NATIONAL
CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

AIR

Can UTeach? Assessing the Relative Effectiveness of STEM Teachers

Ben Backes
Dan Goldhaber
Whitney Cade
Kate Sullivan
Melissa Dodson

WORKING PAPER 173 • December 2016

Can UTeach? Assessing the Relative Effectiveness of STEM Teachers

Ben Backes
American Institutes for Research

Dan Goldhaber
University of Washington Bothell/American Institutes for Research

Whitney Cade
American Institutes for Research

Kate Sullivan
American Institutes for Research

Melissa Dodson
American Institutes for Research

Contents

Acknowledgements…………………………………………………………………...ii

Abstracts……………………………………………………………………………...iii

I. Introduction……………………………………………………………………1

II. UTeach Overview and Prior Research………………………………………...5

III. Data……………………..…………………………………………………..9

IV. Methods..………………………………………………………....................12

V. Results………………………………………………………………………..17

VI. Conclusion…………………………………………………………………31

References………………………………………………………….………......33

Tables and Figures.…………………………………………………….........38

Acknowledgements

The authors gratefully acknowledge financial support from the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100116 to Southwest Educational Development Corporation, an Affiliate of American Institutes for Research. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Neither AIR nor the authors of this paper have financial interests that could be affected by findings from the paper.

We thank the Texas Education Research Center for providing data. The conclusions of this research do not necessarily reflect the opinion or official position of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

Can UTeach? Assessing the Relative Effectiveness of STEM Teachers

Ben Backes, Dan Goldhaber, Whitney Cade, Kate Sullivan, Melissa Dodson

CALDER Working Paper No. 173
December 2016

Abstract

UTeach is a well-known, university-based program designed to increase the number of high-quality science, technology, engineering, and mathematics (STEM) teachers in the workforce. The UTeach program was originally developed by faculty at the University of Texas at Austin but has rapidly spread and is now available at 44 universities in 21 states; it is expected to produce more than 9,000 math and science teachers by 2020. Despite substantial investment and rapid program diffusion, there is little evidence to date about the effectiveness of UTeach graduates. Using administrative data from the state of Texas, we measure UTeach impacts on student test scores in math and science in middle schools and high schools. We find that students taught by UTeach teachers perform significantly better on end-of-grade tests in math and end-of-course tests in math and science by 5% to 12% of a standard deviation on the test, depending on grade and subject. The effect is larger for the founding site at the University of Texas at Austin than for replication UTeach sites, with estimated upper bounds of additional months of learning for students taught by UTeach Austin graduates of 4.0 months in high school math and 5.7 months in high school science. Controlling for the selectivity of the undergraduate institution appears to explain the differential between Austin and replication UTeach sites, but not the overall difference between UTeach and non-UTeach teachers.

I. Introduction

A growing number of policymakers argue that for the U.S. to remain a worldwide economic and technological leader, it must do more to improve the quality of K–12 science, technology, engineering, and mathematics (STEM) education (e.g., Peterson et al., 2011). Given the growing body of evidence that educators are the most important determinant of student achievement outside of family and home influences (Borman & Dowling, 2008; Goldhaber, 2008; Hanushek et al., 2005; Heck, 2009; Ingersoll, 2001; Rice & Schwartz, 2008), it is no surprise that policymakers are focusing on teachers as a lever for improving STEM outcomes.[1] For instance, in fall 2009, President Obama asked his President's Council of Advisors on Science and Technology (PCAST) to draft a series of recommendations regarding the "most important actions that the administration could take to ensure that the United States is a leader in STEM education in the coming decades" (Holdren et al., 2010, p. vii). Among the council's findings was that math and science teachers are the "single most important factor in the K–12 education system…crucial to the strategy of preparing and inspiring students in STEM" (Holdren et al., 2010, p. 57).

The issue of STEM teachers is twofold. First, there are concerns about the quality of the existing STEM teacher workforce, particularly the prevalence of teachers without sufficient training in advanced subjects. For example, 61% of chemistry teachers and 67% of physics teachers do not hold a degree or certificate in those fields (Augustine, 2007). In addition,

[1] Differences between assignment to effective versus ineffective teachers have been found to have profound impacts on students' test scores and later life achievement (Chetty, Friedman, & Rockoff, 2014a, 2014b; Clotfelter, Ladd, & Vigdor, 2007; Hanushek & Rivkin, 2010; Kane & Staiger, 2008).

measures of overall science learning in the U.S. such as the National Assessment of Educational Progress (NAEP) find that science scores for high school students have shown no signs of improvement since 2009.[2] Second, there are longstanding issues of the quantity of STEM teachers and the difficulty of staffing STEM positions. Attracting STEM-trained individuals to the teaching workforce is particularly difficult because of higher paying jobs outside of teaching (West, 2013), and between 2000 and 2012, 20-30% of schools reported difficulty filling STEM vacancies (Cowan et al., 2016).

UTeach is a relatively new program that is designed to address these quality and quantity issues by "transforming the way universities prepare teachers"[3] with an approach that the Obama Administration believes "has shown strong results".[4] President Obama's educational initiatives, such as Race to the Top, Change the Equation, and 100Kin10, place STEM teacher preparation at the center of national education reform efforts, and UTeach is featured in each of these programs. More recently it was recognized by the Obama Administration as a national model for increasing the number of teachers filling hard-to-staff positions in STEM (*Ed Week*, 2010).[5]

The UTeach program was created in 1997 by faculty at the University of Texas at Austin (UT Austin) in an effort to streamline the process of earning a degree in math or science alongside a teaching credential while graduating in a timely manner. Because of the perceived success of the program, it has spread rapidly. In 2014, the National Math and Science Initiative

---

[2] Garcia Mathewson, Tara. "NAEP: 4th, 8th grade science scores are up; 12th grade scores are flat." *Education Dive.* 27 October, 2016.
[3] *National Math and Science Initiative.* http://www.noia.org/wp-content/uploads/2013/03/40100.pdf
[4] https://www.whitehouse.gov/sites/default/files/docs/stem_teachers_release_3-18-13_doc.pdf
[5] Robelen, Erik W. 20 January, 2010. Obama Unveils Projects to Bolster STEM Teaching. *Education Week. http://www.edweek.org/ew/articles/2010/01/20/18stem_ep-2.h29.html*

awarded a $22.5 million grant to continue the expansion of UTeach. Today it is available at 44 universities in 21 states, including state flagship universities such as UC Berkeley, the University of Florida, and West Virginia University, and is expected to produce more than 9,000 math and science teachers by 2020.

A selling point of UTeach is its approach to recruiting STEM majors to become teacher candidates while providing a pathway to have a STEM teaching credential in hand upon graduation. Students in UTeach take courses in their major along with classes for future teachers in a streamlined 4 year degree plan. Thus, UTeach has the potential to improve both the quantity and quality of the STEM teaching workforce by reducing barriers to entry to the teaching profession for STEM majors. As discussed in more detail below, the program touts three key elements that drive its success.[6] First, STEM majors are recruited as early as their freshman year; second, pedagogy courses are designed specifically for the program; and third, master and mentor teachers provide detailed guidance with early and intensive field experiences.

There are several reasons why UTeach teachers may be more effective at teaching STEM courses than the average teacher. First, by drawing from a pool consisting exclusively of math and science majors, the program potentially brings individuals with greater ability into the system than a typical teacher training program. On average, STEM majors who enter the teaching profession score about 100 SAT points higher than non-STEM majors (Goldhaber & Walch, 2013). Indeed, in our sample, the average replication site UTeach graduate scores 0.30-0.40 standard deviations higher on STEM certification exams than the average non-UTeach

[6] http://www.senate.state.tx.us/75r/Senate/commit/c530/meetings/082304/downloads/Charge4_MRankin.pdf

graduate. Second, subject-specific training may improve teacher performance in math and science at the secondary level, because some evidence suggests that greater math and science knowledge of teachers is associated with greater effectiveness at the high school level (Clotfelter, Ladd, & Vigdor, 2010; Goldhaber & Brewer, 1997; Goldhaber et al., 2016). Third, some UTeach-affiliated institutions such as UT Austin are more selective and thus may produce more effective teachers by this selection effect alone (Clotfelter et al., 2010).

Yet while selectivity of UTeach programs (UT Austin in particular) suggests that UTeach may be drawing more academically prepared individuals into the teacher workforce, there is reason to be cautious in thinking that this will necessarily result in significantly better student outcomes. For instance, the evidence on whether measures of college selectivity predicts teacher effectiveness is somewhat mixed (Harris & Sass, 2006).[7] Moreover, recent studies of traditional college and university-based teacher preparation programs (TPPs) suggests limited institutional level differences between TPPs (Goldhaber, Liddle, & Theobald, 2013; Koedel, Parsons, Podgursky, & Ehlert, 2015; von Hippel, Bellows, Osborne, Lincove, & Mills, 2016). In particular, Koedel et al. (2015) and von Hippel et al. (2016) emphasize that observed differences between TPPs are largely due to sampling variability rather than true differences between programs.[8]

[7] There is some evidence, for instance, that the positive findings for Teach For America teachers are partially explained by the selectivity of the Teach For America program (Xu et al., 2011); however, in their random assignment study, Clark et al. (2013) find that measures such as undergraduate selectivity and licensure scores do not explain any of the TFA effectiveness differential.

