows, 2018), there is some experimental evidence (Kane & Staiger, 2008) that the estimates can be accurate when averaged over a sufficiently large number of teachers. As such, using value-added models to evaluate teacher preparation pathways can provide reliable estimates of their efficacy in improving student test scores (Constantine, Player, Silvaa, Grider, & Deke, 2009; Guarino, Santibanez, & Daley, 2006; Wayne & Youngs, 2003). Rather than determine the efficacy of individual TPP’s, we seek to estimate the effects of alternative and standard teacher preparation pathways on student achievement using statewide available in Texas. Given the large number of teachers we include in our study, the estimates we provide for the efficacy of teacher preparation pathways are reliable.

Theoretical Framework

We make use of value-added models, which are a standard tool in addressing questions such as those in this paper. A difficulty of these models, which is common to studies using administrative data rather than random assignment, is that they permit confounding associations.

There are no experimental studies that address directly the questions we pose. Clark et al. (2013) and Kane & Staiger (2008) switched teachers randomly between classrooms in a school. This

randomization would be useful if teachers from some pathway were preferentially assigned to certain sorts of students within each school. We examine this possibility later and find no evidence for it. The main confounding association we actually find is that schools with higher percentages of disadvantaged students are more likely to have alternatively certified teachers.

An experimental study to eliminate this confounding association would need to allow teachers to be hired into schools by the normal process and then, once they had been given their teaching assignments, randomly switch them between schools. The practical and ethical barriers to such a study are so considerable it is unlikely ever to be carried out.

The modern theory of causality provides tools to permit thinking about such problems while designing and analyzing data. Two important figures are Rubin and Pearl (Imbens & Rubin, 2015; Pearl, 2009; Rosenbaum P. & Rubin, 1983; Rubin, 1974, 2005). Texts and review articles summarize their methods for application in the social sciences (Holland, 1986; Morgan & Winship, 2015; Vanderweele, 2015).

A strong recommendation from this literature is for studies such as this one to present relations between variables in graphical form with diagrams where arrows indicate the direction of causality, and absence of an arrow indicates absence of a causal relation. We provide such a diagram for our system in Figure 5. When we construct hierarchical linear models, we will provide four primary variants. The assumptions in these variants are indicated through three causal links that are only present in some of the models as indicated.

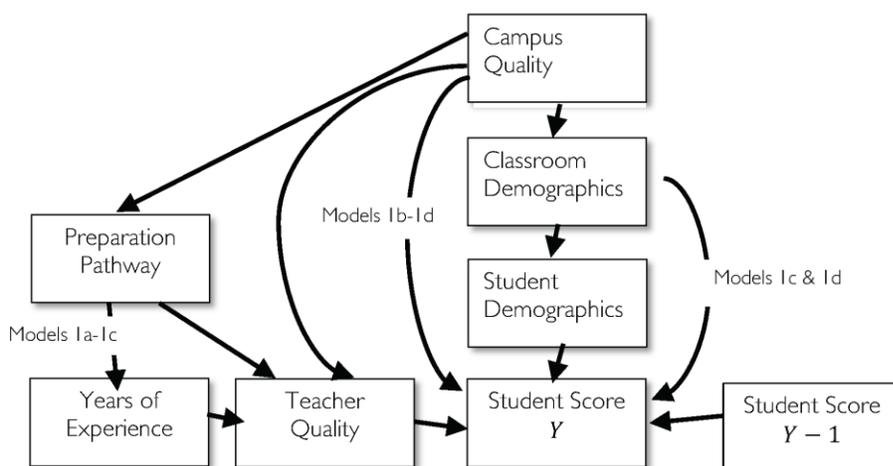


Figure 5. Directed Acyclic Graph indicating the main causal relations between variables in the study. Students Scores are measured in years Y and $Y-1$. The causal assumptions of four primary model variants can be obtained by adding causal links as indicated

In two of the models, student score depends on classroom averages of demographic variables, and in one of the models, teacher quality depends upon years of experience independent of preparation pathway.

The most consequential causal assumption is whether Campus Quality impacts Student Score independent of Teacher Quality. If one assumes that excellent campuses are excellent mainly because of the quality of the teachers, then there is no direct causal link from Campus Quality to Student Score in year Y . This is the formal way that the confounding association between teacher and school quality shows up. We have chosen to deal with it by constructing models that correspond to the two different causal assumptions. For Algebra I, results do not vary much when we do this,

and this gives us confidence in the conclusions. For Biology the results vary much more, which is why we end up more uncertain about the effect of teacher preparation pathways in Biology.

Note in all models the absence of an arrow from Student Demographics to Preparation Pathway. This represents a claim we will substantiate in the next section that within a school, the characteristics of students in a classroom are uncorrelated with the preparation pathway of the teacher.

Two additional concepts are the importance of thinking in concrete terms about interventions, and the need to think through the causal relations between known variables. An example pertinent to the current case may help explain. A robust finding of the literature on teacher quality is that teachers get better with experience, at least up through their first five years (Boyd et al., 2012). We will show that teachers from standard programs are more likely to remain in teaching than those from alternative programs. Therefore, one of the reasons that students of teachers from standard programs might learn more is that teachers from standard programs on average are more experienced. Now imagine some policy intervention that has the effect of increasing the fraction of teachers in a state that comes from standard programs. How should one estimate the effect on students of this intervention?

Causal reasoning says that if staying longer in teaching follows from attending a standard program, and if attending a standard program follows from the policy intervention, then it is fine to attribute increases in student test scores from their experienced teachers to the policy. This means that in constructing a value-added model for teachers where the population is in steady state, one should not control for years of experience. In Figure 5, Years of Experience depends on Preparation Pathway for three of the models, and therefore these models do not include a control for Years of Experience. Formally, controlling for the causal descendent of a factor introduces new confounding associations (Morgan & Winship, 2015, pp. 101–109). This argument is only fully convincing if the age population of teachers from different pathways has reached a steady state, but this is not the case, and for this reason we do present a model including Years of Experience.

Given the competing causal explanations that attach to models with different terms, we have run many different forms of the models, including many variants we do not have space to present. We regard our results as most reliable when they are robust against all our variations.

Data Source

We obtained our data from the Texas Educational Research Center (ERC). This is a data repository populated by a number of different Texas agencies, including the Texas Education Agency and the Texas Higher Education Coordinating Board. We first obtained access in 2014, with a request to study student outcomes in high school STEM disciplines as a function of teacher preparation pathway. The right to access and publish results depends upon staying within the confines of the original request.

The data sources are longitudinal, meaning every teacher and every student has a unique anonymous identifier that follows them through the years. For the purpose of this study, we needed to be able to link students and teachers, so that we would know the preparation background of the teachers and associate them with their students. Although the ERC has test score data going back to as far as 1994, links between students in a class and their teacher only became available for the year 2012. Thus when we first applied for data access we had available one year's worth of data, while at this point we have seven years to work with. The only high school STEM exams offered over this time period are Algebra I and Biology, and these are the exams we study. Essentially all Texas students take these exams; however, tens of thousands of students take Algebra I and its associated exam at the end of middle school in eighth grade. The population of eighth graders taking Algebra I

is quite different from the population taking it in ninth grade or later. The eighth-grade students have much higher prescores on average and much lower incidence of free and reduced lunch. In this study we restrict our attention to the ninth grade Algebra I population only. Almost all students take Biology in ninth grade, so the question of restricting to that population does not arise.

Descriptive Statistics

Teacher Retention

We begin with some straightforward characterizations of teachers coming from different pathways and the schools in which they work. Our first finding is that in recent years STEM teachers coming from standard programs have been more likely to stay in teaching than those from alternative programs. We investigated this by examining teachers appearing in the Texas dataset, excluding those coming in from out of state. Once they appeared on the teaching roster somewhere in the state, we checked if they were still teaching n years later and computed the fraction remaining in teaching as a function of years in service and program pathway. The result appears in Figure 6. We see that for the period 2003-2007, STEM teachers from alternative and standard pathways stayed in teaching at approximately the same rate for any number of years of service. However, for STEM teachers entering the profession between 2008 and 2012, at the five-year mark 5% more from standard pathways were still in teaching, while for those entering after 2013, 10% more were still in teaching at the five-year mark. These findings for STEM teachers are compatible with the results for teacher retention by pathway for all teachers (Ramsay, 2017b).

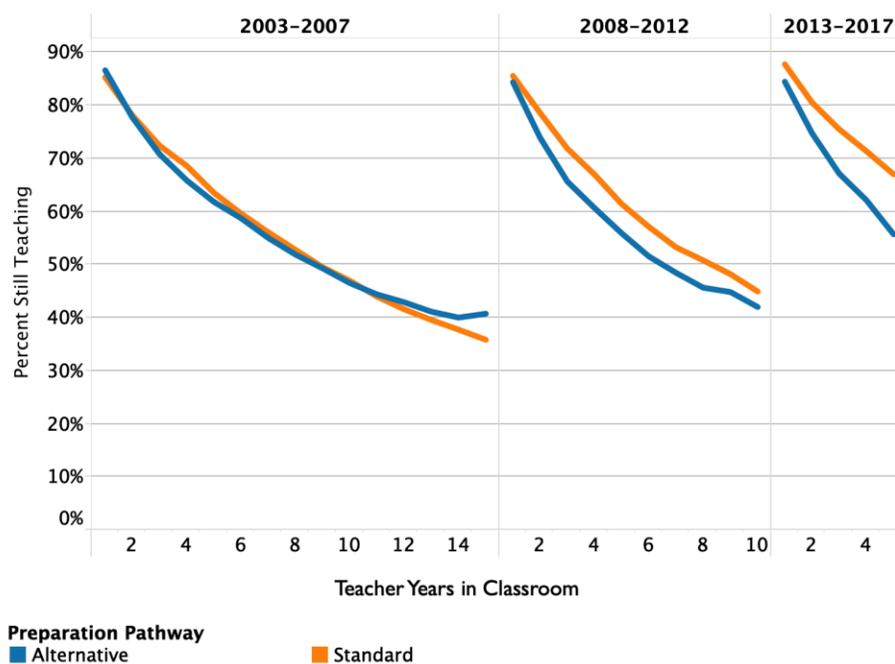


Figure 6. Retention of Texas STEM teachers in teaching by preparation pathway, using three cohorts, the first averaged over teachers entering 2003-2007, the second entering 2008-2012, the third entering 2013-2017 using data through 2018

Data from Texas Education Agency through Education Research Center.

Thus, 15 years ago when the fraction of teachers from alternative pathways was small (Figure 3), the teachers entering in this fashion were just as committed to continuing as those from the standard routes. More recently, as teachers from alternative pathways have become dominant in the state, their likelihood of staying in teaching has dropped further and further behind that of the teachers from standard pathways.

Teacher Population

The specific population of teachers we study in detail is Algebra I and Biology teachers from the years 2011-2012 through 2017-2018. We do not include all the Algebra I and Biology teachers in the state. Our dataset has those in public schools, regular and charter, but no private schools. For the purposes of this study, we exclude teachers who were certified out of state. Furthermore, because the period in which alternative certification programs existed in Texas extends back approximately two decades, we consider teachers with up to 20 years of experience. A cutoff is appropriate because more than 20 years ago the teacher preparation environment in Texas was so different than it is today it does not make sense to use data from teachers prepared then as a guide to the future.

Another sub-population of teachers described by years of service merits special attention. Federal teacher preparation regulations from 2016 specified that teacher preparation programs were to be assessed on the performance of their novice teachers, defined as those with less than four years of experience (US Department of Education, 2016, p. 68). These regulations were rescinded, but Texas and other states are moving to put in place such evaluation systems anyway. Thus, we decided to use two teacher experience groups: novice teachers, defined as those with less than four years of experience, and all teachers with up to twenty years of experience. Table 2 displays the total numbers of teachers in our sample for each academic year, and for these two levels of experience. The number of teachers prepared in Texas teaching either Algebra I or Biology started at around 3000 and increased by around 20% over the seven years we examined. However the number of novice teachers assigned to these courses dropped by nearly half. We will return to this point later.

Table 2

Numbers of teachers in sample for each academic year in study, each subject, and two levels of experience (Years Exp.)

Subject	Years Exp.	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018
Algebra I	<20	2975	3055	3042	3180	3337	3448	3622
Algebra I	<4	814	657	544	502	520	488	466
Biology	<20	3079	3088	3040	3088	3252	3305	3424
Biology	<4	845	677	624	580	580	547	490

Now we look further into teachers assigned to Algebra I and Biology courses each year, and examine the pathways that led them into teaching. Three variables describing them can appear in principle in any combination. Teachers can come from a standard or alternative program, they can come from an Institution of Higher Education (IHE), or not, and they can enter teaching on a Standard or Provisional Certificate. The Standard certificate means they have a student teaching experience; otherwise they enter classrooms as full-time paid teachers on Provisional (or Intern) certificates without having had student teaching. Table 3 shows the likelihood of combinations of these variables for the subjects, years of experience, and academic years we are considering. Matters are not quite as simple as a binary distinction between standard and alternative programs, but cases

that muddle the boundaries are not common. For example, fewer than 10% of teachers come from university post-baccalaureate programs with Provisional certificates. The entities offering alternative certification programs are varied. They include school districts, state-supported education service centers and universities. However the largest providers by far are companies that advertise low cost and provide many services online (“iTeach,” 2019; *Teachers of Tomorrow*, 2018).