[8] Von Hippel et al. (2016) is especially relevant because, like this study, their sample consists of TPPs in Texas. However, while von Hippel et al. (2016) find little to no difference across TPPs in Texas, that does not guarantee that we will fail to find a UTeach effect because we have a much larger sample (four years of data compared to one), allowing for more precise estimates which could potentially mitigate the challenges imposed by sampling variability. In addition, one of our outcomes is performance on end-of-course exams, which are not included in the von Hippel et al. (2016) study and potentially allow for greater differentiation between TPPs due to the more advanced materials covered in EOC exams.

In this paper, we use administrative data covering all math and science teachers and their students in public secondary schools in Texas to assess whether UTeach-affiliated programs in Texas produce teachers that are more effective than the average non-UTeach teacher as measured by student performance on standardized assessments. In doing so, we provide rare estimates of variation in STEM teacher quality in secondary schools, where subject-specific training may be most important (see also Clotfelter et al., 2010; Goldhaber et al., 2016; Jackson, 2014; Xu et al., 2011). We find that, relative to non-UTeach teachers in the state, teachers trained at the UTeach founding site at Austin and the replication sites are more effective as measured by their ability to raise student test scores in math and science. Based on our estimates, the difference between graduates from UTeach replication sites and non-UTeach teachers in the effectiveness with which they teach high school math and science courses is similar to the difference between novice teachers and teachers with 10+ years of experience (about 0.05 standard deviations). For Austin UTeach graduates, the effect is twice as large. For the effectiveness of UTeach as measured by test scores in middle school math, the effect sizes for Austin and replication sites are 0.06 and 0.08 standard deviations, respectively, similar to the difference between a novice teacher and a teacher with 1–2 years of experience. Finally, while not the focus of the paper, we show descriptive evidence that the introduction of UTeach at partner universities has been associated with an increase in the number of STEM teachers produced.

The relatively large samples at UT Austin allow us to examine the extent to which the effectiveness of the UTeach program differs between this founding UTeach site and the replication sites at other universities in Texas (University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler). We do find evidence that UTeach effects are the largest for UT Austin in the more advanced high school subjects, but much of this differential can be explained by our measures of institutional selectivity, such as the SAT math scores of incoming students.

II. UTeach Overview and Prior Research

UTeach introduced an approach that was not typically seen in higher education. UTeach undergraduate students can obtain a teaching certificate and graduate with a math or science degree in 4 years while taking teaching classes designed specifically for math and science teachers. The program streamlines content and pedagogy coursework to combine STEM degrees with secondary certification without adding time or cost to 4-year degrees. This feature is used as a recruitment strategy to attract STEM majors who may view an extra year of education courses as a barrier and a deterrent to becoming certified to teach. According to internal data provided by UTeach, 55% of Austin UTeach graduates have graduated within 4 years, which is slightly higher than the university's overall average of 51-2%.[9] The UTeach model has also been scaled up and replicated nationally in more than 40 universities with the support of both public and private funding.[10] A description of what UTeach sees as the key characteristics of its program follows.

Recruitment and Selection Strategies. Undergraduate STEM majors are recruited into the UTeach program as early as their freshman year with no selection criteria for entry. The

[9] UTeach average obtained through personal correspondence with Michael Marder, Co-Director of UTeach, 9 December 2016. Overall Austin average obtained through 2007 and 2009 entering classes from IPEDS: http://nces.ed.gov/collegenavigator/?q=Austin&s=TX&pg=2&id=228778
[10] These funders include the National Math and Science Initiative (NMSI), Exxon Mobil, Howard Hughes Medical Institute (HHMI), and state and federal resources.

program offers compact and flexible degree plans that allow STEM majors to complete their degrees and certification in 4 years. In addition, UTeach provides interested undergraduates with two one-credit-hour, field-based courses free of charge, allowing undergraduates to try teaching before committing to completing the teaching option.[11] Based on their experiences with these courses, undergraduates either choose to continue in the program or self-select out of the teaching option early in their college career.

*Preparation and Support for Preservice Teachers***.** In addition to the content courses required for their major, students who continue with the UTeach program complete a set of STEM-specific pedagogy courses that emphasize inquiry-based instruction, connections between the theory of the pedagogy and the practice of teaching, the interconnections between math and science, and the importance of diverse historical and methodological perspectives.

*Highly Structured Field Experiences***.** STEM majors enrolled in UTeach courses engage in approximately 40 hours of structured field experiences before student teaching, all of which are supervised by master teachers and trained classroom mentor teachers. Before entering the student teaching semester, students are paired with local teachers who are trained to supervise and observe the UTeach student, offering multiple points for teacher candidates to reflect on their strengths and needs.

To guide program implementation at expansion sites, UTeach staff created the UTeach *Elements of Success,* a set of critical program components. UTeach has also developed

[11] According to UTeach staff, between 2008 and 2012, 21% (lowest year) and 38% (highest year) of students who enrolled in this initial free field-based course at Austin eventually graduated from UTeach.

resources and support materials for all operational and instructional aspects of the UTeach model. Universities replicating UTeach receive direct and individualized support, including access to the *UTeach Operations Manual,* UTeach curriculum, student work samples, support materials, and support events, including course workshops and retreats, topical webcasts, the annual UTeach Conference, and UTeach Open House.

Research on UTeach has been conducted primarily by UTeach-affiliated faculty and their graduate students, with little or no third-party evaluation. Peer-reviewed studies of UTeach programs in the last 10 years have employed descriptive or correlational designs that relied on surveys, interviews, observations, reviews of student and teacher discourse, and reviews of transcripts, lesson plans, and other artifacts. One category of studies focuses on preservice and in-service teachers' knowledge, use, and perceptions of the efficacy of specific instructional approaches learned in the UTeach courses (Confrey, Makar, & Kazak, 2004; Dickinson & Summers, 2010; Marshall & Young, 2006). Another set of studies explores preservice teachers' development and use of mathematical and statistical discourse (Ares, Stroup, & Schademan, 2009; Makar & Confrey, 2005). Finally, studies by Stroup, Hills, and Carmona (2011) and Marder and Walkington (2014) focus on exploring statistical approaches and methods for analyzing administrative and qualitative teacher and student data. The latter finds, for example, that the classroom observation protocol developed for UTeach is fairly weakly correlated with value-added scores.

III. Data

We use detailed student-level administrative data that link students in Texas to their teachers for 4 school years (2011–12 through 2014–15).[12] Texas has the second largest public K–12 enrollment in the United States, and large minority and disadvantaged student populations: about 52% of its students are Hispanic, 13% Black, and 30% White, and about 60% of students identified as economically disadvantaged.[13]

The student-level longitudinal data we use in the analysis contain math and science scores as primary outcome variables.[14] For math, these include both end-of-grade (EOG) and end-of-course (EOC) exams, with the bulk of EOC scores coming from Algebra I (76%) and the remainder Geometry (18%) and Algebra II (6%).[15] The share of Geometry and Algebra II students is relatively small because these tests were not administered in year during the time frame of our analyses. For science, we include EOC scores as outcome measures (80% Biology, 17% Chemistry, and 3% Physics), with Grade 8 EOG science used as a control for students in ninth grade.[16] Although estimating models with science test scores as an outcome variable in value-added models has not been as thoroughly vetted as math and reading, we perform a number of robustness checks, such as controlling for eighth-grade science scores for upper-

---

[12] We also use test score data from 2010–11 as prior test scores for regressions using the 2011–12 data for the outcome measure.
[13] tea.texas.gov/acctres/Enroll_2013-14.pdf
[14] In the 2010–11 school year, all students between grades 3 and 11 took the TAKS in math. However, with the introduction of EOC exams in 2011–12, the STAAR test began to be phased in with the 2011-12 school year. Since then, STAAR has been administered in Grades 3–8 with additional EOC exams in Algebra I and Biology. In 2012 and 2013 only, there were also EOC exams for Algebra II, Geometry, Chemistry, and Physics. For students taking an EOC exam, we consider their most recent test score to be their lagged test score. A Grade 8 science EOG test has been administered since 2012. See http://tea.texas.gov/student.assessment/staar/ for more information.
[15] We also use students' achievement on reading tests as a control variable.
[16] We do not use EOG science scores as an outcome because this test is not administered to sixth- or seventh-graders.

grade students and estimating UTeach effects on a sample of schools that do not seem to group students of similar ability by classroom. Our results for EOC science are similar when performing these checks. Finally, students' scale scores are standardized to have mean 0 and standard deviation 1 at the subject-grade-year-test level within the state.

In addition to standardized test scores, we observe a variety of student characteristics: race/ethnicity, gender, free- or reduced-price lunch (FRL) eligibility, gifted status, limited English proficiency (LEP) status, and disability status, which are used as covariates in our analyses. In addition, all students are linked to teachers based on course enrollment.[17]

Teacher personnel files contain information on teachers' experience, undergraduate and graduate institutions, demographics, and other supplemental background variables. These will likewise be used as covariates for some of the models in the analysis that follows. UTeach teachers are identified by combining where each teacher earned his or her degree, graduation year, and subject of teaching certificate. According to UTeach, it is possible to obtain undergraduate training to become a STEM-certified teacher from UTeach universities only by going through the UTeach program.

We begin by describing placement patterns of UTeach graduates by UTeach site, year, and subject; subjects include EOG math (Grades 6–8), EOC math (Algebra I, Algebra II, and Geometry), and EOC science (Biology, Chemistry, and Physics). A teacher is counted in a given sample if he or she teaches a student in that sample; thus, it is possible for teachers to appear in multiple samples. For example, if a teacher taught both eighth-grade math and Algebra I,

[17] Teachers of record in students' core math and science courses are linked to them for the analysis. Student observations linked to multiple teachers (e.g., due to coteaching, student mobility) are weighted in proportion to the amount of time spent with each teacher, based on available enrollment data (Hock & Isenberg, 2012).

then he or she would appear in both the EOG math and EOC math samples. We display placement patterns in this manner because we obtain estimates separately for these three samples.

Counts of UTeach teachers by campus and year are shown in Table 1. Two patterns are readily apparent. The first is that the number of teachers from UT Austin is relatively steady over time, whereas the number of teachers from the replication sites grows substantially. Austin's program dates back to 1997, well before the coverage of our data, suggesting the number of UTeach teachers graduating from Austin has more or less stabilized, whereas many of the other sites began their UTeach replication relatively recently and are ramping up graduate numbers. For example, although the number of UTeach graduates in EOC math classrooms from Austin was virtually the same in 2012 and 2015 (96 teachers vs. 93), the number increased from 15 to 41 at Houston and from 11 to 56 at North Texas during the same time frame. Second, Austin graduates are disproportionately concentrated in EOC subjects instead of the EOG grades. In 2015, only 21 Austin UTeach graduates taught students who took EOG tests, while 159 taught math and science EOC subjects. No other UTeach sites had such relatively few teachers in EOG.

While this paper is primarily interested in the question of whether UTeach programs produce more effective teachers than other TPPs, given the above discussion about the difficulty of staffing STEM positions, it is also important to consider whether the introduction of UTeach is also associated with an increase in the number of STEM teachers produced. Table 2 displays the number of teachers who appear in our STEM analysis sample by campus and graduation calendar year. At the two replication sites whose first UTeach graduates finished in

2010, we see substantial increases in the number of STEM teachers. At Houston, for example, the three years prior to UTeach saw 20, 21, and 13 teachers enter the workforce (average of 18 per year), while in the three years after, there was an increase to 29, 24, and 41 teachers (31 per year). At UNT, the increase was even larger, from 13, 19, and 15 teachers (16 per year) to 25, 41, and 44 teachers (37 per year). Thus, it does appear that UTeach is fulfilling its goal of recruiting more students to become STEM teachers.

Table 3 presents descriptive statistics of the students taught by UTeach and non-UTeach teachers included in the study. As with the counts discussed above, the samples used here are, of necessity, limited to grades and subjects in which standardized tests are administered to students. We keep the same groupings of teachers by EOG math, EOC math, and EOC science, while also splitting the sample into Austin UTeach, replication (non-Austin) UTeach, and non-UTeach teachers. For Austin graduates relative to non-UTeach graduates, although the racial composition of students taught is broadly similar, Austin graduates are less likely to teach FRL-eligible students and more likely to teach gifted students and students whose prior achievement was substantially higher. For prior math achievement, Austin UTeach graduates teach students who score 0.18 standard deviations higher in math EOG and EOC and 0.13 higher in EOC science subjects.[18] On the other hand, UTeach graduates from the replication

[18] Prior math scores do not average 0 across all students in the EOC math sample due to selection into advanced math courses (test scores are standardized among all students in the state). For example, if students progress to Geometry only if they have sufficiently high Algebra I scores, then the prior test scores for students taking Geometry will be higher than the mean of zero because of this selection mechanism. Results with EOC scores as the outcome are similar when including 8th grade EOG math and reading scores as additional controls.

sites are more likely to teach black students, LEP students, and students with lower prior achievement, compared to graduates from Austin.[19]

Teacher characteristics of UTeach and non-UTeach teachers are shown in Table 4. The typical UTeach teacher—whether from Austin or not—has fewer years of experience than the typical non-UTeach teacher. This is especially pronounced for the non-Austin group because, as discussed previously, these programs are relatively new: in all subjects, the percentage of teachers from replication sites in their first through third year of teaching is about 80%. In addition, the average selectivity of the undergraduate institution attended is higher for both UTeach samples (Austin and replication) —dramatically so for Austin. UTeach graduates are also more likely to be present in the certification database as STEM-certified; in fact, all UTeach teachers in our sample have a STEM certification because of the way we construct the UTeach variable, in which a teacher has to graduate from a UTeach campus with a STEM certification. In contrast, about half of EOG math teachers and one quarter of EOC math/science teachers are either not in the certification database or do not have a STEM-specific certification. Finally, math UTeach teachers are substantially more likely to be Hispanic than the average teacher in Texas.

IV. Methods

Our baseline analysis measures the difference in relative effectiveness between UTeach-trained teachers and comparison teachers who teach math or science to secondary students. Our approach follows similar studies of individual TPPs such as Teach For America and the New

[19] The patterns for school-level measures of student demographics and ability for schools with UTeach present vs. all schools are similar to the student-level patterns in Table 3.

York City Teaching Fellows Program (Boyd, Lankford, Loeb, & Wyckoff, 2006; Hansen, Backes, Brady, & Xu, 2015; Kane, Rockoff, & Staiger, 2008). We estimate the following equations:

$$y_{ist} = \beta_0 + \beta_1 y_{ist-1} + \beta_2 X_i + \beta_3 UTeach_i + \beta_4 T_j + \varepsilon_{ist}\ , (1)$$

where y_{ist} indicates the score on an EOG or EOC math or science exam (with separate regressions for each) for student *i* in school *s* in year t, y_{ist-1} a vector of cubic functions of prior year test scores in math and reading (and science, when science performance is the outcome measure), $UTeach_i$ is an indicator for whether student *i* was taught by a UTeach graduate in the tested subject, X_i contains a vector of student *i*'s characteristics, including race, gender, eligibility for FRL, special education status, and gifted status, and T_j a vector of controls for teacher characteristics, which in most models consists solely of experience.[20] Students with missing prior year scores are assigned a value of 0 for prior score with an additional control for missing prior year scores.[21] For students who took multiple tests in the prior year (e.g., Algebra I and EOG eighth-grade math), we use EOG scores as the measure of prior year achievement in regression models. In addition, ε_{ist} represents a randomly distributed error term.

The coefficient of interest, β_3, represents the average differential effectiveness of UTeach graduates relative to other teachers in the state. Both experimental work and nonexperimental tests suggest that controlling for prior test scores as in Equation (1) is sufficient for estimating teacher effects with little bias (Bacher-Hicks, Kane, & Staiger, 2014; Chetty et al., 2014a; Kane, McCaffrey, Miller, & Staiger, 2013; Kane & Staiger, 2008), with the

[20] In models where EOC tests are pooled together as an outcome variable (e.g., EOC math score as the outcome variable), we interact subject with all control variables to allow the association between these variables and the outcome to vary by test type.

[21] EOC results for students in higher grades (10 and above) are similar when controlling for eighth-grade EOG scores in addition to prior year EOC scores.

caveat that these studies do not examine high school teachers. Obtaining unbiased estimates of the effect of certain teacher characteristics on student achievement at the high school level is likely to be more challenging given the greater prevalence of specialty high schools and ability tracking (Jackson, 2014). We attempt to account for the potential that students with unobserved attributes correlated with test achievement are tracked into schools or classes (Jackson, 2014) by estimating models that include school or track effects. In these models, the effects of UTeach teachers are identified based on comparisons within the same school or track, where a track is defined to be all students within the same school who take the same set of courses in the same year.

Specifically, to account for school effects, we add school fixed effects to Equation (1) and thus compare UTeach graduates to other teachers within the same school rather than to all non-UTeach teachers in the state. Such a model ensures that UTeach trained teachers are compared to teachers in similar school settings. However, it could be the case that the UTeach effectiveness differential (if any) varies across school types. We experiment with an alternate model in which, instead of school fixed effects, we include school factors (e.g., percentage minority, percentage FRL-eligible) in the model and the relative effectiveness of UTeach teachers is allowed to vary along these measures through the use of interaction terms. Finally, in further specifications, we include track fixed effects to test whether estimates are sensitive to the sorting of students to tracks at the high school level (Jackson, 2014). In all analyses, standard errors are clustered at the school-cohort level (Chetty et al., 2014a).