Table 3

Percentages of teachers who followed various pathways. Percentages in each block and each column sum to 100%, apart from rounding. Rows shaded in grey are included in our standard category, while all others will be designated as alternative. First Certification (First Cert) is Standard (Std) or Probationary (Prob). Certification Program (Cert Prog) is Standard (Std), Alternative (Alt) or Post-Bacc (PB). Type is Institute of Higher Education (IHE) or not (N)

Subj.	Years Exp.	First Cert	Cert Prog	Type	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018
Alg. I	<20	Std	Std	IHE	35%	36%	38%	38%	38%	38%	39%
Alg. I	<20	Std	PB	IHE	6%	5%	5%	4%	4%	4%	3%
Alg. I	<20	Prob	Alt	N	42%	43%	43%	43%	44%	44%	45%
Alg. I	<20	Prob	Alt	IHE	8%	7%	6%	7%	7%	6%	5%
Alg. I	<20	Prob	PB	IHE	7%	6%	6%	6%	6%	5%	5%
Alg. I	<20	Std	Alt	N	1%	1%	1%	1%	1%	1%	2%
Alg. I	<20	Std.	Alt	IHE	1%	1%	1%	1%	1%	1%	1%
Alg. I	<4	Std	Std	IHE	16%	19%	20%	22%	24%	22%	21%
Alg. I	<4	Std	PB	IHE	5%	5%	5%	4%	3%	3%	3%
Alg. I	<4	Prob	Alt	N	64%	65%	64%	62%	61%	60%	62%
Alg. I	<4	Prob	Alt	IHE	6%	5%	3%	2%	4%	3%	3%
Alg. I	<4	Prob	PB	IHE	5%	2%	2%	3%	4%	5%	4%
Alg. I	<4	Std	Alt	N	2%	2%	4%	4%	2%	4%	5%
Alg. I	<4	Std	Alte	IHE	2%	1%	1%	1%	1%	1%	-
Bio	<20	Std	Std	IHE	25%	25%	25%	24%	24%	24%	24%
Bio	<20	Std	PB	IHE	8%	8%	7%	7%	7%	6%	6%
Bio	<20	Prob	Alt	N	48%	49%	50%	51%	52%	53%	54%
Bio	<20	Prob	Alt	IHE	8%	8%	8%	8%	8%	7%	7%
Bio	<20	Prob	PB	IHE	8%	8%	7%	7%	7%	7%	6%
Bio	<20	Std	Alt	N	1%	1%	1%	2%	1%	2%	2%
Bio	<20	Std	Alt	IHE	1%	1%	1%	1%	1%	1%	1%
Bio	<4	Std	Std	IHE	10%	12%	13%	14%	15%	13%	14%
Bio	<4	Std	PB	IHE	8%	6%	6%	6%	4%	4%	5%
Bio	<4	Prob	Alt	N	66%	64%	65%	61%	64%	67%	63%
Bio	<4	Prob	Alt	IHE	7%	7%	7%	6%	6%	5%	4%
Bio	<4	Prob	PB	IHE	6%	5%	4%	5%	4%	6%	6%
Bio	<4	Std	Alt	N	2%	3%	4%	4%	3%	3%	6%
Bio	<4	Std	Alt	IHE	1%	2%	1%	2%	1%	1%	2%

Therefore, rather than studying the eight possible pathway scenarios, we reduced them to two. We define standard programs to include teachers prepared by Institutions of Higher Education (IHE) enrolled in standard or post-baccalaureate programs and obtaining Standard first certificates. Everyone else, including some of the students from universities, we attribute to an alternative program. We ran variants of the analysis, for example including graduates of university-based alternative programs who began teaching with a standard certificate in the standard group. However as the numbers of teachers in such classifications were small, and none of the conclusions in our analysis were affected, we report only results from the comparison groups we have just described. The great majority of teachers come either from standard university programs with student teaching or from alternative programs without it.

For teachers with up to 20 years of experience in our sample, around 40% come from standard programs and 60% from alternative programs. For novice teachers, 30% or less come from standard programs, and 70% or more from alternative programs. These percentages illustrate the scale to which alternative certification has grown in Texas. However, please note that the detailed results in this section describe the teachers in our sample assigned to teach particular high-stakes courses, not teachers in the state overall.

Student Population

Next we turn attention to the characteristics of students, and the associations between teacher pathway and student demographics. Table 4 provides descriptions of students appearing in our sample, comparing the classrooms of teachers from standard and alternative pathways. We aggregate all years together, since changes over time were not worth remarking. Standard and alternatively certified teachers have significantly different student populations. The columns labeled “Standard” and “Alternative” report the percentage of teachers’ students with a certain characteristic. For example, 3.4% of the students of teachers from standard pathways were designated as Gifted. The alternatively certified teachers have a higher fraction of their students who are eligible for free and reduced lunch and who are Black and Hispanic. In general, for any factor that tends to lead to lower student outcomes, alternatively certified teachers have more of these students. This makes it important to control for these student characteristics in the analysis.

We investigated whether the difference in student populations of the teachers from the two pathways is mainly within schools or between schools. To do this we constructed hierarchical linear models (Bolker et al., 2016; Gelman & Hill, 2007) of the form

$$X_i = C_i + \text{StdCert}_i + \epsilon_i; C_i \sim N(\mu_C; \sigma_C^2), \quad (0)$$

where X_i is a demographic characteristic of student i , C_i is their campus, StdCert_i is the certification pathway of their teacher, ϵ_i here and elsewhere is a random term making X_i normally distributed, and N is the normal distribution. The results appear in the final column of Table 4. They show that after one controls for the campus, the difference between student populations of teachers from standard and alternative pathways becomes much smaller, often insignificant, and may reverse sign. The difference in student populations is mainly due to the schools in which teachers from standard and alternative pathways are likely to work. Within a given school the teachers from standard and alternative programs are not preferentially assigned to one sort of student or another. The final row of the table describes students who are tracked into Algebra I in eighth grade.

Table 4

Average characteristics of students for standard and alternative teachers with up to 20 years of experience over all years in the study. All differences between columns labeled Standard and Alternative are significant with $p < 0.001$. Final columns show difference between classroom demographics of teachers from standard and alternative pathways after controlling for school identity with Eq. (0). Numbers in parentheses show one standard uncertainty

Category	Discipline	Standard	Alternative	Standard - Alternative, School Control	
EcoDis	Algebra I	52.1%	58.7% (0.1%)	-0.2% (0.1%)	**
Gifted	Algebra I	3.4%	3.6% (0.0%)	0.1% (0.0%)	***
SpecEd	Algebra I	5.3%	5.4% (0.0%)	-0.5% (0.0%)	***
LEP	Algebra I	7.2%	9.6% (0.0%)	0.3% (0.0%)	***
Asian	Algebra I	2.0%	1.9% (0.0%)	0.2% (0.0%)	***
Black	Algebra I	12.0%	16.1% (0.0%)	-0.2% (0.0%)	***
Hispanic	Algebra I	52.8%	56.5% (0.1%)	0.1% (0.1%)	**
White	Algebra I	30.9%	23.4% (0.1%)	0.0% (0.1%)	
EcoDis	Biology	44.4%	51.1% (0.1%)	0.1% (0.1%)	
Gifted	Biology	11.1%	10.9% (0.0%)	-0.1% (0.0%)	
SpecEd	Biology	3.9%	3.9% (0.0%)	0.1% (0.0%)	***
LEP	Biology	5.2%	6.9% (0.0%)	0.3% (0.0%)	***
Asian	Biology	4.1%	4.0% (0.0%)	0.0% (0.0%)	
Black	Biology	10.6%	14.1% (0.0%)	-0.1% (0.0%)	**
Hispanic	Biology	48.3%	52.3% (0.1%)	0.1% (0.1%)	
White	Biology	34.5%	27.4% (0.1%)	0.0% (0.1%)	
Track	Biology	28.2%	26.4% (0.1%)	0.0% (0.1%)	

* $|t| > 1.96$, ** $|t| > 2.33$, *** $|t| > 3.09$

Methods

Multilevel Models for Student Test Scores

We studied changes in student test scores through multilevel models where students are nested within classroom, classrooms are nested within teacher, teachers are nested within campus, we control for each student's pre-score, an array of demographic information about student and campus, and estimate the effect of teacher pathway on student test scores. Thus, to the extent possible, the models compare teachers with other teachers teaching the same subject in the same school and attempt to compensate for differences in school and classroom populations.

We start with data from the 2011-2012 academic year using pre-scores from 2010-2011 and proceed through 2017-2018. The 2011-2012 academic year was the first year that student-teacher links became available in the Texas statewide dataset, and the 2017-2018 academic year provides the most recent data available².

During our study period, Texas was transitioning between sets of high-stakes standardized exams, from TAKS to STAAR. The only high school STEM exams offered during this entire period were STAAR Algebra I and STAAR Biology. In 2011-2012, pre-scores came from TAKS eighth-

² Because the study makes use of de-identified student data in highly aggregated form, institutional IRB has ruled this study exempt.

grade mathematics and TAKS eighth-grade science, while for later years scores came from STAAR eighth-grade mathematics and science.

We kept only cases where the student had a valid ninth-grade score in year Y and a valid eighth-grade score in the previous year. There were tens of thousands of students who took Algebra I in eighth grade, mainly high-achieving students in suburban middle schools. We decided not to include an analysis of this population. There were several accommodations available to students, including provisions for English-language learners, vision-impaired students, and a modified exam for students with learning disabilities. We could not simply group these students in with other students because most of them were taking a substantially different exam. Thus, in most of our analyses, we exclude all students who received any of these accommodations. However, we included a large subset of them in the following way. In our report on student sub-populations, we create a multilevel model for all students who took the alternate exam in year $Y - 1$ and also took the alternate exam in year Y .

Test score effect sizes are customarily obtained by dividing exam scores by the standard deviation. For ninth graders who took Algebra I in 2011-2012, the standard deviation was 0.17. For these same students the standard deviation of their mathematics scores the year before in 8th grade was 0.15, and other years are similar. We express all model results in units of the exam standard deviations, which is around 0.16.

Any given student test score result could end up in our data set from one to six times. The test scores appeared multiple times when the student took classes with separate identification numbers in separate semesters, when more than one teacher was associated with the class section, or if the student changed schools in Texas. We weighted every student record inversely with the number of times they appeared, so if a student was taught by several teachers during the year, each of them shared equally, and so that student did not contribute more to the final results than a student who appeared only once.

Model Specifications

We explored many different multilevel models, using `lmer` in R (Bolker et al., 2016). In our first collection of models, which are variants of Eq. (1), all data from 2011-12 through 2017-18 were included at once. At the top level, S_{iY} is the score of student i in some year Y and $S_{i,Y-1}$ is the same student's score on the exam in the same subject the previous year in a cubic polynomial. Teacher j of student i contributes through the random intercept $T_{j[i]}$. By modeling the teacher in this way, each teacher should contribute equally to the estimate of the effect of their pathway to teaching (Koedel, Parsons, et al., 2015). The campus k contributes random intercept $C_{k[i]}$ as does the class n through $Class_{n[i]}$, meaning that student i is in class n . Coefficients for student-level demographic factors X range over Gifted, racial and ethnic groups, Limited English Proficiency (LEP), Free/Reduced Lunch Eligibility (EcoDis), and Special Education. Here $g[i]$ is the value of assigned group affiliation for student i .

We modeled the influence of tracking, as recommended by Jackson (2014). The most important form of student tracking in Texas is placing students in Algebra I in eighth grade. To control for this, we removed students enrolled in Algebra 1 during their eighth-grade year from our study of mathematics and created a flag for them when modeling biology. Some variants of Eq. (1) control for classroom-level averages of the demographic variables through $\gamma_X \bar{X}_{n[i]}$ (Chetty, Friedman, & Rockoff, 2014; Friedman, Rockoff, & Chetty, 2014), and others include a variable E to control for teacher years of experience, assuming binned values of 0-4, 4-10, and 10-20 years of experience.