Equation (1) yields an estimate of the average difference in achievement between students who were taught by UTeach graduates and those who were not. To investigate heterogeneity across programs, we decompose the UTeach coefficient into separate coefficients for different campuses (UT Austin, University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler) to assess whether different UTeach sites produce teachers of varying effectiveness:

$$y_{ist} = \beta_0 + \beta_1 y_{ist-1} + \beta_2 X_i + \sum_j \alpha^j \, Campus_i^j + \beta_4 T_j + \varepsilon_{ist} \text{ , (2)}$$

In Equation (2) above, α^j represents the coefficient estimate for each separate UTeach campus *j* and measuring the average difference between campus *j* and the average non-UTeach teacher. Variation in the α^j coefficients would indicate the extent to which graduates of different UTeach sites are differentially effective. It could be the case, for example, that graduates trained at the founding site, UT Austin, are more effective than graduates from the replication sites as a result of higher implementation fidelity (as the founding site) or because UT Austin is the most selective of the UTeach campuses. We test for this explicitly by grouping UTeach schools into Austin and replication (non-Austin) campuses.[22]

An important question is the extent to which UTeach effects are driven by the UTeach program itself rather than general institution or selectivity effects. For example, as of 2014, 41% of UTeach graduates nationwide were trained at UT Austin, ranked as a “highly competitive” university in Barron’s Profiles of American Colleges.[23] In their study of Teach for America (TFA), a selective teacher training program, Xu et al. (2011) found that a substantial portion of the greater effectiveness of TFA instructors relative to other teachers can be explained by TFA’s selection of candidates with better observable characteristics, such as graduating from a more

[22] In practice, there are not enough observations for each campus to obtain informative estimates for individual campuses based on Equation (2), so we group UTeach campuses into Austin and replication (non-Austin) sites for most of the analyses described below.
[23] UTeach program data: “UTeach and UTeach Expansion” from uteach-institute.org.

selective university and having higher licensure test (Praxis) scores. Thus, one may expect UTeach teachers from UT Austin to be the most effective, given that UT Austin is more selective than the other UTeach campuses. We thus perform a series of tests to investigate the question of program versus institution effects. First, we investigate whether selectivity can explain the UTeach effect by exploring the sensitivity of results to the addition of selectivity measures at the campus (SAT scores of incoming freshman students) and teacher (licensure scores) levels. Second, for replication sites, we compare the relative performance of UTeach graduates to those who graduated from those same institutions prior to the introduction of UTeach to see if the introduction of UTeach was associated with an increase in the performance of graduates from a given replication site.

We then turn to heterogeneity based on student demographic characteristics. In our sample, UTeach graduates at both Austin and replication sites are disproportionately likely to be Hispanic relative to other teachers in the state. Because there is some evidence that students taught by an own-race teacher tend to perform better on standardized tests (Dee, 2004; Goldhaber & Hansen, 2010), this could be an important avenue through which UTeach teachers have a differential impact, especially because Texas has a substantially higher share of minority students than the national average. We also investigate how the UTeach effect varies with student characteristics such as initial math proficiency, age/grade, gender, and FRL eligibility.

V. Results

Before displaying our estimates of UTeach effectiveness, we first note two ancillary findings that place our UTeach estimates in context. First, we estimate the dispersion of teacher effects by subject – sometimes referred to as the teacher "effect size" – by estimating models with teacher fixed effects and shrinking these estimates using an Empirical Bayes procedure. The standard deviations for these teacher fixed effects for each subject are as follows: EOG math 0.22, EOC math 0.44, EOC science 0.29, EOG reading 0.14, and EOC reading 0.27, with the results for the math and science subjects here being consistent with Goldhaber et al.'s (2016) estimates from the state of Washington. Second, using Lipsey et al.'s (2012) estimates of annual learning by subject and grade, taking the most conservative (i.e., largest standard deviation) estimates for translating test score gains to months of learning, the average student gains 0.32 standard deviations per year in middle school math, 0.25 standard deviations per year in high school math, and 0.22 standard deviations per year in high school science.[24]

A. Baseline Findings

We begin by displaying the results of our basic estimating equation for math subjects in Table 5. Each math test—EOG math, Algebra I, Algebra II, and Geometry—is used as an outcome variable in six specifications, with each column representing the results from a different specification.[25] The first three columns show results with no fixed effects but different choices for controls, and the last three columns contain various levels of fixed effects: school,

[24] In all months of learning calculations, we use a 9 month school year.
[25] Appendix Table 1 shows selected coefficient estimates from column 2 of the three most common subjects: EOG math, Algebra I, and Biology. For the sake of space we do not report the coefficients in the main tables of the paper.

school-track, and student. Results are mostly consistent across the different tests: without fixed effects, the typical student in a UTeach classroom scores 4% to 10% of a standard deviation higher, depending on the subject and model.

When adding fixed effects, there is a reduction in effect size to 0% to 6% of a standard deviation in the fixed effect models. It is likely that the shrinkage of the UTeach effect size when adding school fixed effects (whether at the school level or school-track level) is due to teachers of similar effectiveness sorting into the same schools. Thus, for example, graduates from Austin sort into schools where comparison teachers are more effective on average. This is the opposite of the result sometimes observed in TFA evaluations, in which the TFA effect increases with school fixed effects because the comparison teachers in the disadvantaged schools where TFA corps members are placed tend to be below average (e.g., Hansen et al., 2015). We discuss the interpretation of the models with and without fixed effects in further detail below.

We display the same models for the science subjects (Biology, Chemistry, and Physics) in Table 6 with similar results, albeit consistently larger in the models without school fixed effects. Although the ordinary least squares (OLS) results for Physics in Table 6 are large, this constitutes a very small share of the sample, and this reflected by the large standard errors. Because results are generally consistent across EOC tests within the same subject, for the remainder of the results section we group results into the three categories used in the summary statistics tables: EOG math, EOC math, and EOC science.[26] For EOC math, observations largely come from Algebra I (75% of observations), whereas for EOC science,

[26] In these specifications, we interact coefficients by test type to allow their association with the outcome to vary by test type. For example, in Table 5, in results not shown, prior math scores are differentially predictive depending on the outcome test under examination.

Biology (80%) makes up the bulk of observations. The vast majority of science EOC observations come from Biology and Chemistry, where results are very similar for the two subjects.

The effect sizes shown in Tables 5 and 6 are large relative to the returns to teacher experience and to how much a typical student learns in a year. In EOG math, column 2's 0.08 standard deviations is similar to the difference between a teacher with 0 years of experience and one with 3–5 years of experience (see Appendix Table 1). In a 9 month school year, 0.08 standard deviations translates to 2.25 months of additional learning. For Algebra I and Biology, the UTeach effect is larger than the difference between teachers with more than 10 years of experience and teachers with 0 years of experience and is equivalent to 3.6 months of learning in math and 4.9 months in science. Although our estimates of returns to experience may appear small (0.04 for teachers with more than 10 years of experience in Algebra I and Biology), they are not substantially different than other estimates of the returns to experience using teachers in high school subjects (e.g., Clotfelter et al., 2010, Xu et al., 2011).

B. Accounting for Teacher Characteristics

The first three columns of Tables 5 and 6 control for different teacher characteristics. Column 1 includes no teacher characteristics. Column 2 adds teacher experience. In six of the seven tests used as outcomes, including teacher experience increases the magnitude of the UTeach estimate. This is not surprising because UTeach teachers have less experience on average than other teachers, so accounting for this experience differential makes UTeach performance look relatively stronger.

One potential explanation for the effectiveness of UTeach teachers relative to comparison teachers is that all UTeach teachers are STEM-certified, while about half of EOG teachers and one quarter of EOC teachers are either certified in a non-STEM-specific field or absent from the state's certification database. Although the literature presents mixed evidence about the relationship between certification and student achievement (e.g., Rockoff, Jacob, Kane, & Staiger, 2011), we nevertheless explore how the addition of control for STEM certification would affect our UTeach estimates. Results are shown in column 3 of Tables 5 and 6. Interestingly, the STEM certification coefficient for Algebra I in column 3 (0.128; not shown but available from authors) is very similar to the corresponding result for Algebra and Geometry teachers in North Carolina (0.127) from Clotfelter et al. (2010). In addition, the association between Biology score and STEM certification in our study (0.051) is also similar to Clotfelter et al. (2010)'s result (0.029).[27] Returning to UTeach estimates, relative to columns 1 and 2 of Tables 5 and 6, the coefficients for UTeach in the two math subjects with the largest number of teachers—EOG math and Algebra I—are modestly attenuated when including STEM certification as a control. However, even with these controls, the coefficients remain statistically significant and positive. Thus, providing a guided route to certification may explain a small portion of the effectiveness of UTeach in the classroom but UTeach teachers remain more effective than the average teacher in Texas even after accounting for STEM certification.

[27] We do not show the coefficients in the tables for sake of brevity. They are as follows: EOG M 0.025, Algebra I 0.128, Geometry 0.030, Algebra II -0.035 (not significant), Biology 0.051, Chemistry -0.028, Physics -0.037 (not significant). All are statistically significant unless otherwise indicated.

C. Results for Austin vs. Other UTeach Sites

Next we present separate results for the two types of UTeach campuses: the original site at Austin and all other replication sites at universities in Texas, as shown in Table 7. The results indicate that achievement gains associated with UTeach classrooms in the EOC subjects in math and science (panels 2 and 3) are stronger for the UT Austin graduates than for graduates from other UTeach sites. For Austin, the EOC achievement boost is estimated to be 11% of a standard deviation in math (4.0 months of additional learning) and 14% in science (5.7 months) without school fixed effects or controls for STEM certification (column 2), and 2% to 7% with school fixed effects (column 4). For the non-Austin replication sites, the corresponding gains are 7% in EOC math (2.2 months of learning) and 6% in EOC science (2.5 months) without fixed effects, and close to zero with fixed effects, with the fixed-effects estimates for the non-Austin campuses not statistically significant. For EOG math (panel 1), results are similar across the Austin and non-Austin campuses (they are not statistically significantly different from each other), although it is only for non-Austin campuses that the coefficient is statistically significantly greater than zero with school fixed effects. Our EOG math point estimates for Austin are 0.06 (1.7 months of learning) and 0.08 for replication sites (2.25 months). When disaggregating non-Austin sites by campus, shown in Appendix Table 2, standard errors become too large to be confident in the findings, although the point estimates for UT Dallas are very large and positive.