The main item of interest, certification pathway StdCert_m for teacher j of student i out of program m enters as a fixed effect. Finally, the second level of the model has random intercepts for teacher T , campus C , and class section Class . See Gelman & Hill (2007, Chapter 12.5) for notation.

$$S_i = \sum_{Y=2012}^{2016} \sum_{\beta=1}^3 \lambda_{\beta Y} S_{i,Y-1}^{\beta} + T_{j[i]} + E_{j[i]} + C_{k[i]} + \text{Class}_n[i] + \text{StdCert}_m[j[i]] \\ + \sum_X X_{g[i]} + \sum_X \gamma_X \bar{X}_n[i] + \epsilon_i; \quad (1)$$

$$T_j \sim N(\mu_T; \sigma_T^2); \quad C_k \sim N(\mu_C; \sigma_C^2); \quad \text{Class}_n \sim N(\mu_L; \sigma_L^2) \quad .$$

Tables 5 and 6 provide results from these models. We present results from four different variants of this general form so as to display the effect upon the variable of interest, StdCert , of progressively adding terms. The causal assumptions of these four models were illustrated in Figure 5. Model (1a) lacks a campus intercept and also lacks averages of classroom demographics. The causal assumption in this model is that high quality campuses with students who perform well are high quality mainly because their teachers are of high quality. Model (1b) adds the campus intercept as a random effect, allowing high quality campuses to improve student scores through means other than teacher quality. (1c) adds classroom averages of demographic variables, and (1d) adds a control for years of service, which is justified by assuming it does not result from teacher pathway. Model (1c) is probably the most persuasive of the model specifications, since classroom averages of demographic variables do turn out to affect the results. While controlling for years of service in (1d) is debatable on causal grounds, adding it or not does not make much of a difference.

We also created estimates in which each year was treated separately. Model 2 is similar to model (1c) with random intercepts for campus, class, and teacher, and controls both for student demographics, classroom averages of student demographics, and tracking.

$$S_{i,Y} = \sum_{\beta=1}^3 \lambda_{\beta} S_{i,Y-1}^{\beta} + T_{j[i]} + C_{k[i]} + \text{Class}_n[i] + \text{StdCert}_m[j[i]] \\ + \sum_X X_{g[i]} + \sum_X \gamma_X \bar{X}_n[i] + \epsilon_i \quad . \quad (2, \text{Random})$$

$$T_j \sim N(\mu_T; \sigma_T^2); \quad C_k \sim N(\mu_C; \sigma_C^2); \quad \text{Class}_n \sim N(\mu_L; \sigma_L^2) \quad .$$

Model 3 has the form

$$S_{i,Y} = \sum_{\beta=1}^3 \lambda_{\beta} S_{i,Y-1}^{\beta} + T_{j[i]} + C_{k[i]} + \text{Class}_n[j[i]] + \text{StdCert}_m[j[i]] \\ + \sum_X X_{g[i]} + \sum_X \gamma_X \bar{X}_n[i] + \epsilon_i \quad (3, \text{Fixed})$$

$$T_j \sim N(\mu_T; \sigma_T^2); \quad \text{Class}_n \sim N(\mu_L; \sigma_L^2)$$

This is the same as the previous, except that campus is treated as a fixed effect at the top level, rather than being modeled as a random effect at the second level. This model is less appropriate for finding the contribution of teacher pathway because in cases where a campus has teachers from only a single pathway, the campus fixed effect subtracts them off rather than comparing them with teachers in similar campuses as the campus random effect model does.

Model 4 is

$$S_{i,Y} = \sum_{\beta=1}^3 \lambda_{\beta} S_{i,Y-1}^{\beta} + T_{j[i]} + \text{Class}_n[j[i]] + \text{StdCert}_m[j[i]] + \sum_X X_{g[i]} + \sum_X \gamma_X \bar{X}_{n[i]} + \epsilon_i \quad (4, \text{None})$$

$$T_j \sim N(\mu_T; \sigma_T^2); \text{Class}_n \sim N(\mu_L; \sigma_L^2)$$

In this case there is no campus intercept. One would use this model if one adopts the view that the difference between campus performance is mainly due to the teachers and not to other non-student factors.

We had a fifth model which we applied to subgroups of students.

$$S_{i,Y} = \sum_{\beta=1}^3 \lambda_{\beta} S_{i,Y-1}^{\beta} + T_{j[i]} + \text{Class}_n[j[i]] + \text{StdCert}_m[j[i]] + \sum_X \gamma_X \bar{X}_{n[i]} + \epsilon_i \quad (5, \text{Subgroup})$$

$$T_j \sim N(\mu_T; \sigma_T^2); \quad C_k \sim N(\mu_C; \sigma_C^2); \quad \text{Class}_n \sim N(\mu_L; \sigma_L^2)$$

This model was applied after being restricted to a demographic subset of our sample, for example to the subgroup of economically disadvantaged students, gifted students, or Special Needs students taking an alternate test. This model allowed us to focus on the effect of teacher pathway on a specific group of students, while also controlling for the broader demographic compositions of their classes.

Results

Multi-level Models

We begin our discussion of results with Model (1). The random and fixed effects are given in Table 5 for Algebra I and Table 6 for Biology. The coefficients related to student subgroups for the models of Algebra I and Biology are quite similar to each other.

Among the random effects, the largest is the difference between campuses, with a standard deviation of 0.26 for Algebra I and 0.28 in Biology. The standard deviation of the difference between teachers is around in 0.25 Algebra I and 0.2 in Biology. That is, we find slightly larger differences between campuses than within them. The standard deviation of classes taught by the same teacher is between 0.13 and 0.17 in both subjects.

We report the teacher pathway effect (StdCert) as positive when students of teachers from standard programs get higher scores than those from alternative programs. As shown in Table 5, students with teachers from standard pathways gain around 0.03–0.05 in Algebra I.

Table 5

Model (1) coefficients for Algebra I. Numbers in parentheses are standard uncertainties. Different variants of the model include terms as shown. The main result of interest is in the shaded row labeled StdCert, which gives gains in standard deviation units for students of teachers from standard programs.

Random Effects	1a		1b		1c		1d	
Campus SD			0.261		0.253		0.253	
Teacher SD	0.319		0.250		0.245		0.244	
Class SD	0.188		0.175		0.172		0.172	
Fixed Effects	1a		1b		1c		1d	
StdCert	0.054	(0.007) ***	0.041	(0.006) ***	0.037	(0.006) ***	0.033	(0.006) ***
EcoDis	-0.074	(0.001) ***	-0.073	(0.001) ***	-0.070	(0.001) ***	-0.070	(0.001) ***
Gifted	0.218	(0.003) ***	0.219	(0.003) ***	0.204	(0.003) ***	0.204	(0.003) ***
SpecEd	-0.294	(0.002) ***	-0.298	(0.002) ***	-0.272	(0.003) ***	-0.272	(0.003) ***
LEP	-0.152	(0.002) ***	-0.153	(0.002) ***	-0.149	(0.002) ***	-0.149	(0.002) ***
Asian	0.148	(0.042) ***	0.145	(0.042) ***	0.133	(0.041) ***	0.133	(0.041) ***
Black	-0.081	(0.041) *	-0.080	(0.041)	-0.081	(0.041)	-0.081	(0.041)
Hispanic	-0.054	(0.041)	-0.053	(0.041)	-0.055	(0.041)	-0.055	(0.041)
White	-0.017	(0.041)	-0.017	(0.041)	-0.020	(0.041)	-0.020	(0.041)
AveEcoDis					-0.087	(0.006) ***	-0.087	(0.006) ***
AveGifted					0.390	(0.013) ***	0.389	(0.013) ***
AveSpecEd					-0.107	(0.005) ***	-0.107	(0.005) ***
AveLEP					-0.017	(0.006) **	-0.017	(0.006) **
AveAsian					0.270	(0.021) ***	0.270	(0.021) ***
AveBlack					-0.069	(0.010) ***	-0.068	(0.010) ***
AveHispanic					-0.037	(0.008) ***	-0.037	(0.008) ***
AveWhite					-	-	-	-
Tracked					-	-	-	-
Yrs: 4-10							0.019	(0.004) ***
Yrs: 11-20							0.026	(0.005) ***
$\lambda_1(2012)$	1.235	(0.077) ***	1.259	(0.077) ***	1.188	(0.077) ***	1.209	(0.077) ***
$\lambda_1(2013)$	3.051	(0.079) ***	3.042	(0.079) ***	2.946	(0.079) ***	2.964	(0.079) ***
$\lambda_1(2014)$	1.798	(0.074) ***	1.778	(0.074) ***	1.694	(0.074) ***	1.698	(0.074) ***
$\lambda_1(2015)$	1.093	(0.073) ***	1.064	(0.073) ***	0.989	(0.073) ***	0.987	(0.073) ***
$\lambda_1(2016)$	2.948	(0.078) ***	2.897	(0.078) ***	2.834	(0.078) ***	2.826	(0.078) ***
$\lambda_1(2017)$	2.405	(0.075) ***	2.339	(0.074) ***	2.304	(0.074) ***	2.290	(0.074) ***
$\lambda_1(2018)$	2.756	(0.072) ***	2.706	(0.072) ***	2.721	(0.072) ***	2.700	(0.072) ***
$\lambda_2(2012)$	-0.082	(0.163)	-0.157	(0.162)	-0.050	(0.162)	-0.082	(0.162)
$\lambda_2(2013)$	-0.250	(0.167)	-0.246	(0.167)	-0.071	(0.167)	-0.105	(0.167)
$\lambda_2(2014)$	3.829	(0.158) ***	3.844	(0.158) ***	3.981	(0.158) ***	3.974	(0.158) ***

Table 5 (Cont'd.)

Fixed Effects	1a			1b			1c			1d		
$\lambda_2(2015)$	5.633	(0.150)	***	5.663	(0.150)	***	5.771	(0.150)	***	5.775	(0.150)	***
$\lambda_2(2016)$	1.840	(0.163)	***	1.912	(0.163)	***	2.001	(0.163)	***	2.016	(0.163)	***
$\lambda_2(2017)$	4.113	(0.156)	***	4.209	(0.156)	***	4.230	(0.156)	***	4.254	(0.156)	***
$\lambda_2(2018)$	2.128	(0.144)	***	2.193	(0.144)	***	2.105	(0.144)	***	2.139	(0.144)	***
$\lambda_3(2012)$	2.111	(0.101)	***	2.158	(0.101)	***	2.098	(0.100)	***	2.113	(0.101)	***
$\lambda_3(2013)$	0.914	(0.110)	***	0.915	(0.110)	***	0.808	(0.110)	***	0.828	(0.110)	***
$\lambda_3(2014)$	-1.951	(0.106)	***	-1.952	(0.106)	***	-2.033	(0.106)	***	-2.029	(0.106)	***
$\lambda_3(2015)$	-3.041	(0.098)	***	-3.050	(0.098)	***	-3.110	(0.098)	***	-3.112	(0.098)	***
$\lambda_3(2016)$	-0.634	(0.108)	***	-0.668	(0.108)	***	-0.721	(0.108)	***	-0.730	(0.108)	***
$\lambda_3(2017)$	-2.549	(0.102)	***	-2.599	(0.102)	***	-2.609	(0.102)	***	-2.621	(0.102)	***
$\lambda_3(2018)$	-0.939	(0.090)	***	-0.968	(0.090)	***	-0.912	(0.090)	***	-0.930	(0.090)	***

* $|t| > 1.96$, ** $|t| > 2.33$, *** $|t| > 3.09$. AveEcoDis refers to fraction of class that is economically disadvantaged, and other variables beginning with Ave are defined similarly. Coefficients λ are defined in Equation 1 and refer to powers of pretest score, computed separately for each year

This result is significant in all the model specifications chosen, although the effect is nearly twice as large in the models with fewer controls. That is not surprising, since we know from Table 4 that teachers from standard programs find themselves overall in classrooms with fewer economically disadvantaged students, and more white students. In Biology, as shown in Table 6, results are significant in models 1a and 1b, but not in models 1c and 1d that include averages of classroom demographics and an intercept for each campus.

Thus the case for improved student learning in classrooms with teachers from standard programs is strong for Algebra I, and robust against all variation of the models we examined. The case is weaker in Biology than in Algebra I, where we measure a significant advantage for teachers from standard pathways only in models that attribute most of the student learning gains in schools that are high-performing overall to the teachers rather than to other factors.

The difference between models 1c and 1d is that 1d includes a control for years of experience, and 1c does not. We include the control for teacher years of experience despite the evidence that years of experience depends on pathway because it has been conventional in previous studies. One sees that whether it is included or not turns out not to make much of a difference.

We now turn to a more detailed analysis year by year and for various subgroups. Table 7 presents results for the three different models that vary the way campus effects are treated. The fixed effect model (Eq. 2, Fixed) which maximizes how much of an effect is attributed to the campus tends to give the smallest results, and the model without campus effects (Eq. 3, None) tends to give the largest results, showing that strong Algebra I teachers are associated with strong campuses. In every year and for each of the models, students gain about 0.03–0.05 in classes of Algebra I teachers from standard programs with up to 20 years of experience. The point estimates in classrooms of novice Algebra I teachers with up to four years of experience are similar, but the sample sizes are much smaller, and few of the differences are statistically significant. For Biology classrooms, there are fewer statistically significant results, and the point estimates are scattered between positive and negative. If one removes the classroom averages of demographic variables from models (2)-(5), there are many cases where students Biology teachers from standard programs have significantly higher learning gains, but these models are harder to defend than the ones we have used, and we do not report their results.