The results in Tables 5–7 indicate that the decision whether to include school fixed effects leads to meaningful differences in the magnitude and, at times, statistical significance of the estimates. In models without school fixed effects—thus comparing UTeach graduates to all other teachers in the state without accounting for the school a teacher is assigned to—both the founding site and replication sites perform better than the average non-UTeach teacher in all subjects (although the point estimate for replication sites' EOC science is not statistically significant). However, when comparing UTeach graduates from replication sites to other teachers in their schools, replication site graduates fare about the same as non-UTeach teachers in EOC math and science. As discussed in Goldhaber et al. (2013), it is difficult to know which set of results to privilege. On one hand, the fixed effects model is theoretically attractive because it removes potential time-invariant biasing factors such as principal or neighborhood quality. On the other hand, if UTeach graduates are sorted into schools where both UTeach and comparison teachers are truly more effective, then the addition of school fixed effects obscures the true effectiveness of UTeach teachers by restricting the comparison group to teachers who are more effective than the average teacher in the state. Smaller UTeach coefficient estimates when school fixed effects are introduced is the same pattern observed in Goldhaber et al. (2013), where TPP program estimates are attenuated in fixed effects specifications. In the remainder of this paper, we display estimates both with and without fixed effects in the interest of transparency but largely focus the discussion on the models without fixed effects.

D. Exploring Heterogeneity Across UTeach Sites

In this section, we investigate potential explanations for why the observed achievement gains associated with being in a UTeach classroom are stronger for the Austin graduates than for graduates from the replication sites. We run through three potential explanations that could plausibly explain the Austin-replication site differential. First, we investigate institution-level selectivity measures (SAT scores of incoming students); second, we test individual-level ability measures (teacher certification scores); third, we investigate whether the Austin effect is driven by its program having been in place longer by looking for evidence of the replication sites improving over time. Before proceeding, we emphasize that we cannot cleanly distinguish selection effects from training effects. Thus, when we measure an “Austin effect,” we are measuring the result of the combined process of the sorting of students into Austin, the sorting of Austin students into training, and the training itself.

The first two potential explanations pertain to selection. We experiment with two ways of accounting for selection and then discuss them together. First we control for the selectivity of the undergraduate institution attended, as measured by the 75th percentile of math and reading SAT scores of incoming freshmen. Specifically, we add dummy variables for each selectivity quintile for both math and reading SAT scores to allow for potential nonlinearities between selectivity and effectiveness, with an additional dummy indicating missing scores (either because the undergraduate institution does not require SAT scores for admission or because we cannot identify in the state database which institution a teacher attended).[28] Although SAT scores are only one dimension by which students are selected into college, they are nevertheless useful for sorting colleges into broad selectivity tiers (Hoxby, 2009). Results are shown in Table 8. The coefficients for Austin are now attenuated in all models but remain statistically significant in all models where EOC scores are an outcome and fixed effects are not included. Where the Austin EOC effect size is 11% of a standard deviation in EOC math and 14%

[28] We experimented with breaking the selectivity groups into finer groupings (e.g., deciles rather than quintiles) but heaping at certain SAT score cutoffs makes creating these smaller groups of equal sizes difficult.

in EOC science based on models that do not control for selectivity, the effects become 8% and 8%, respectively, with the institution-level selectivity controls. In addition, the difference in coefficients between the Austin and non-Austin UTeach sites is no longer statistically significant in the EOC models in column 2, which may be suggestive evidence that the Austin effect relative to other UTeach sites could be driven in part either by the selection of more talented prospective teachers or the advantage in institutional resources that Austin enjoys relative to other institutions in Texas.

Is the effectiveness of Austin teachers relative to non-UTeach teachers purely due to admitting teaching candidates from a selective institution? To investigate, we restrict the sample to teachers who attended schools whose selection criteria are similar to Austin's, as measured by incoming students' SAT math scores. Specifically, we take the 75th percentile SAT math score for each teacher's undergraduate institution and restrict the sample to teachers whose undergraduate institution places them in the top 10% of teachers by this measure. Of these top 10–percentile teachers, Austin lies directly in the middle (i.e., Austin graduates are at the 95th percentile by undergraduate selectivity among all teachers in Texas). In results not shown, we find that Austin graduates fare better in this comparison with graduates of other highly selective institutions by 6% to 7% of a standard deviation in all subjects without fixed effects. However, this comparison may be somewhat misleading because the bulk of comparison teachers from selective institutions are drawn from state institutions because of a lack of comparably selective institutions within Texas. Like TPPs in general (Boyd et al., 2005), Austin UTeach graduates are more likely to teach in schools close to where they were prepared, whereas teachers drawn from out of state are teaching far from their training institutions, by definition.[29] Goldhaber et al. (2013) provide evidence that math teachers are more effective when teaching within 10–25 miles of their training program. Thus, it is unclear whether comparing Austin graduates to graduates of other highly selective institutions is measuring a UTeach Austin effect, a proximity effect, or some combination of both.[30]

As a second way to account for selectivity, we add controls for teacher licensure scores in STEM subjects as a measure of individual aptitude.[31] Specifically each teacher's first recorded score on a STEM certification exam and standardize these scores to have mean 0, standard deviation 1 within each subject–year combination and include them as an additional control. Results are shown in Table 9. Relative to our main estimates in Table 7, coefficients are attenuated to a small degree, but even though Austin graduates score substantially above average on certification exams, estimates of their effectiveness are large even after controlling for certification scores. As an auxiliary point of interest, we find that a teacher's certification scores are more predictive of student performance in EOC subject tests than in EOG tests.[32] As with the STEM certification regressions in column 3 of Tables 5 and 6, despite a different sample and different model, our coefficient for certification score in EOC math is similar to

[29] For example, about 20% of Austin UTeach graduates teach in Austin ISD even though Austin ISD comprises about 1% of total observations in the data.
[30] We investigate further by comparing Austin graduates to graduates of Texas A&M, the other selective public Texas university. In results available from the authors, Austin graduates tend to perform better than Texas A&M graduates on EOC subjects. However, as with the results discussed above, this is not a perfect comparison because Austin is more selective than Texas A&M: Austin's 25th and 75th percentile SAT math scores are 600 and 720, compared to 550 and 670 for Texas A&M.
[31] For math teachers, the standardized scores are 0.71 for Austin, 0.39 for replication sites, and 0.08 for non-UTeach teachers. For science, the corresponding scores are 0.72 for Austin, 0.46 for replication sites, and 0.06 for non-UTeach. Overall mean scores are not zero because the standardization process happens before linking to students, so if teachers who score poorly are less likely to enter or persist in teaching, then the analysis sample will have teachers who score above average. This is the same standardization procedure followed by Clotfelter et al. (2010).
[32] Coefficients on certification scores: math EOG 0.024, math EOC 0.037, science 0.041 (all statistically significant).

Clotfelter et al. (2010)'s estimate for Algebra I, whereas our coefficient for EOC science is somewhat larger than their estimate for Biology.

To probe whether UT Austin graduates' effectiveness can be fully explained by all observable characteristics in our data, we include the undergraduate institution selectivity scores used in Table 8, the certification scores used in Table 9, and additional controls for certification route. Specifically, we include separate dummy variables for whether a teacher obtained a certification through an alternate route and for whether a teacher is missing in the certification database. Results are shown in Table 10, with the UTeach coefficients representing UTeach estimates relative to other teachers who come from standard university-based programs. In column 3, which also controls for whether the teacher observed holds a STEM credential, only two of the six coefficients are statistically significant at the 95% level. However, even after accounting for certification route, institutional selectivity, certification scores, and possession of a STEM credential, students of Austin graduates score 5% of a standard deviation higher in EOC math (not statistically significant) and 7% higher in EOC science.

Another possible reason the estimated effectiveness of Austin graduates might be higher than other sites' is that the Austin program has existed for substantially longer. Thus, it is possible that, once other programs have been in place as long as Austin's, they too will be similarly strong. To investigate whether the other programs show evidence of improving over time, we include program * year interaction terms, with results displayed in Table 11. Results for the replication sites in the omitted year, 2012, should not be taken seriously due to very few observations in those years (see Table 1). However, looking at changes from 2013 to 2014 and from 2014 to 2015, there is some weak evidence of improvement over time in math. In EOG math, for example, the relative coefficient increases from -0.10 in 2013 to -0.07 in 2014 to -0.03 in 2015. However, these are imprecisely estimated and the interaction terms are generally not statistically significantly distinguishable from each other. This pattern is worth continuing to explore as more data become available in the future.

E. Further Exploring Program vs. Institution Effects

In this section, we investigate two additional pieces of evidence in an effort to separate whether the effectiveness of UTeach graduates is due to the UTeach program itself or the fact that UTeach teachers are coming from a select set of TEP institutions. Specifically, for replication sites, we compare the relative performance of UTeach graduates to graduates of the same institutions before UTeach was founded. In addition, we estimate the performance of non-UTeach (i.e., non-STEM) graduates of UTeach-affiliated institutions.

We first estimate the effectiveness of STEM teachers who graduated from UTeach replication sites before UTeach was adopted by those sites; unfortunately, we are not able to do the same for Austin graduates prior to UTeach because UTeach has been at Austin for nearly 20 years. Specifically, we estimate the coefficients of fixed effects that identify whether a teacher graduated from UTeach Austin, from UTeach at a replication site, or from a replication site in the period before UTeach was introduced. Results are shown in Table 12. A comparison of the “Other UTeach” to the “Replication Pre” coefficients reveals that replication site UTeach graduates (“Other UTeach”) are estimated to be consistently more effective than other STEM teachers who graduated from the same institutions before UTeach was implemented (“Replication Pre”) in all subjects in the models without fixed effects. In EOG math, where

UTeach replication site graduates are concentrated and have the largest point estimates, the pre-UTeach replication site graduates are no more effective at teaching EOG math than the average non-UTeach teacher. This pattern would be consistent with the introduction of UTeach at a campus improving the average quality of STEM teachers who graduate from that university.

As an additional test, we replicate Table 12 on a sample of non-STEM (i.e., non-UTeach) teachers from UTeach-affiliated universities. If the improvement in teacher quality at UTeach replication sites associated with the introduction of UTeach in Table 12 were driven by general university-specific trends rather than UTeach itself, one might expect teachers who have graduated from these UTeach-affiliated institutions in the time since UTeach was introduced ("Other UTeach") to be larger than the pre-UTeach graduates ("Replication Pre"). Results are shown in Table 13. For EOG reading performance of the replication site graduates, *where non-STEM graduates are not part of the UTeach program at any campus*, we do not see evidence that teachers who graduated from replication sites are more effective in the time period during which UTeach has been operating. Thus, while there is improvement in EOC reading in the replication sites after the introduction of UTeach in some models, it is not the case that non-STEM teachers at replication sites were uniformly better across all subjects in the more recent UTeach period.

Turning to the Austin coefficient for non-STEM teachers in Table 13, relative to our main results for STEM teachers in Table 7, results are strikingly similar, where results for EOG math are slightly larger than EOG reading, and results for EOC math are identical to the results for

EOC reading.[33] However, the interpretation of STEM vs. non-STEM results at Austin is complicated by Austin housing the only non-STEM UTeach site: UTeach-Liberal Arts at UT Austin.[34] Thus, Austin is the only campus where a comparison of STEM to non-STEM teachers is not a clean UTeach versus non-UTeach comparison.

F. Exploring Teacher–Student and Teacher–Class Match Effects

The website for the UTeach site at Austin touts its approach to diversity, in both its training[35] and in producing a more diverse set of teachers than other programs.[36] To investigate whether UTeach graduates are differentially effective at teaching minority students, we include interaction terms between UTeach and student demographics. Results are not shown (for the sake of brevity) but there are few clear patterns with the exception of Black students appearing to perform relatively worse when taught by Austin UTeach graduates and FRL-eligible students performing better when taught by replication site graduates.[37] To explore weather UTeach teachers are differentially effective at teaching high-ability classrooms, we compare UTeach teachers in the top 20% of classrooms (as defined by prior math scores) to other teachers in top classrooms. In results available from the authors, we find that students in these high-ability

[33] As with Table 12, we do not estimate pre-UTeach Austin because of the length of time that UTeach has been at Austin.
[34] For more information, see https://liberalarts.utexas.edu/uteach/.
[35] https://institute.uteach.utexas.edu/sessions/three-university-perspectives-weaving-equity-diversity-and-current-issues-classroom
[36] http://www.sheeo.org/sites/default/files/Weds%201115%20Marcus%20Lingenfelter%202013UTeachGradReportFinal120204.pdf
[37] When interacting UTeach with school characteristics, we find suggestive evidence that increases in the school share of black students is associated with reductions in the effectiveness of Austin grads, but these estimates are very imprecise. These results are available from the authors upon request.

classrooms disproportionately benefit from having either an Austin or UTeach replication site graduate in EOC math subjects, with point estimates on the order of 0.20–0.26.

G. Additional Tests of Robustness

To this point, our sample has included all teachers regardless of the amount of time elapsed since they completed their training. If it were the case that TPP effects are strongest for teachers who recently graduated and entered the workforce and that they dissipate as time elapses, as determined by Goldhaber et al. (2013), one would expect the UTeach effects estimated above to be an attenuated version of what might be found on a sample of novice teachers. We ran our basic models on a sample of teachers in their first, second, or third year of teaching and display the results in Table 14. Comparing column 2 of Table 7 to column 2 of Table 14, three of the four EOC coefficients are larger when the sample is restricted to novice teachers, as in Table 14. For example, for EOC math, the coefficient increases from 0.11 to 0.13 for Austin and 0.06 to 0.08 for replication sites after the sample restriction. This pattern is consistent with TPP effects being most pronounced for teachers who have recently completed their training for teachers in advanced subjects.

Our main results in Tables 6 through 7 have two additional columns that include different forms of fixed effects: school-track fixed effects (column 5) and student fixed effects (column 6). Results indicate that the choice between school fixed effects in column 4 and school-track fixed effects in column 5 is unimportant to our substantive findings. Finally, in results available from authors, when we restrict our sample of high schools to schools that do not appear to track by ability grouping (about 80% of schools),[38] UTeach coefficient estimates are modestly smaller for Austin graduates teaching EOC math (0.08) and unchanged for Austin graduates teaching EOC science and for replication site graduates in all subjects.

VI. Conclusion

Relative to other teachers in the state, we find that graduates of both the UTeach founding site at Austin and the replication sites are more effective as measured by their ability to raise student test scores in math and science. In some cases, these effect sizes are very large, with high school math and science students taught by UTeach Austin graduates estimated to accrue 4-6 months of additional learning in a 9 month school year. We conduct several tests to assess whether the positive UTeach effects we find might be driven by overall university-level effects such as the selection of students to universities rather than UTeach program effects. While it is difficult to definitely distinguish between true program effects and general institution-level effects, these tests suggest that the introduction of UTeach to a given university is associated with an increase in the performance of the STEM teachers produced by that university. In addition, the inclusion of proxies for selectivity such as the SAT scores of incoming students cannot fully explain the UTeach effects we find.

While our tests are suggestive of true UTeach program effects, there are several reasons that the findings in this paper are still important even if fully driven by institution effects. First, a primary goal of UTeach is to increase the production of STEM teachers from partner

[38] Specifically, we retain a sample of schools where the average deviation in prior math test scores at the classroom level from the overall school average is less than 0.05 standard deviations.

universities. While not the focus of the paper, we find evidence that UTeach partner universities do produce more STEM teachers after the implementation of UTeach. Thus, even if the UTeach program itself does not improve the quality of a given teacher candidate, by producing more teachers from universities with above average teachers, the program would still improve the overall quality of the STEM workforce. Second, our results suggest that whether a teacher candidate graduated from a UTeach program is a signal of an applicant's quality. While distinguishing selection effects from training effects is important for policymakers who operate at the state and university levels, for leaders of secondary schools interested in the learning of their students, the mechanism driving the effectiveness of UTeach-trained teachers is not particularly relevant. Finally, UTeach offers a 4 year degree plan that condenses the certification courses that were offered in the previous programs at UTeach-affiliated institutions prior to UTeach implementation. Our results suggest that condensing these courses has not resulted in detrimental performance once teachers enter the classroom

References

Ares, N., Stroup, W. M., & Schademan, A. R. (2009). The power of mediating artifacts in group-level development of mathematical discourses. *Cognition and Instruction*, *27*(1), 1–24.

Augustine, N. R. (2007). Is America falling off the flat earth?. National Academies Press.

Bacher-Hicks, A., Kane, T. J., & Staiger, D. O. (2014). *Validating teacher effect estimates using changes in teacher assignments in Los Angeles* (No. w20657). Cambridge, MA: National Bureau of Economic Research.

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, *78*(3), 367–409.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy, 1*(2), 176–216.

Boyd, D., Lankford, H., Loeb, P. & Wyckoff, J. (2005). The draw of home: How teachers' preferences for proximity disadvantage urban youth. *Journal of Policy Analysis and Management*, *24*(1), 113–132.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *The American Economic Review*, *104*(9), 2593–2632.

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

Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs. Princeton, NJ: Mathematica Policy Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655–681.

Confrey, J., Makar, K., & Kazak, S. (2004). Undertaking data analysis of student outcomes as professional development for teachers. *ZDM: International Reviews on Mathematics Education*, *36*(1), 32–40.

Cowan, J., Goldhaber, D., Hayes, K., & Theobald, R. (2016). Missing elements in the discussion of teacher shortages. *Educational Researcher*, 0013189X16679145.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, *86*(1), 195–210.

Dickinson, G., & Summers, E. J. (2010). (Re)Anchored, video-centered engagement: The transferability of preservice training to practice. *Contemporary Issues in Technology and Teacher Education*, *10*(1), 106–118.

Goldhaber, D. (2008). Teachers matter, but effective teacher quality policies are elusive. In *Handbook of research in education finance and policy*. (pp. 146–165).

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, 505-523.

Goldhaber, D., Gratz, T., & Theobald, R. (2016). What's in a Teacher Test? Assessing the Relationship Between Teacher Licensure Test Scores and Student STEM Achievement and Course-Taking. *Working paper.*

Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, *47*(1), 218–251.