Table 6

Model (1) coefficients for Biology. Numbers in parentheses are standard uncertainties. Different variants of the model include terms as shown. The main result of interest is in the shaded row labeled StdCert, which gives the gains in standard deviation units for students of teachers from standard programs.

Random Effects	1a		1b		1c		1d	
Campus SD			0.228		0.201		0.201	
Teacher SD	0.276		0.199		0.174		0.174	
Class SD	0.163		0.153		0.136		0.136	
Fixed Effects	1a		1b		1c		1d	
StdCert	0.036 (0.007)	***	0.013 (0.005)	**	0.006 (0.005)		0.004 (0.005)	
EcoDis	-0.082 (0.001)	***	-0.078 (0.001)	***	-0.067 (0.001)	***	-0.067 (0.001)	***
Gifted	0.218 (0.001)	***	0.221 (0.001)	***	0.173 (0.001)	***	0.173 (0.001)	***
SpecEd	-0.289 (0.002)	***	-0.292 (0.002)	***	-0.248 (0.002)	***	-0.248 (0.002)	***
LEP	-0.234 (0.002)	***	-0.234 (0.002)	***	-0.219 (0.002)	***	-0.219 (0.002)	***
Asian	0.108 (0.034)	***	0.104 (0.034)	**	0.090 (0.006)	***	0.090 (0.006)	***
Black	-0.049 (0.034)		-0.049 (0.034)		-0.038 (0.006)	***	-0.038 (0.006)	***
Hispanic	-0.042 (0.034)		-0.042 (0.034)		-0.035 (0.006)	***	-0.035 (0.006)	***
White	0.031 (0.034)		0.029 (0.034)		0.031 (0.006)	***	0.031 (0.006)	***
AveEcoDis					-0.166 (0.005)	***	-0.166 (0.005)	***
AveGifted					0.265 (0.005)	***	0.264 (0.005)	***
AveSpecEd					-0.170 (0.005)	***	-0.170 (0.005)	***
AveLEP					-0.089 (0.006)	***	-0.089 (0.006)	***
AveAsian					0.138 (0.012)	***	0.138 (0.012)	***
AveBlack					-0.173 (0.008)	***	-0.172 (0.008)	***
AveHispanic					-0.107 (0.006)	***	-0.107 (0.006)	***
AveWhite					-	-	-	-
Tracked					0.156 (0.001)	***	0.156 (0.001)	***
Yrs: 4-10							-0.003 (0.003)	
Yrs: 11-20							0.010 (0.004)	**
$\lambda_1(2012)$	1.040 (0.061)	***	1.077 (0.061)	***	1.191 (0.061)	***	1.204 (0.061)	***
$\lambda_1(2013)$	-1.513 (0.060)	***	-1.489 (0.060)	***	-1.225 (0.059)	***	-1.209 (0.059)	***
$\lambda_1(2014)$	0.279 (0.059)	***	0.278 (0.059)	***	0.540 (0.059)	***	0.545 (0.059)	***
$\lambda_1(2015)$	-0.147 (0.059)	**	-0.171 (0.059)	**	0.048 (0.058)		0.051 (0.058)	
$\lambda_1(2016)$	0.051 (0.058)		0.007 (0.058)		0.240 (0.058)	***	0.239 (0.058)	***
$\lambda_1(2017)$	-0.544 (0.059)	***	-0.621 (0.059)	***	-0.403 (0.059)	***	-0.409 (0.059)	***
$\lambda_1(2018)$	-0.760 (0.060)	***	-0.857 (0.060)	***	-0.545 (0.060)	***	-0.555 (0.060)	***
$\lambda_2(2012)$	-3.713 (0.123)	***	-3.791 (0.123)	***	-3.779 (0.122)	***	-3.798 (0.123)	***
$\lambda_{d2}(2013)$	7.316 (0.119)	***	7.259 (0.119)	***	6.813 (0.118)	***	6.786 (0.118)	***
$\lambda_2(2014)$	3.616 (0.116)	***	3.603 (0.116)	***	3.149 (0.115)	***	3.141 (0.115)	***
$\lambda_2(2015)$	4.529 (0.113)	***	4.556 (0.113)	***	4.245 (0.112)	***	4.240 (0.112)	***
$\lambda_2(2016)$	4.913 (0.113)	***	4.973 (0.113)	***	4.633 (0.112)	***	4.636 (0.112)	***
$\lambda_2(2017)$	7.045 (0.115)	***	7.164 (0.114)	***	6.869 (0.114)	***	6.878 (0.114)	***
$\lambda_2(2018)$	8.518 (0.117)	***	8.686 (0.117)	***	8.122 (0.116)	***	8.138 (0.116)	***
$\lambda_3(2012)$	5.161 (0.073)	***	5.209 (0.073)	***	5.035 (0.073)	***	5.044 (0.073)	***

Table 6 (Cont'd.)

Fixed Effects	1a		1b		1c		1d	
$\lambda_3(2013)$	-2.733 (0.074)	***	-2.697 (0.074)	***	-2.581 (0.074)	***	-2.567 (0.074)	***
$\lambda_3(2014)$	-0.971 (0.071)	***	-0.962 (0.071)	***	-0.838 (0.071)	***	-0.834 (0.071)	***
$\lambda_3(2015)$	-1.279 (0.068)	***	-1.290 (0.068)	***	-1.271 (0.067)	***	-1.269 (0.067)	***
$\lambda_3(2016)$	-1.881 (0.068)	***	-1.908 (0.068)	***	-1.882 (0.067)	***	-1.883 (0.067)	***
$\lambda_3(2017)$	-3.349 (0.069)	***	-3.409 (0.069)	***	-3.416 (0.068)	***	-3.421 (0.068)	***
$\lambda_3(2018)$	-4.477 (0.071)	***	-4.570 (0.071)	***	-4.394 (0.070)	***	-4.402 (0.070)	***

* $|t| > 1.96$, ** $|t| > 2.33$, *** $|t| > 3.09$. AveEcoDis refers to fraction of class that is economically disadvantaged, and other variables beginning with Ave are defined similarly. Coefficients λ are defined in Equation 1 and refer to powers of pretest score, computed separately for each year

Table 7

Estimates from Models (2)-(4) for teachers from standard certification programs, broken down by discipline, school year, and teacher experience. Numbers in parentheses are standard uncertainties.

Discipline	Year	Exp	Random		Fixed		None			
Algebra I	11-12	<20	0.017	(0.011)		0.011	(0.012)	0.029	(0.014)	*
Algebra I	12-13	<20	0.031	(0.010)	**	0.033	(0.011)	0.034	(0.013)	**
Algebra I	13-14	<20	0.021	(0.009)	**	0.016	(0.010)	0.03	(0.011)	**
Algebra I	14-15	<20	0.027	(0.009)	**	0.02	(0.010)	0.034	(0.011)	**
Algebra 1	15-16	<20	0.041	(0.011)	***	0.031	(0.012)	0.065	(0.014)	***
Algebra 1	16-17	<20	0.046	(0.010)	***	0.037	(0.011)	0.073	(0.014)	***
Algebra 1	17-18	<20	0.038	(0.011)	***	0.036	(0.013)	0.054	(0.014)	***
Biology	11-12	<20	0.004	(0.009)		0.008	(0.010)	-0.003	(0.011)	
Biology	12-13	<20	0.013	(0.010)		0.008	(0.011)	0.024	(0.013)	
Biology	13-14	<20	0.014	(0.008)		0.013	(0.008)	0.017	(0.010)	
Biology	14-15	<20	0.008	(0.008)		0.008	(0.009)	0.014	(0.011)	
Biology	15-16	<20	0.008	(0.008)		0.012	(0.009)	0.014	(0.010)	
Biology	16-17	<20	0.018	(0.008)	*	0.019	(0.009)	0.033	(0.010)	***
Biology	17-18	<20	-0.001	(0.009)		-0.002	(0.010)	0.011	(0.011)	
Algebra I	11-12	<4	-0.033	(0.026)		-0.011	(0.034)	-0.046	(0.030)	
Algebra I	12-13	<4	0.005	(0.029)		0.06	(0.038)	-0.025	(0.034)	
Algebra I	13-14	<4	0.02	(0.029)		0.023	(0.040)	0.018	(0.033)	
Algebra I	14-15	<4	0.039	(0.028)		0.012	(0.038)	0.051	(0.033)	
Algebra I	15-16	<4	0.076	(0.039)	*	0.083	(0.067)	0.074	(0.041)	
Algebra 1	16-17	<4	0.113	(0.037)	**	0.113	(0.056)	0.129	(0.042)	**
Algebra 1	17-18	<4	0.067	(0.042)		0.07	(0.067)	0.069	(0.046)	
Biology	11-12	<4	0.036	(0.022)		0.008	(0.027)	0.053	(0.026)	*
Biology	12-13	<4	-0.017	(0.026)		-0.033	(0.036)	-0.015	(0.030)	
Biology	13-14	<4	0.002	(0.024)		-0.006	(0.031)	-0.003	(0.027)	
Biology	14-15	<4	-0.012	(0.026)		0.023	(0.037)	-0.02	(0.028)	
Biology	15-16	<4	-0.018	(0.026)		-0.061	(0.038)	-0.003	(0.029)	
Biology	16-17	<4	0.011	(0.031)		0.06	(0.049)	0.001	(0.034)	
Biology	17-18	<4	-0.054	(0.033)		-0.091	(0.054)	-0.039	(0.035)	

* $|t| > 1.96$, ** $|t| > 2.33$, *** $|t| > 3.09$

Tables 8 and 9 provide results for each combination of teacher pathway, subject, and a variety of student subgroups, using the model in Eq. 5 for teachers with up to 20 years of experience and novice teachers, respectively. Almost all of the estimates in Algebra I by subgroup indicate that students of teachers with standard certification gain between 0.03 and 0.08 in standard deviation units more than their counterparts with alternatively certified teachers. The largest differences are for students flagged as gifted, but the results from students eligible for free and reduced lunch (FRL) and those of limited English proficiency (LEP) are also noteworthy. The effects are stronger in Algebra I for teachers with up to 20 years of experience than they are for novice teachers. In Biology there are few statistically significant results, except for novice teachers in 2011-2012 and experienced teachers in 2016-2017. For both Algebra I and Biology the majority of the point estimates favor teachers from standard programs, and there are only two statistically significant results favoring teachers from alternative pathways, which are for LEP and Hispanic students of novice Biology teachers in 2017-2018.

The Algebra I results from this section are summarized in graphical form in Figure 7. The results for all students come from the model in Eq. (2), and the rest from Eq. (5). Only results for teachers with up to 20 years of experience are shown in the graph.

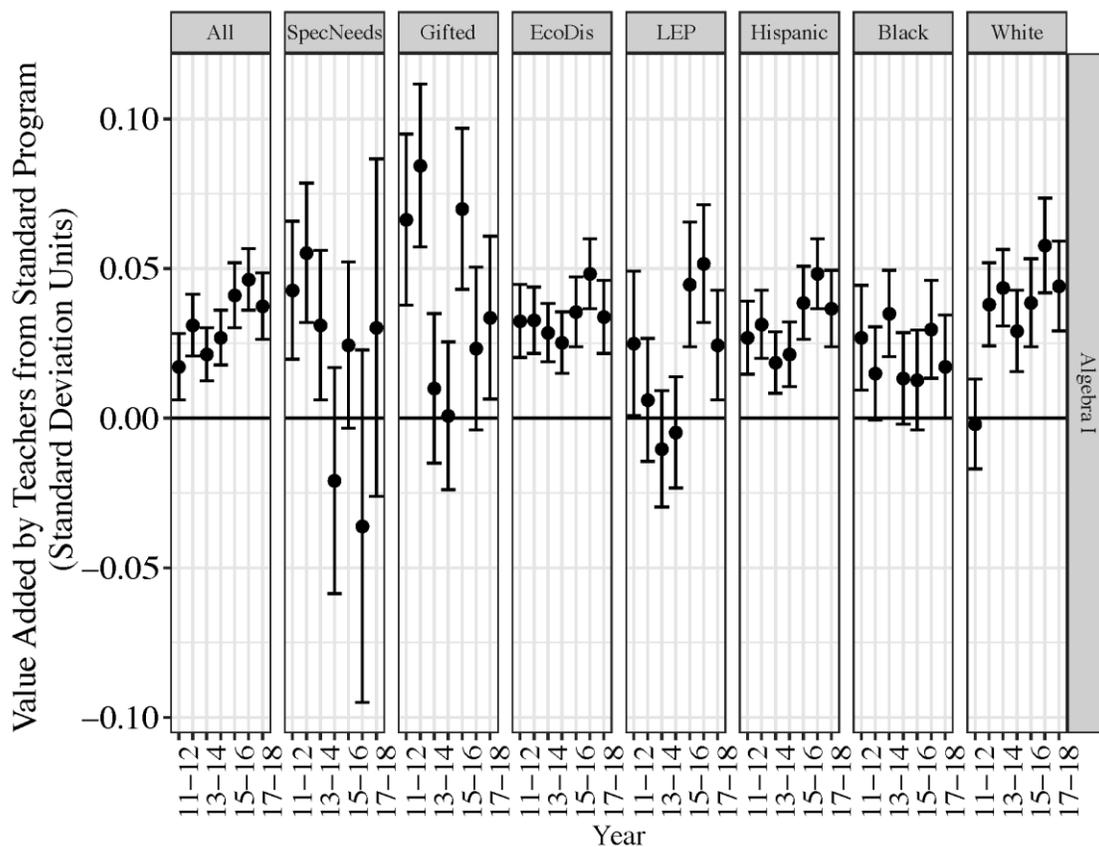


Figure 7. Value-added model gains in standard deviation units for all Algebra I teachers from standard programs as compared with teachers from alternative programs, overall and for student subgroups, showing change over time. Bars show standard uncertainties
Data from Texas Education Agency through Education Research Center.

Table 8

Estimates from Model (5) for student subgroups in classrooms of standard certified teachers with up to 20 years of experience. Numbers in parentheses are standard uncertainties

Year	Subgroup	Algebra			Biology		
11-12	Gifted	0.066	(0.029)	**	0.018	(0.018)	
11-12	Special Needs	0.043	(0.023)		0.026	(0.024)	
11-12	FRL	0.033	(0.012)	**	0.005	(0.010)	
11-12	LEP	0.025	(0.024)		0.01	(0.023)	
11-12	Black	0.027	(0.018)		0.019	(0.015)	
11-12	Hispanic	0.027	(0.012)	*	0.003	(0.010)	
11-12	White	-0.002	(0.015)		-0.008	(0.012)	
12-13	Gifted	0.084	(0.027)	***	-0.008	(0.017)	
12-13	Special Needs	0.055	(0.023)	**	-0.028	(0.023)	
12-13	FRL	0.033	(0.011)	**	-0.003	(0.010)	
12-13	LEP	0.006	(0.021)		0.026	(0.022)	
12-13	Black	0.015	(0.016)		0.015	(0.014)	
12-13	Hispanic	0.031	(0.011)	**	-0.005	(0.010)	
12-13	White	0.038	(0.014)	**	-0.001	(0.011)	
13-14	Gifted	0.010	(0.025)		0.031	(0.014)	*
13-14	Special Needs	0.031	(0.025)		0.008	(0.024)	
13-14	FRL	0.029	(0.010)	**	0.018	(0.009)	*
13-14	LEP	-0.010	(0.019)		0.003	(0.021)	
13-14	Black	0.035	(0.014)	**	0.003	(0.013)	
13-14	Hispanic	0.019	(0.010)		0.015	(0.009)	
13-14	White	0.044	(0.013)	***	0.007	(0.010)	
14-15	Gifted	0.001	(0.025)		0.018	(0.014)	
14-15	Special Needs	-0.021	(0.038)		0.033	(0.040)	
14-15	FRL	0.025	(0.01)	**	0.011	(0.009)	
14-15	LEP	-0.005	(0.019)		-0.004	(0.014)	
14-15	Black	0.013	(0.015)		-0.005	(0.015)	
14-15	Hispanic	0.021	(0.011)	**	0.011	(0.009)	
14-15	White	0.029	(0.014)	*	0.009	(0.011)	
15-16	Gifted	0.070	(0.027)	**	0.023	(0.014)	
15-16	Special Needs	0.024	(0.028)		0.051	(0.034)	
15-16	FRL	0.036	(0.012)	**	0.010	(0.009)	
15-16	LEP	0.045	(0.021)	*	0.009	(0.019)	
15-16	Black	0.013	(0.017)		-0.003	(0.014)	
15-16	Hispanic	0.039	(0.012)	***	0.006	(0.009)	
15-16	White	0.046	(0.015)	***	0.008	(0.011)	
16-17	Gifted	0.039	(0.028)		0.011	(0.013)	
16-17	Special Needs	-0.036	(0.059)		0.000	(0.071)	
16-17	FRL	0.048	(0.012)	***	0.025	(0.009)	**
16-17	LEP	0.052	(0.02)	**	0.032	(0.021)	
16-17	Black	0.030	(0.016)		0.030	(0.014)	*
16-17	Hispanic	0.048	(0.012)	***	0.021	(0.007)	*
16-17	White	0.058	(0.014)	***	0.016	(0.011)	

Table 8 (Cont'd.)

Year	Subgroup	Algebra		Biology	
17-18	Gifted	0.034	(0.027)	-0.004	(0.014)
17-18	Special Needs	0.030	(0.057)	0.055	(0.063)
17-18	FRL	0.034	(0.012) **	0.002	(0.010)
17-18	LEP	0.025	(0.018)	0	(0.018)
17-18	Black	0.018	(0.017)	-0.009	(0.015)
17-18	Hispanic	0.037	(0.013) **	0.001	(0.010)
17-18	White	0.044	(0.015) **	-0.006	(0.012)

* |t|>1.96, ** |t|>2.33, *** |t|>3.09

Table 9

Estimates from Model (5) for student subgroups in classrooms of standard certified teachers with less than 4 years of experience. Numbers in parentheses are standard uncertainties

Year	Subgroup	Algebra		Biology	
11-12	Gifted	0.063	(0.064)	0.033	(0.039)
11-12	Special Needs	-0.08	(0.056)	0.115	(0.051) *
11-12	FRL	-0.022	(0.028)	0.043	(0.023)
11-12	LEP	-0.074	(0.05)	0.113	(0.05) **
11-12	Black	0.01	(0.042)	0.066	(0.033) *
11-12	Hispanic	-0.036	(0.028)	0.05	(0.024) *
11-12	White	-0.069	(0.04)	0.008	(0.029)
12-13	Gifted	0.063	(0.07)	-0.066	(0.045)
12-13	Special Needs	-0.012	(0.063)	-0.041	(0.055)
12-13	FRL	-0.004	(0.031)	-0.009	(0.028)
12-13	LEP	0.001	(0.051)	0.015	(0.055)
12-13	Black	-0.037	(0.041)	-0.033	(0.037)
12-13	Hispanic	-0.002	(0.031)	-0.013	(0.029)
12-13	White	0.02	(0.039)	0.004	(0.034)
13-14	Gifted	0.008	(0.069)	0.032	(0.043)
13-14	Special Needs	0.018	(0.069)	0.036	(0.066)
13-14	FRL	0.044	(0.031)	0.001	(0.025)
13-14	LEP	-0.007	(0.055)	0	(0.052)
13-14	Black	0.024	(0.041)	-0.055	(0.036)
13-14	Hispanic	0.042	(0.032)	0.01	(0.026)
13-14	White	0.06	(0.04)	-0.008	(0.032)
14-15	Gifted	-0.058	(0.067)	0.038	(0.043)
14-15	Special Needs	0.18	(0.092) *	0.072	(0.085)
14-15	FRL	0.037	(0.03)	-0.029	(0.029)
14-15	LEP	-0.012	(0.05)	0.008	(0.051)
14-15	Black	0.104	(0.04) **	-0.043	(0.035)
14-15	Hispanic	0.022	(0.031)	-0.004	(0.028)
14-15	White	-0.013	(0.042)	0.007	(0.034)
15-16	Gifted	0.096	(0.083)	0.017	(0.040)
15-16	Special Needs	0.081	(0.081) *	0.062	(0.092)
15-16	FRL	0.063	(0.041)	-0.043	(0.030)
15-16	LEP	0.086	(0.061)	-0.096	(0.052)

Table 9 (Cont'd.)

Year	Subgroup	Algebra		Biology	
15-16	Black	0.078	(0.052)	-0.019	(0.043)
15-16	Hispanic	0.069	(0.043)	-0.024	(0.029)
15-16	White	0.1	(0.047) *	0.011	(0.036)
16-17	Gifted	0.186	(0.088) *	-0.033	(0.048)
16-17	Special Needs	-0.15	(0.149)	-0.202	(0.168)
16-17	FRL	0.104	(0.039) **	0.013	(0.033)
16-17	LEP	0.167	(0.058) **	0.059	(0.061)
16-17	Black	0.113	(0.049) *	0.044	(0.045)
16-17	Hispanic	0.123	(0.042) **	0.024	(0.033)
16-17	White	0.139	(0.052) **	0.005	(0.048)
17-18	Gifted	0.086	(0.085)	0.054	(0.048)
17-18	Special Needs	0.215	(0.173)	0.266	(0.166)
17-18	FRL	0.067	(0.043)	-0.056	(0.036)
17-18	LEP	0.072	(0.060)	-0.124	(0.056) **
17-18	Black	0.077	(0.057)	0.019	(0.048)
17-18	Hispanic	0.049	(0.043)	-0.075	(0.035) **
17-18	White	0.097	(0.057)	0.001	(0.040)

* $|t| > 1.96$, ** $|t| > 2.33$, *** $|t| > 3.09$

Teacher Assignment

We considered whether our results might be affected by the way teachers were selected to teach courses with high-stakes exams. Because of the very high stakes for schools and their personnel associated with these exams, one could expect principals to monitor past results carefully, and assign teachers with a good track record for raising student test scores to Algebra I and Biology (Dieterle, Guarino, Reckase, & Wooldridge, 2015).

We find evidence for such assignment bias, and it shows up in several ways. We constructed a two-stage model for the probability of being assigned to teach. The first stage of the model is Equation 1, which computes a value-added coefficient T for every teacher. The second stage is a binomial logistic regression model that computes the probability a teacher was assigned to teach Algebra I or Biology as a function of the value-added score in the course in the same school the year before. Thus, the probability of being assigned ($a = 1$) to a course given value-added score T in the previous year and certification pathway StdCert is

$$P(a|T, \text{StdCert}) = 1 / (1 + \exp[-(a_0 + a_T T + a_c \text{StdCert})]) \quad (5)$$

Here T is the value-added score we compute for each teacher normalized by the standard deviation of value-added scores and StdCert=1 corresponds to a teacher who came from a standard program. The coefficients of this model for Algebra I and Biology appear in Table 10.

The results are significant every year. For example, if an Algebra I teacher from a standard program in 2011-2012 had a value-added score 1.5 standard deviations above the mean, they had a 60% chance of returning to teach the course the next year as opposed to a 41% chance if their value-added score was 1.5 standard deviations below the mean. We also find that teachers from standard pathways were more likely to be reassigned to teach than those from alternative pathways,

after controlling for their value-added scores. Perhaps this is because department heads and principals saw something of value in their practice the scores did not capture.

Table 10

Probability of being reassigned to teach Algebra I or Biology in the same school given value-added score the previous year and certification pathway. Numbers in parentheses are standard uncertainties

Year	Discipline	$\mathbf{a_0}$			$\mathbf{a_T}$			$\mathbf{a_c}$		
2012-2013	Algebra I	-0.20	(0.05)	***	0.20	(0.03)	***	0.18	(0.07)	*
2013-2014	Algebra I	-0.08	(0.05)		0.25	(0.04)	***	0.11	(0.07)	
2014-2015	Algebra I	0.06	(0.05)		0.22	(0.03)	***	0.19	(0.07)	**
2015-2016	Algebra I	0.03	(0.05)		0.20	(0.04)	***	0.18	(0.07)	*
2016-2017	Algebra I	0.10	(0.05)	*	0.26	(0.04)	***	0.22	(0.07)	**
2017-2018	Algebra I	0.21	(0.04)	***	0.25	(0.04)	***	0.17	(0.07)	*
2012-2013	Biology	0.23	(0.04)	***	0.27	(0.03)	***	0.05	(0.07)	
2013-2014	Biology	0.25	(0.04)	***	0.28	(0.04)	***	0.20	(0.07)	**
2014-2015	Biology	0.31	(0.04)	***	0.21	(0.04)	***	0.08	(0.08)	
2015-2016	Biology	0.36	(0.04)	***	0.25	(0.04)	***	0.27	(0.08)	***
2016-2017	Biology	0.46	(0.04)	***	0.16	(0.04)	***	0.20	(0.08)	*
2017-2018	Biology	0.39	(0.04)	***	0.26	(0.04)	***	0.12	(0.08)	

* $p < .05$, ** $p < .01$, *** $p < .001$

School principals and department heads did not of course have access to the specific value-added scores we have computed, but they had in their possession all the raw data about student test scores that go into making them up and appear to have acted accordingly (Dieterle et al., 2015; Grissom, Loeb, & Nakashima, 2014).