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.

Goldhaber, D., & Walch, J. (2013). *Rhetoric versus reality: Is the academic caliber of the teacher workforce changing?* (Vol. 4). CEDR Working Paper 2013.Hansen, M., Backes, B., Brady, V., & Xu, Z. (2015). *Examining spillover effects from Teach For America corps members in*

Miami-Dade County Public Schools. National Center for Analysis of Longitudinal Data in Education Research Working Paper, 113.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, *100*(2), 267–271.

Harris, D., & Sass, T. (2006). *The effects of teacher training on teacher value-added*. Florida State University.

Heck, R. H. (2009). Teacher effectiveness and student achievement: Investigating a multilevel cross-classified model. *Journal of Educational Administration*, *47*(2), 227–249.

Hock, H., & Isenberg, E. (2012). Methods for Accounting for Co-Teaching in Value-Added Models. *Working Paper*. Mathematica Policy Research, Inc.

Holdren, J. P., Lander, E. S., & Varmus, H. (2010). Prepare and inspire: K–12 education in science, technology, engineering, and math (STEM) for America's future. *Executive Report).* Washington, DC: President's Council of Advisors on Science and Technology.

Hoxby, C. M. (2009). The changing selectivity of American colleges. *The Journal of Economic Perspectives*, *23*(4), 95–118.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American educational research journal*, *38*(3), 499–534.

Jackson, C. K. (2014). Teacher quality at the high-school level: The importance of accounting for tracks. *Journal of Labor Economics*, 32(4), 645-684.

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.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). Cambridge, MA: National Bureau of Economic Research.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. In *Research Paper. MET Project.* Seattle, WA: Bill & Melinda Gates Foundation.

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*. http://www.mitpressjournals.org/doi/pdf/10.1162/EDFP_a_00172.

Lipsey, Mark W., Kelly Puzio, Cathy Yun, Michael A. Hebert, Kasia Steinka-Fry, Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick. (2012). "Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms." National Center for Special Education Research.

Makar, K., & Confrey, J. (2005). “Variation-talk”: Articulating meaning in statistics. *Statistics Education Research Journal*, *4*(1), 27–54.

Marder, M., & Walkington, C. (2014). Classroom observation and value-added models give complementary information about the quality of mathematics teaching. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measuring Effective Teaching project* (pp. 234–277). New York, NY: Wiley.

Marshall, J., & Young, E. S. (2006). Preservice teachers’ theory development in physical and simulated environments. *Journal of Research in Science Teaching*, *43*(9), 907–937.

Peterson, P. E., Woessmann, L., Hanushek, E. A., & Lastra-Anadón, C. X. (2011). *Globally challenged: Are U.S. students ready to compete? The latest on each state’s international standing in math and reading*. PEPG 11-03. Program on Education Policy and Governance, Harvard University.

Rice, J. K., & Schwartz, A. E. (2008). Toward an understanding of productivity in education. In *Handbook of research in education finance and policy* (pp. 131–145).

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education*, *6*(1), 43–74.

Stroup, W., Hills, T., & Carmona, G. (2011). Computing the average square: An agent-based introduction to aspects of psychometric practice. *Technology, Knowledge and Learning*, *16*(3), 199–220.

Von Hippel, P. T., Bellows, L., Osborne, C., Lincove, J. A., & Mills, N. (2016). Teacher quality differences between teacher preparation programs: How big? How reliable? Which programs are different? *Economics of Education Review*, *53*, 31–45.

West, M. (2013). Do Math and Science Teachers Earn More Outside of Education? *Brookings*. https://www.brookings.edu/research/do-math-and-science-teachers-earn-more-outside-of-education/

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach For America in high school. *Journal of Policy Analysis and Management*, *30*(3), 447–469.

Table 1: Number of UTeach Teachers by Campus and Sample Year

	EOG M				EOC M				EOC S			
	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015
Arlington	0	0	0	≤5	0	0	0	7	0	0	0	7
Austin	24	27	22	21	96	115	100	93	60	76	57	66
Dallas	≤5	≤5	6	13	≤5	≤5	11	17	0	≤5	6	10
Houston	17	27	46	53	15	28	33	41	≤5	8	15	22
UNT	8	17	23	35	11	37	56	57	≤5	6	9	11
Rio Grande	0	0	0	8	0	0	0	9	0	0	0	≤5
Tyler	0	0	≤5	9	0	0	0	7	0	0	0	0

Notes: Cell values indicate the number of unique teachers linked to students with valid test scores. "EOG M" denotes teachers of EOG math (Grades 6–8), "EOC M" teachers of EOC math (Algebra I, Geometry, and Algebra II), and "EOC S" denotes teachers of EOC science (Biology, Chemistry, and Physics).

Table 2: Number of Teachers in Analysis Sample by Campus and Graduation Year

	Arlington	Austin	Dallas	Houston	Rio Grande	Tyler	UNT
1995	19	15	14	20	68	16	22
1996	29	17	15	19	59	8	18
1997	14	9	≤5	14	48	7	18
1998	13	8	12	21	48	7	20
1999	8	≤5	11	14	57	7	19
2000	13	11	≤5	15	43	≤5	18
2001	8	21	11	21	41	11	19
2002	9	8	8	9	45	11	14
2003	8	13	≤5	12	34	≤5	6
2004	8	13	6	≤5	27	6	9
2005	11	26	8	9	31	≤5	14
2006	8	22	15	15	27	≤5	15
2007	11	21	10	20	39	≤5	13
2008	15	33	6	21	30	≤5	19
2009	23	35	10	13	33	6	15
2010	22	27	6	29	42	9	25
2011	33	44	10	24	31	8	41
2012	18	38	13	41	25	15	44
2013	19	29	17	35	18	8	39
2014	25	23	19	27	21	9	19

Notes: Each cell denotes the number of STEM teachers in our analysis sample who graduated from a given campus in a given calendar year. Numbers below horizontal lines represent teachers who graduated after UTeach had been implemented at a given campus.

Table 3: Summary Statistics of Students Taught by UTeach and Non-UTeach Teachers

	EOG M			EOC M			EOC S		
	Non-U	Austin	Replication	Non-U	Austin	Replication	Non-U	Austin	Replication
Male	0.51	0.46	0.51	0.49	0.51	0.50	0.49	0.50	0.51
Black	0.11	0.09	0.17	0.12	0.12	0.15	0.12	0.11	0.19
Hispanic	0.52	0.53	0.52	0.48	0.48	0.51	0.48	0.43	0.51
Asian	0.03	0.06	0.06	0.04	0.07	0.06	0.04	0.05	0.07
White	0.32	0.29	0.23	0.32	0.30	0.24	0.32	0.37	0.19
Other	0.02	0.03	≤0.02	0.02	0.02	0.02	0.02	0.03	0.02
LEP	0.16	0.15	0.23	0.09	0.11	0.14	0.09	0.09	0.17
FRL	0.59	0.53	0.60	0.51	0.49	0.57	0.51	0.43	0.57
Spec Ed	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.03	0.05
Gifted	0.09	0.14	0.09	0.12	0.13	0.10	0.11	0.13	0.09
Grade	6.93	7.11	6.99	9.01	9.14	9.00	9.28	9.39	9.13
	(0.77)	(0.82)	(0.77)	(0.63)	(0.55)	(0.59)	(0.49)	(0.53)	(0.33)
Lag m	-0.03	0.18	-0.08	0.16	0.18	0.09	-0.00	0.13	-0.15
	(0.50)	(0.53)	(0.52)	(0.70)	(0.73)	(0.75)	(0.51)	(0.52)	(0.47)
Lag r	-0.02	0.19	-0.05	0.10	0.13	-0.01	0.05	0.23	-0.14
	(0.46)	(0.52)	(0.51)	(0.59)	(0.62)	(0.59)	(0.53)	(0.56)	(0.52)
Lag s							0.03	0.28	-0.11
							(0.60)	(0.65)	(0.56)
Miss lagm	0.04	0.04	0.04	0.05	0.05	0.05	0.06	0.06	0.07
Miss lagr	0.04	0.04	0.04	0.06	0.07	0.06	0.07	0.08	0.07
Miss prior s							0.07	0.07	0.09

Notes: “Non-U” denotes non-UTeach, “Austin” denotes the UT Austin site, and “Replication” denotes UTeach replication (non-Austin) sites. “EOG M” denotes teachers of EOG math (Grades 6–8), “EOC M” teachers of EOC math (Algebra I, Geometry, and Algebra II), and “EOC S” denotes teachers of EOC science (Biology, Chemistry, and Physics). Standard deviations in parentheses.