While value-added score was the strongest single predictor we found of whether a teacher was assigned twice in a row to teach Algebra I or Biology, many characteristics of the teacher population changed. It is worth noting that the STAAR exams we use as a post-test were employed for the first time in the spring of 2012; at that time high school students were expected to take 15 exams in order to graduate. In the spring of 2013, the Texas legislature reduced the required exams from 15 to 5 and abolished all the STEM exams but Algebra I and Biology. One result was a dramatic shift in the experience distribution of teachers assigned to Algebra I and Biology. In 2011-2012, 35% of all Texas mathematics teachers had 0-5 years of experience (Ramsay, 2017a), and 39% of the Algebra I teachers had 0-5 years of experience. But by 2014-2015, when the percentage of Texas mathematics teachers with 0-5 years of experience was essentially unchanged at 37%, the percentage of Algebra I teachers with 0-5 years of experience dropped to 17%. As shown in Figure 8, the drop-in novice Algebra I teachers was accompanied by a rise in teachers with 10-20 years of experience. Placement of biology teachers was equally well predicted by value-added scores, and the distribution of Biology teachers changed in a very similar fashion, from a distribution characteristic of science teachers overall, to a distribution greatly weighted towards teachers with 10 to 20 years of experience.

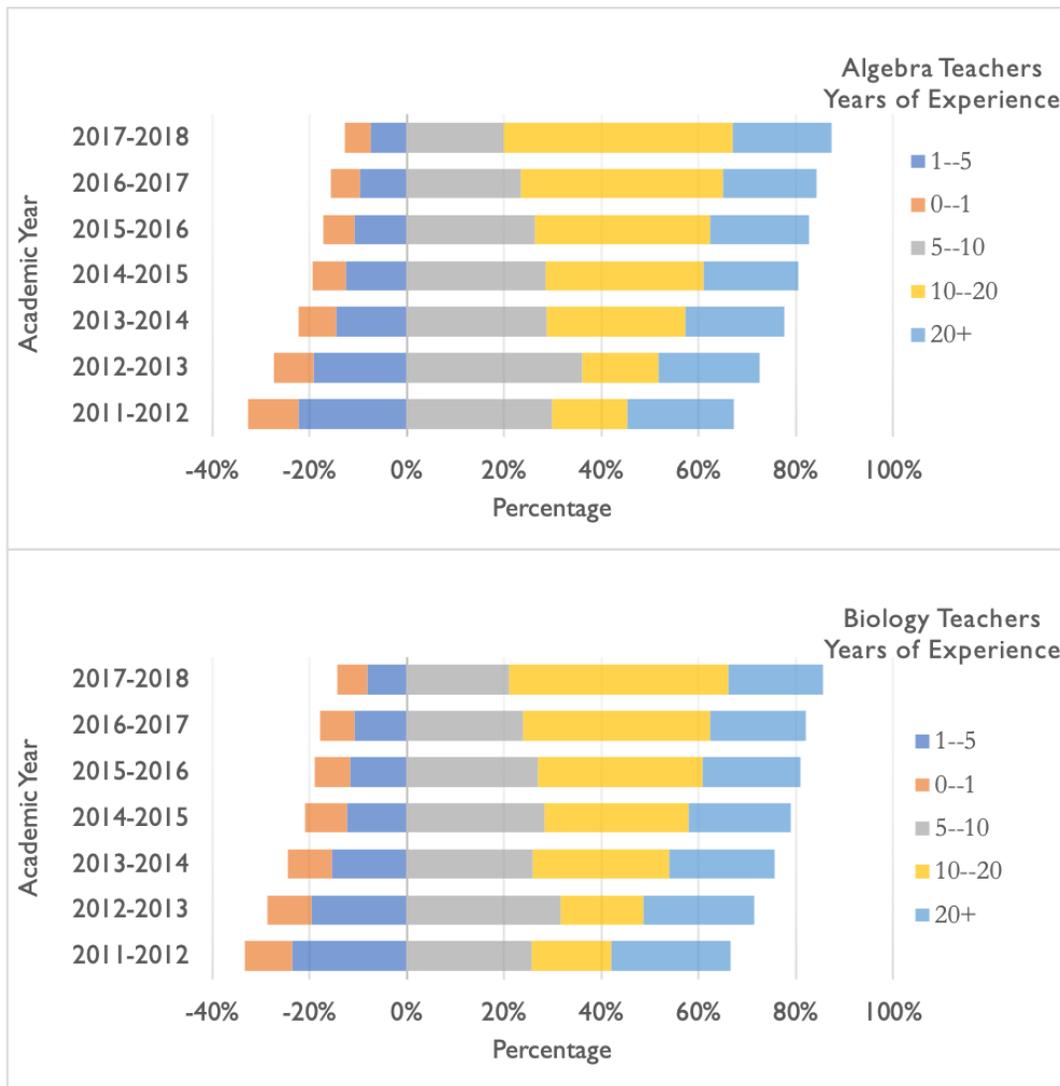


Figure 8. Distribution of teacher years of experience in Algebra I and Biology. Data from Texas Education Agency through Education Research Center

Discussion

Some previous studies have concluded that characteristics of teacher education are too small to detect or too small to matter in raising student test scores (Aaronson, Barrow, & Sander, 2007; Gordon et al., 2006; Harris & Sass, 2011; Rivkin et al., 2005; Staiger & Rockoff, 2010; von Hippel & Bellows, 2018). We found significant effects on the order of 0.03 to 0.05 standard deviations for ninth-grade students of Algebra I in favor of teachers with standard certification. Effects in Biology are weaker in the models with the strongest controls, although the occasional significant differences favor teachers with standard preparation. The pathway effects we find in Tables 5 through 9 are

small compared to the typical deviations between teachers and schools, although they are consistent with teacher pathway effects found in other studies (Boyd et al., 2009).

Whether gaining 0.03 in standard deviation units is an important educational difference merits additional discussion. It corresponds to a 25% greater chance of getting one more problem right on a 50-question exam, since the standard deviation on the exam is $0.16 \times 50 = 8$ questions, and 3% of that is a quarter of a question. This may seem too small to matter. However, if sustained over time, the magnitude of this effect is comparable to that of living in poverty. For example, in Tables 5 and 6, the coefficient for EcoDis, free and reduced lunch eligibility, is around -0.07 . As an additional illustration, Figure 9, employing methods of (Bendinelli & Marder, 2012) shows that if one groups students according to their mathematics scores and free/reduced lunch status in fourth grade, and then follows the students through 11th grade, the difference between the well-off and low-income students develops to around 0.06 in standard deviation units and it takes around three years to develop. That is, the difference in test score results due to having a math teacher from a standard program for a year is of the same order as the effect on test results over a year associated with living in poverty. One could conclude that the standardized tests are not very sensitive either to instruction (Popham, 2007; Stroup, 2009) or to poverty, but this does not mean the tests are completely incapable of detecting them.

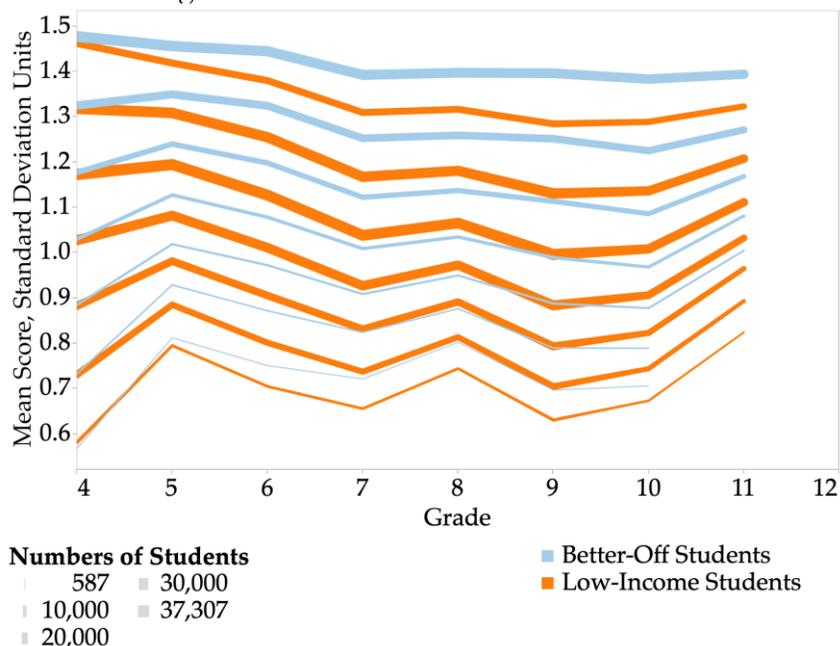


Figure 9. Mathematics exam averages in standard deviation units over time for cohorts of students from the Class of 2011 grouped both by their scores in 4th grade, and by their free/reduced lunch status. The graph tracks the students grade by grade over time
Data from Texas Education Agency through Education Research Center

Because the tests are not very sensitive to what we wish to measure, it takes a large number of teachers and students to arrive at reliable values. As seen in Tables 5 and 6, for a single teacher teaching multiple sections of the same class at the same time, the typical variation from one section of the class to another is around 0.18 standard deviations. We estimate that to find the effect of any particular type of teacher preparation pathway with uncertainty less than 0.03 standard deviations, one must average results from around 1000 teachers. We accomplished this by aggregating together

preparation programs in groups with similar practices rather than focusing on effects at the level of a single program. Such grouping also is a feature of the study of National Board certification in Cowan and Goldhaber (2016), and of pathways in New York City by Boyd et al. (2009).

Overall Algebra I teachers from standard certification pathways improved student test scores by 0.03 to 0.05 standard deviations. For subgroups including gifted students, students eligible for free and reduced lunch, and Black and Hispanic students, students of standard teachers gained 0.03 to 0.08 standard deviations. In Biology the evidence for positive effects for teachers from standard programs overall is not robust, although there are scattered positive results on the order of 0.03 standard deviations in Table 8 for teachers with up to 20 years of experience, and mainly positive but also some negative results in Table 9 for novice teachers. The columns in Table 6 for Models 1a and 1b also find significant test score gains for Biology students of teachers from standard programs. These are the models based on the assumption that the main reason students in some schools have higher scores is that the teachers are better.

We also find, as often found before, that there is more variation of student outcomes within teacher preparation pathways than between. This finding has been used in support of policies that reduce barriers for new people to enter teaching, but make it difficult for them to continue unless they can demonstrate favorable student outcomes (Gordon et al., 2006). While such policies might make sense in cases where there are more people wishing to become teachers than there are positions available, they are less justifiable for shortage areas such as secondary STEM. It is hard to imagine that young people or career changers will be attracted to secondary teaching by the prospect of having tenure and merit decisions made by value-added models. Newspaper accounts such as those of Bonner (2016) conclude that use of value-added scores for merit and promotion has been exacerbating teacher shortages. Teacher shortages may not directly impact high-stakes subjects such as Algebra I and Biology; schools have to staff them or face severe penalties. Shortages show up for subjects such as computer science where there are no high-stakes assessments, and where only a small fraction of high schools even offer a course (Guzdial, n.d.). It is tempting to consider policies that make it difficult for teachers with low value-added scores to continue teaching altogether, in hopes of capturing some of the 0.18 standard deviations advantage for the best teachers in Tables 5 and 6. However, reducing the stability of teaching careers will impact which individuals decide to enter teaching or settle instead on other careers. There is no assurance that secondary students will benefit in the end.

We remark in passing that our model estimates in Tables 5 and 6 contain interesting results beyond those directly pertaining to teacher certification pathway. For example, both in Algebra and Biology, a student identified as Economically Disadvantaged gains about 0.07 standard deviations less per year than one who is not. However the effect of poverty concentration is larger; in Algebra I, a class where none of the students is Economically Disadvantaged gains 0.09 standard deviations in comparison with a class where all the students are disadvantaged, and in Biology concentrated poverty can be responsible for a drop as large as 0.17 standard deviations. This observation should be taken into account in connection with studies of school choice involving lotteries (Angrist, Hull, Pathak, & Walters, 2017) because it estimates the effect students have on each other. If charter schools concentrate students with supportive families independent of ethnicity and economic need, then effects attributed to school organization and pedagogy might instead be due to influence of students on each other. Comparing charter school students to students denied admission by lottery back in the regular public schools does not address this problem.

The change we found in teacher assignment over time provides reason to worry about how high-stakes tests are being used. The tests were designed to measure student mastery of academic material. They are now being used to allow students to advance academically, to judge the performance of individual teachers, to judge the performance of schools and school administrators,

and finally to judge the programs that prepare the teachers. For test results to provide an unbiased estimate of preparation programs, principals would have to ignore teachers' track record when assigning them to classes with high stakes assessments, even when the future of both students and the administrators are at risk, and when administrators are constantly impressed with the importance of data-driven decision making (Houston, 2013). This is not realistic.

It has frequently been stated that that variance between classrooms within schools is larger than variance between schools (Nye, Konstantopoulos, & Hedges, 2004). Staiger & Rockoff (2010, p 103) conclude that "School leaders have very little ability to select effective teachers during the initial hiring process" and present as evidence "the fact that most of the variation in teacher effects occurs among teachers hired into the same school." Our results are different. In Tables 5 and 6, variation between schools is the largest random effect, followed by variation between teachers in schools, then followed by different classrooms of the same teacher. Thus, if we take into account both the variance between schools and the way we found teachers to be assigned, evidence indicates that principals hire teachers and assign them to classes based on information about their effectiveness.

This point may be particularly significant. The parallel teacher preparation system in Texas, spreading now to other states, represents an evolution of systems that enabled the growth of New York City Teaching Fellows and Teach for America (Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Laczko-Kerr & Berliner, 2002). These programs were intended to demonstrate that smart students carefully chosen and rapidly prepared could obtain better student results than conventionally prepared teachers (Clark et al., 2013; Decker et al., 2004). The scale at which this could be carried out has had limits because it depends upon philanthropic contributions of hundreds of millions of dollars per year (Education Week, 2016). Texas-style alternative certification does not have this limit because companies have established a viable business model where in-service candidates pay the program costs. The quality control provided by Teach for America selection criteria appears to be missing, but that impression is almost certainly mistaken. It is instructive to examine accountability reports for the educator preparation programs (Texas Education Agency, 2018). In 2016-2017 the two largest alternative programs had over 30,000 applicants, admitted about 14,000 of them, had over 55,000 listed as program participants without having completed, and of those around 7000 had positions in schools. This indicates that district human resource departments and school principals are quite selective in whom they take from the alternative programs. The selection is based upon interviews and upon a more extensive examination of a candidate's vita than is apparent in any state datasets. The hiring process plays a critical role in mediating between preparation programs and student outcomes.

Conclusions

Our results lead to nuanced policy prescriptions, and in particular we must distinguish between implications in Texas and implications for other states. In Texas, the differences we found for students in Algebra I and Biology because of their teachers' pathway do not justify disruptions that would come from an abrupt policy change impacting alternative certification programs. As alternative certification developed steadily over the last 20 years to become a major contributor to Texas' teacher supply, the schools learned how to hire and assign teachers to increase student test-score gains. In Algebra I there are many subgroups of students for whom the advantages of having a teacher from a standard university program are significant, but in Biology the effects are weaker. One must keep in mind because of the shortage of STEM teachers that it is difficult to justify reducing teachers from any pathway. Slightly increasing the scores of low-income students on

Algebra I exams but reducing the number able to take Physics or Chemistry at all would almost certainly be a very poor trade.

On the other hand, our results do not provide strong incentive for other states to follow Texas's lead in establishing a large for-profit alternative certification sector. As shown in Figures 3 and 4, the growth of alternative certification in Texas since the early 2000s has not led in the end to an increase in the production of STEM teachers. At best it has stemmed the decline. And if alternative certification were to grow much more rapidly in other states, as a disruptive innovation (Christensen, 1997), than it has in Texas, based on the for-profit companies' experiences but without corresponding ability in schools to select alternatively certified teachers wisely, the results might be quite different.

Caution is merited on all sides. Texas should be cautious about damaging the ecosystem that developed over the past two decades to prepare teachers. Other states should be cautious about rapidly introducing a parallel for-profit teacher preparation system that competes with universities. University faculty engaged in teacher preparation should be cautious about ignoring or dismissing alternative pathways with the capacity to prepare huge numbers of teachers in new ways, and be attentive to the ways that school hiring practices impact quality.

Around 700,000 undergraduates obtain STEM degrees from U.S. universities each year. This is an enormous pool; persuading just 1.5% more to obtain a teaching certificate along with their degree each year would add over 10,000 new STEM teachers. On balance, our evidence says the standard way of preparing teachers at universities in person is still the best. Teachers from standard pathways stay in teaching longer and their students learn more. If alternative certification does expand to ameliorate shortages, let this happen slowly and carefully. In view of the substantial extra time students spend preparing in standard pathways before full-time teaching, universities should consider what lessons can be learned from the alternative programs. At the same time, we urge renewed support for the preparation of STEM teachers through standard university pathways as the most efficient, scalable, and high-quality way to address the critical need for improved STEM education and to address the shortage of STEM teachers.

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>
- American Association for Employment in Education. (2016). *Educator Supply and Demand Report 2014-15*. Sycamore, Illinois.
- Amrein-Beardsley, A. (2009). Value-added tests: Buyer, be aware. *Educational Leadership*, 67(3), 38–42. <https://doi.org/10.1007/s00253-014-5773-9>
- Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132(2), 871–919. <https://doi.org/10.1093/qje/qjx001>
- Augustine, N. (2006). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington DC: The National Academies Press. Retrieved from <http://www.nap.edu/catalog/11463.html>
- 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>
- Bendinelli, A., & Marder, M. (2012). Visualization of longitudinal student data. *Physical Review Special Topics - Physics Education Research*, 8, 020119/1-15.
- Bobronnikov, E., Price, C., Gamse, B., Parsad, A., Roy, R., & Velez, M. (2014). Preliminary findings

- from the Noyce program evaluation and possible future directions. Retrieved July 17, 2019, from <http://app.nsfnoyce.org/wp-content/uploads/2014/07/Session-3.10-Ellen-Bobronikov-2014-Abt-Noyce-Prog-Eval-6-20-14-Final-for-posting.pdf>
- Bolker, B., Bates, D., Maechler, M., Walker, S., Christensen, R. H. B., Singmann, H., ... Green, P. (2016). Package lme4. Retrieved from <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Bonner, L. (2016). Enrollment plunges at UNC teacher prep programs. *The Charlotte Observer*. Retrieved from <http://www.charlotteobserver.com/news/local/education/article58311793.html>
- Boyd, D., Goldhaber, D., Lankford, H., & Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *Future of Children*, 17(1), 45–68. <https://doi.org/10.1353/foc.2007.0000>
- Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2012). Recruiting effective math teachers: Evidence from New York City. *American Educational Research Journal*, 49(6), 1008–1047. <https://doi.org/10.3102/0002831211434579>
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, 31, 416–440. <https://doi.org/10.3102/0162373709353129>
- 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>
- Christensen, C. M. (1997). *Innovator's Dilemma*. Boston, MA: Harvard Business Review Press.
- Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Puma, K. S. A. E. M., Warner, E., & Schmidt, S. (2013). The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs (NCEE 2013-4015). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/pubs/20134015/>
- College Board. (2016). AP program participation and performance data 2016. Retrieved from <https://research.collegeboard.org/programs/ap/data/participation/ap-2016>
- Constantine, J., Player, D., Silvaa, T., Grider, K. H. M., & Deke, J. (2009). An evaluation of teachers trained through different routes to certification, final report. Retrieved from ies.ed.gov/ncee/pubs/20094043/pdf/20094043.pdf
- Cowan, J., & Goldhaber, D. (2016). National Board certification and teacher effectiveness: Evidence from Washington state. *Journal of Research on Educational Effectiveness*. <https://doi.org/10.1080/19345747.2015.1099768>
- 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>
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*. <https://doi.org/10.14507/epaa.v13n42.2005>
- Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of Teach For America on students: Findings from a national evaluation (MPR 8792-750)*. Princeton NJ. Retrieved from <https://www.mathematica-mpr.com/-/media/publications/pdfs/teach.pdf>
- Dieterle, S., Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management*, 34, 32–58. Retrieved from <http://www.scopus.com/inward/record.url?scp=84916210067&partnerID=8YFLogxK>
- Duncan, A. (2010). Teacher preparation: Reforming the uncertain profession. Retrieved from

- <http://www.ed.gov/news/speeches/teacher-preparation-reforming-uncertain-profession>
Education Week. (2016). *Teach for America by the numbers*. Retrieved from
<https://www.edweek.org/ew/section/multimedia/teach-for-america-by-the-numbers.html>
- Etheredge, D. K. (2015). *Alternative Certification Teaching Programs in Texas: A Historical Analysis (Doctoral Dissertaion)*. University of North Texas. Retrieved from
<https://digital.library.unt.edu/ark:/67531/metadc799511/>
- Friedman, J. N., Rockoff, J. E., & Chetty, R. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.
<https://doi.org/10.3386/w19423>
- Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.
- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, *34*, 29–44.
<https://doi.org/10.1016/j.econedurev.2013.01.011>
- Gordon, R., Kane, T. J., & Staiger, D. O. (2006). Identifying effective teachers using performance on the job. Washington DC: Brookings Institution. Retrieved from
https://www.brookings.edu/wp-content/uploads/2016/06/200604hamilton_1.pdf
- Greenberg, J., McKee, A., & Walsh, K. (2013). Teacher prep review: A review of the nation’s teacher preparation programs. Retrieved from
http://www.nctq.org/dmsView/Teacher_Prep_Review_2013_Report
- Grimmett, P. P., & Young, J. C. (2012). *Teacher Certification and the Professional Status of Teaching in North America: The New Battleground for Public Education*. Charlotte, NC: Information Age Publishing.
- Grissom, J. A., Loeb, S., & Nakashima, N. A. (2014). Strategic involuntary teacher transfers and teacher performance: Examining equity and efficiency. *Journal of Policy Analysis and Management*, *33*(1), 112–140. <https://doi.org/10.1002/pam.21732>
- Grossman, P., & Loeb, S. (2008). *Alternative routes to teaching: Mapping the new landscape of teacher education*. Harvard Education Press.
- Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: a review of the recent empirical literature. *Review of Educational Research*, *76*(2), 173–208.
<https://doi.org/10.3102/00346543076002173>
- Guzdial, M. (n.d.). Why isn’t there more computer science in US high schools? [web log comment]. Retrieved July 1, 2017, from <http://cacm.acm.org/blogs/blog-cacm/156531-why-isnt-there-more-computer-science-in-u-s-high-schools/fulltext>
- Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, *61*, 280–288.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7–8), 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. https://doi.org/10.1162/EDFP_a_00134
- Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014). The effects of teacher entry portals on student achievement. *Journal of Teacher Education*, *65*(1), 7–23. <https://doi.org/10.1177/0022487113503871>
- Hess, F. M. (2002). Tear down this wall: The case for a radical overhaul of teacher certification. *Educational Horizons*, *80*, 169–183. Retrieved from <http://www.jstor.org/stable/42927125>
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960. <https://doi.org/10.1080/01621459.1986.10478354>

- Houston, P. (2013). *Using data to improve schools: What's working*. Arlington, VA. Retrieved from http://www.aasa.org/uploadedFiles/Policy_and_Advocacy/files/UsingDataToImproveSchools.pdf
- Imbens, G., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York, NY: Cambridge University Press.
- iTeach. (2019). Retrieved from <http://iteach.net>
- Jackson, C. K. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics*, 32, 645–684. <https://doi.org/10.1086/676017>
- Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6), 615–631. <https://doi.org/10.1016/j.econedurev.2007.05.005>
- Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation (NBER Working Paper No. 14607)*. Retrieved from <http://www.nber.org/papers/w14607>
- 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, 508–534. https://doi.org/10.1162/EDFP_a_00172
- Laczko-Kerr, I., & Berliner, D. (2002). The effectiveness of “Teach for America” and other under-certified teachers. *Education Policy Analysis Archives*, 10(37), 1–53. <https://doi.org/10.14507/epaa.v10n37.2002>
- Marder, M. (2012). Measuring Teacher Quality with Value-added Modeling. *Kappa Delta Pi Record*, 48, 156–161.
- Mellor, L., Lummus-Robinson, M., Brinson, V., & Dougherty, C. (2008). Linking teacher preparation programs to student achievement, unpublished report, available on request from corresponding author of this article.
- Morgan, S., & Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Second Ed.). New York, NY: Cambridge University Press.
- National Commission on Excellence in Education. (1983). A nation at risk. Retrieved from <http://www2.ed.gov/pubs/NatAtRisk/index.html>
- National Science Foundation. (2016). Robert Noyce Teacher Scholarship Program. Retrieved from <http://www.nsf.gov/pubs/2016/nsf16559/nsf16559.pdf>
- NRC. (2010). *Preparing Teachers: Building Evidence for Sound Policy*. Washington DC: The National Academies Press. Retrieved from <https://doi.org/10.17226/12882>
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26, 237–257. <https://doi.org/10.3102/01623737026003237>
- Pearl, J. (2009). *Causality* (Second Ed.). New York, NY: Cambridge University Press.
- 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>
- Ramsay, M. (2017a). *Experience of math and science teachers 2012-2016*. Retrieved from <http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=51539614334&libID=51539614334>
- Ramsay, M. (2017b). *Teacher retention by route 2012-2016*. Retrieved from <http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=51539614330&libID=51539614330>
- Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417–458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>

- Rosenbaum P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66, 688–701. <https://doi.org/10.1037/h0037350>
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. 2004 Fisher Lecture. *Journal of the American Statistical Association*, 100, 322–331. <https://doi.org/10.1198/016214504000001880>
- Sanders W.L., & R. J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement: Research progress report*. Knoxville, TN: University of Tennessee.
- Sass, T. R. (2011). Certification Requirements and teacher quality: A comparison of alternative routes to teaching. Working Paper 64. *National Center for Analysis of Longitudinal Data in Education Research*.
- Shiva Mungal, A. (2015). Hybridized teacher education programs in NYC: A missed opportunity? *Education Policy Analysis Archives*, 23, 1–28. <https://doi.org/10.14507/epaa.v23.2096>
- Shiva Mungal, A., Trujillo, T., & Scott, J. (2016). Teach for America, Relay graduate school, and the charter school networks: The making of a parallel education structure. *Education Policy Analysis Archives*. <https://doi.org/10.14507/epaa.24.2037>
- Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, 24(3), 97–117. <https://doi.org/10.1257/jep.24.3.97>
- Stroup, W. M. (2009). What it means for mathematics tests to be insensitive to instruction. In *Plenary Address at the Thirty-First Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*.
- 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, 1–36. <https://doi.org/10.14507/epaa.27.3696>
- Tatto, M. T., Savage, C., Liao, W., Marshall, S. L., Goldblatt, P., & Contreras, L. M. (2016). The emergence of high-stakes accountability policies in teacher preparation: An examination of the U.S. Department of Education’s proposed regulations. *Education Policy Analysis Archives*, 24(21), 1–57. <https://doi.org/10.14507/epaa.24.2322>
- Teachers of Tomorrow*. (2018). Retrieved from <http://teachersoftomorrow.org>
- Texas Education Agency. (2018). *2016-2017 Accountability System for Educator Preparation (ASEP) Annual Reports*. Retrieved from <https://tea.texas.gov/WorkArea/DownloadAsset.aspx?id=51539626053>
- Texas Education Agency. (2019). *Texas Education Code 21.045*. Retrieved from <https://statutes.capitol.texas.gov/Docs/ED/htm/ED.21.htm#21.045>
- Turner, H. M., Goodman, D., Adachi, E., Brite, J., & Decker, L. E. (2012). *Evaluation of Teach For America in Texas Schools*. San Antonio, TX.
- US Department of Education. (2016). Teacher preparation issues: Final regulations. Retrieved from <http://www2.ed.gov/documents/teaching/teacher-prep-final-regs.pdf>
- US Department of Education. (2018a). Title II Higher Education Act data files. Retrieved from <https://title2.ed.gov/Public/DataTools/Files.aspx>
- US Department of Education. (2018b). Title II Higher Education Act data tables. Retrieved from <https://title2.ed.gov/Public/DataTools/Tables.aspx>
- Vanderweele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York, NY: Oxford University Press.
- 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*.
<https://doi.org/10.1016/j.econedurev.2018.01.005>
- von Hippel, P. T., Osborne, C., Lincove, J., Bellow, L., & 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. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2506935
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89–122. <https://doi.org/10.3102/00346543073001089>

About the Authors

Michael Marder

The University of Texas at Austin

marder@mail.utexas.edu

Michael Marder is Professor of Physics and Codirector 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.

Bernard David

The University of Texas at Austin

bdavid@austin.utexas.edu

Bernard David is a doctoral candidate in STEM Education at The University of Texas at Austin, where he researches the effects of market-based education reforms upon student outcomes in STEM disciplines. Bernard was formerly a secondary science teacher in Washington, D.C.

Caitlin Hamrock

E3 Alliance

chamrock@e3alliance.org

Caitlin Hamrock is the Director of Research at E3 Alliance where she studies education in Central Texas. She and her team conduct research alongside and for an audience of practitioners, focusing on education trends from cradle to career.

education policy analysis archives

Volume 28 Number 27

February 24, 2020

ISSN 1068-2341



Readers are free to copy, display, distribute, and adapt this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, the changes are identified, and the same license applies to the derivative work. More details of this Creative Commons license are available at <https://creativecommons.org/licenses/by-sa/4.0/>. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education 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).

Please send errata notes to Audrey Amrein-Beardsley at audrey.beardsley@asu.edu

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Lead Editor: **Audrey Amrein-Beardsley** (Arizona State University)

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Melanie Bertrand, David Carlson, Lauren Harris, Eugene Judson, Mirka Koro-Ljungberg, Daniel Liou, Scott Marley, Molly Ott, Iveta Silova** (Arizona State University)

Cristina Alfaro

San Diego State University

Gary Anderson

New York University

Michael W. Apple

University of Wisconsin, Madison

Jeff Bale

University of Toronto, Canada

Aaron Bevanot SUNY Albany

David C. Berliner

Arizona State University

Henry Braun Boston College

Casey Cobb

University of Connecticut

Arnold Danzig

San Jose State University

Linda Darling-Hammond

Stanford University

Elizabeth H. DeBray

University of Georgia

David E. DeMatthews

University of Texas at Austin

Chad d'Entremont Rennie

Center for Education Research & Policy

John Diamond

University of Wisconsin, Madison

Matthew Di Carlo

Albert Shanker Institute

Sherman Dorn

Arizona State University

Michael J. Dumas

University of California, Berkeley

Kathy Escamilla

University of Colorado, Boulder

Yariv Feniger Ben-Gurion

University of the Negev

Melissa Lynn Freeman

Adams State College

Rachael Gabriel

University of Connecticut

Amy Garrett Dikkers University
of North Carolina, Wilmington

Gene V Glass

Arizona State University

Ronald Glass University of

California, Santa Cruz

Jacob P. K. Gross

University of Louisville

Eric M. Haas WestEd

Julian Vasquez Heilig California

State University, Sacramento

Kimberly Kappler Hewitt

University of North Carolina

Greensboro

Aimee Howley Ohio University

Steve Klees University of Maryland

Jaekyung Lee SUNY Buffalo

Jessica Nina Lester

Indiana University

Amanda E. Lewis University of

Illinois, Chicago

Chad R. Lochmiller Indiana

University

Christopher Lubienski Indiana

University

Sarah Lubienski Indiana

University

William J. Mathis

University of Colorado, Boulder

Michele S. Moses

University of Colorado, Boulder

Julianne Moss

Deakin University, Australia

Sharon Nichols

University of Texas, San Antonio

Eric Parsons

University of Missouri-Columbia

Amanda U. Potterton

University of Kentucky

Susan L. Robertson

Bristol University

Gloria M. Rodriguez

University of California, Davis

R. Anthony Rolle

University of Houston

A. G. Rud

Washington State University

Patricia Sánchez University of

University of Texas, San Antonio

Janelle Scott University of

California, Berkeley

Jack Schneider University of

Massachusetts Lowell

Noah Sobe Loyola University

Nelly P. Stromquist

University of Maryland

Benjamin Superfine

University of Illinois, Chicago

Adai Tefera

Virginia Commonwealth University

A. Chris Torres

Michigan State University

Tina Trujillo

University of California, Berkeley

Federico R. Waitoller

University of Illinois, Chicago

Larisa Warhol

University of Connecticut

John Weathers University of

Colorado, Colorado Springs

Kevin Welner

University of Colorado, Boulder

Terrence G. Wiley

Center for Applied Linguistics

John Willinsky

Stanford University

Jennifer R. Wolgemuth

University of South Florida

Kyo Yamashiro

Claremont Graduate University

Miri Yemini

Tel Aviv University, Israel

archivos analíticos de políticas educativas consejo editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editores Asociados: **Felicitas Acosta** (Universidad Nacional de General Sarmiento), **Armando Alcántara Santuario** (Universidad Nacional Autónoma de México), **Ignacio Barrenechea**, **Jason Beech** (Universidad de San Andrés), **Angelica Buendia**, (Metropolitan Autonomous University), **Alejandra Falabella** (Universidad Alberto Hurtado, Chile), **Veronica Gottau** (Universidad Torcuato Di Tella), **Carolina Guzmán-Valenzuela** (Universidad de Chile), **Cesar Lorenzo Rodríguez Uribe** (Universidad Marista de Guadalajara), **Antonio Luzon**, (Universidad de Granada), **María Teresa Martín Palomo** (University of Almería), **María Fernández Mellizo-Soto** (Universidad Complutense de Madrid), **Tiburcio Moreno** (Autonomous Metropolitan University-Cuajimalpa Unit), **José Luis Ramírez**, (Universidad de Sonora), **Paula Razquin**, **Axel Rivas** (Universidad de San Andrés), **Maria Veronica Santelices** (Pontificia Universidad Católica de Chile)

Claudio Almonacid

Universidad Metropolitana de Ciencias de la Educación, Chile

Miguel Ángel Arias Ortega

Universidad Autónoma de la Ciudad de México

Xavier Besalú Costa

Universitat de Girona, España

Xavier Bonal Sarro Universidad Autónoma de Barcelona, España

Antonio Bolívar Boitia

Universidad de Granada, España

José Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez

Instituto Nacional para la Evaluación de la Educación, México

Gabriela de la Cruz Flores

Universidad Nacional Autónoma de México

Marco Antonio Delgado Fuentes

Universidad Iberoamericana, México

Inés Dussel, DIE-CINVESTAV,

México

Pedro Flores Crespo Universidad

Iberoamericana, México

Ana María García de Fanelli

Centro de Estudios de Estado y Sociedad (CEDES) CONICET, Argentina

Juan Carlos González Faraco

Universidad de Huelva, España

María Clemente Linuesa

Universidad de Salamanca, España

Jaume Martínez Bonafé

Universitat de València, España

Alejandro Márquez Jiménez

Instituto de Investigaciones sobre la Universidad y la Educación, UNAM, México

María Guadalupe Olivier Tellez, Universidad Pedagógica Nacional, México

Miguel Pereyra Universidad de

Granada, España

Mónica Pini Universidad Nacional de San Martín, Argentina

Omar Orlando Pulido Chaves

Instituto para la Investigación Educativa y el Desarrollo Pedagógico (IDEP)

José Ignacio Rivas Flores

Universidad de Málaga, España

Miriam Rodríguez Vargas

Universidad Autónoma de Tamaulipas, México

José Gregorio Rodríguez

Universidad Nacional de Colombia, Colombia

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM, México

José Luis San Fabián Maroto

Universidad de Oviedo, España

Jurjo Torres Santomé, Universidad de la Coruña, España

Yengny Marisol Silva Laya

Universidad Iberoamericana, México

Ernesto Treviño Ronzón

Universidad Veracruzana, México

Ernesto Treviño Villarreal

Universidad Diego Portales Santiago, Chile

Antoni Verger Planells

Universidad Autónoma de Barcelona, España

Catalina Wainerman

Universidad de San Andrés, Argentina

Juan Carlos Yáñez Velazco

Universidad de Colima, México

arquivos analíticos de políticas educativas conselho editorial

Editor Consultor: **Gustavo E. Fischman** (Arizona State University)

Editoras Associadas: **Andréa Barbosa Gouveia** (Universidade Federal do Paraná), **Kaizo Iwakami Beltrao**, (Brazilian School of Public and Private Management - EBAPE/FGV), **Sheizi Calheira de Freitas** (Federal University of Bahia), **Maria Margarida Machado**, (Federal University of Goiás / Universidade Federal de Goiás), **Gilberto José Miranda**, (Universidade Federal de Uberlândia, Brazil), **Marcia Pletsch, Sandra Regina Sales** (Universidade Federal Rural do Rio de Janeiro)

Almerindo Afonso
Universidade do Minho
Portugal

Alexandre Fernandez Vaz
Universidade Federal de Santa
Catarina, Brasil

José Augusto Pacheco
Universidade do Minho, Portugal

Rosanna Maria Barros Sá
Universidade do Algarve
Portugal

Regina Célia Linhares Hostins
Universidade do Vale do Itajaí,
Brasil

Jane Paiva
Universidade do Estado do Rio de
Janeiro, Brasil

Maria Helena Bonilla
Universidade Federal da Bahia
Brasil

Alfredo Macedo Gomes
Universidade Federal de Pernambuco
Brasil

Paulo Alberto Santos Vieira
Universidade do Estado de Mato
Grosso, Brasil

Rosa Maria Bueno Fischer
Universidade Federal do Rio Grande
do Sul, Brasil

Jefferson Mainardes
Universidade Estadual de Ponta
Grossa, Brasil

Fabiany de Cássia Tavares Silva
Universidade Federal do Mato
Grosso do Sul, Brasil

Alice Casimiro Lopes
Universidade do Estado do Rio de
Janeiro, Brasil

Jader Janer Moreira Lopes
Universidade Federal Fluminense e
Universidade Federal de Juiz de Fora,
Brasil

António Teodoro
Universidade Lusófona
Portugal

Suzana Feldens Schwertner
Centro Universitário Univates
Brasil

Debora Nunes
Universidade Federal do Rio
Grande do Norte, Brasil

Lílian do Valle
Universidade do Estado do Rio de
Janeiro, Brasil

**Geovana Mendonça Lunardi
Mendes** Universidade do Estado de
Santa Catarina

Alda Junqueira Marin
Pontifícia Universidade Católica de
São Paulo, Brasil

Alfredo Veiga-Neto
Universidade Federal do Rio Grande
do Sul, Brasil

Flávia Miller Naethe Motta
Universidade Federal Rural do Rio de
Janeiro, Brasil

Dalila Andrade Oliveira
Universidade Federal de Minas
Gerais, Brasil