Table 4: Summary Statistics for Teachers

	EOG M			EOC M			EOC S		
	Non-U	Austin	Replication	Non-U	Austin	Replication	Non-U	Austin	Replication
Yrs exp	9.57	4.76	0.93	10.68	3.79	0.97	10.29	3.50	0.52
	(8.68)	(3.82)	(1.11)	(9.38)	(3.49)	(0.98)	(9.41)	(2.85)	(0.81)
1st yr teacher	0.09	0.11	0.44	0.08	0.16	0.35	0.09	0.14	0.55
2-3rd yr teacher	0.14	0.24	0.42	0.13	0.28	0.46	0.13	0.28	0.30
Missing tch exp	0.03	0.04	0.05	0.03	0.03	0.11	0.03	0.05	0.13
SAT 75th pct M	585	720	630	589	720	626	592	720	638
	(63)	(0)	(30)	(65)	(0)	(33)	(66)	(0)	(37)
SAT 75th pct R	569	690	607	572	690	606	575	690	612
	(58)	(0)	(26)	(60)	(0)	(28)	(60)	(0)	(35)
Missing SAT M scores	0.45	0.03	0.01	0.43	0.01	0.01	0.43	0.01	0.04
STEM certified	0.48	1	1	0.75	1	1	0.78	1	1
Black	0.10	≤0.02	0.15	0.08	0.08	0.08	0.08	0.04	0.08
Hispanic	0.16	0.34	0.26	0.16	0.33	0.23	0.15	0.13	0.17
White	0.48	0.46	0.53	0.54	0.50	0.56	0.54	0.61	0.50
Teacher-year obs	54513	94	274	48199	404	335	37392	259	104

Notes: "Non-U" denotes non-UTeach, "Austin" denotes the UT Austin site, and "Replication" denotes UTeach replication (non-Austin) sites. "EOG M" denotes teachers of EOG math (Grades 6–8), "EOC M" teachers of EOC math (Algebra I, Geometry, and Algebra II), and "EOC S" denotes teachers of EOC science (Biology, Chemistry, and Physics). Standard deviations in parentheses.

Table 5: Coefficient Estimates of UTeach by Subject, Math

	1	2	3	4	5	6
Panel 1: EOG Math						
UTeach	0.04**	0.08***	0.07***	0.04**	0.02*	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Panel 2: Algebra I						
UTeach	0.10***	0.10***	0.09***	0.01	0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)
Panel 3: Geometry						
UTeach	0.03	0.05*	0.04	0.04*	0.06***	0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)
Panel 4: Algebra II						
UTeach	0.04	0.07	0.08	0.04	0.05	0.00
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.00)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: Additional controls include teacher experience, a dummy variable for missing teacher experience, identifiers for FRL, gifted, LEP, and special education, cubic functions of prior reading and prior match scores, and indicators for missing test scores. Standard errors displayed in parentheses are clustered at the school-cohort level.

Table 6: Coefficient Estimates of UTeach by Subject, Science

	1	2	3	4	5	6
Panel 1: Biology						
UTeach	0.11***	0.12***	0.12***	0.06**	0.04**	0.02
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Panel 2: Chemistry						
UTeach	0.09*	0.11**	0.11**	0.04	0.04	-0.00
	(0.05)	(0.05)	(0.05)	(0.03)	(0.03)	(0.00)
Panel 3: Physics						
UTeach	0.11	0.16	0.16	-0.06	-0.01	-0.00
	(0.13)	(0.13)	(0.13)	(0.09)	(0.05)	(0.00)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5.

Table 7: Coefficient Estimates for UT Austin and UTeach Replication Sites

	1	2	3	4	5	6
Panel 1: EOG Math						
Austin	0.05*	0.06**	0.05*	0.03	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Other UTeach	0.04*	0.08***	0.07***	0.05**	0.02*	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Panel 1: EOC Math						
Austin	0.10***	0.11***	0.10***	0.01	0.02	-0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)
Other UTeach	0.05*	0.06**	0.06**	-0.00	-0.00	-0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Panel 1: EOC Science						
Austin	0.13***	0.14***	0.14***	0.06***	0.06***	0.07**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Other UTeach	0.02	0.06	0.06	-0.02	-0.02	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. "EOG M" denotes teachers of EOG math (Grades 6–8), "EOC M" teachers of EOC math (Algebra I, Geometry, and Algebra II), and "EOC S" denotes teachers of EOC science (Biology, Chemistry, and Physics).

Table 8: Coefficient Estimates for UTeach When Controlling for Undergraduate Institution Selectivity

	1	2	3	4	5	6
Panel 1: EOG Math						
Austin	0.02	0.03	0.02	0.02	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Other UTeach	0.03	0.07***	0.06**	0.05**	0.03**	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Panel 1: EOC Math						
Austin	0.08***	0.08***	0.08***	0.01	0.02	-0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Other UTeach	0.05*	0.06**	0.06**	-0.00	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Panel 1: EOC Science						
Austin	0.07**	0.08***	0.08***	0.04	0.04*	0.06*
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Other UTeach	0.01	0.05	0.05	-0.02	-0.02	0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. “EOG M” denotes teachers of EOG math (Grades 6–8), “EOC M” teachers of EOC math (Algebra I, Geometry, and Algebra II), and “EOC S” denotes teachers of EOC science (Biology, Chemistry, and Physics). Additional controls include quintiles of 75th percentile SAT math and reading scores of the undergraduate institution the teacher attended.

Table 9: Coefficient Estimates for UTeach When Controlling for STEM Certification Scores

	1	2	3	4	5	6
Panel 1: EOG Math						
Austin	0.04	0.05*	0.04	0.02	0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Other UTeach	0.04*	0.08***	0.07***	0.05**	0.02*	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Panel 1: EOC Math						
Austin	0.08***	0.09***	0.08***	0.00	0.02	-0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)
Other UTeach	0.03	0.05*	0.04	-0.01	-0.01	-0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Panel 1: EOC Science						
Austin	0.11***	0.11***	0.12***	0.05**	0.05***	0.07**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Other UTeach	-0.01	0.04	0.04	-0.03	-0.03	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. "EOG M" denotes teachers of EOG math (Grades 6–8), "EOC M" teachers of EOC math (Algebra I, Geometry, and Algebra II), and "EOC S" denotes teachers of EOC science (Biology, Chemistry, and Physics). Additional controls include a cubic function of a teacher's STEM certification score.

Table 10: Coefficient Estimates for UTeach When Controlling for Undergraduate Institution Selectivity, STEM Certification Scores, and Certification Route (Updated)

	1	2	3	4	5	6
Panel 1: EOG Math						
Austin	0.02	0.01	0.00	0.01	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Other UTeach	0.03	0.06**	0.05**	0.04**	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Panel 1: EOC Math						
Austin	0.06**	0.05	0.04	-0.01	0.01	-0.02
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Other UTeach	0.03	0.03	0.02	-0.02	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Panel 1: EOC Science						
Austin	0.06**	0.07**	0.07**	0.04	0.04**	0.06
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.04)
Other UTeach	-0.01	0.04	0.04	-0.02	-0.02	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. “EOG M” denotes teachers of EOG math (Grades 6–8), “EOC M” teachers of EOC math (Algebra I, Geometry, and Algebra II), and “EOC S” denotes teachers of EOC science (Biology, Chemistry, and Physics). Additional controls include quintiles of 75th percentile SAT math and reading scores of the undergraduate institution the teacher attended, a cubic function of a teacher’s STEM certification score, and controls for certification route (standard, alternative, other/missing). UTeach coefficients are relative to the omitted group of other teachers whose certification route is from a standard university-based program.

Table 11: Interactions with Year

	EOG M		EOC M		EOC S	
	1	2	3	4	5	6
Austin	0.04	0.01	0.06	-0.02	0.17**	0.09
	(0.06)	(0.05)	(0.06)	(0.03)	(0.08)	(0.06)
Other UTeach	0.13**	0.01	-0.02	-0.02	0.18***	0.11***
	(0.07)	(0.04)	(0.08)	(0.06)	(0.02)	(0.03)
Austin * 2013	-0.02	-0.04	0.03	0.04	-0.02	-0.00
	(0.07)	(0.06)	(0.07)	(0.04)	(0.08)	(0.07)
Austin * 2014	0.04	-0.00	0.10	0.09*	-0.01	-0.01
	(0.08)	(0.06)	(0.11)	(0.05)	(0.09)	(0.07)
Austin * 2015	0.10	0.11	0.09	0.01	-0.07	-0.08
	(0.10)	(0.08)	(0.09)	(0.05)	(0.09)	(0.07)
Replication * 2013	-0.10	-0.01	0.06	0.01	-0.09**	-0.16**
	(0.08)	(0.06)	(0.09)	(0.07)	(0.05)	(0.07)
Replication * 2014	-0.07	0.02	0.12	0.02	-0.09	-0.11
	(0.08)	(0.06)	(0.10)	(0.07)	(0.11)	(0.10)
Replication * 2015	-0.03	0.06	0.09	0.01	-0.14***	-0.13**
	(0.07)	(0.05)	(0.10)	(0.07)	(0.05)	(0.06)
School FEs		x		x		x

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. “EOG M” denotes teachers of EOG math (Grades 6–8), “EOC M” teachers of EOC math (Algebra I, Geometry, and Algebra II), and “EOC S” denotes teachers of EOC science (Biology, Chemistry, and Physics).

Table 12: Coefficient Estimates for UTeach and Replication Campuses Prior to UTeach

	1	2	3	4	5	6
Panel 1: EOG Math						
Austin	0.05*	0.06**	0.05*	0.03	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Other UTeach	0.04*	0.08***	0.07***	0.05**	0.02*	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Replication Pre	-0.00	-0.01	-0.01**	-0.00	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Panel 1: EOC Math						
Austin	0.11***	0.11***	0.10***	0.01	0.02	-0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)
Other UTeach	0.05*	0.06**	0.06**	-0.00	-0.00	-0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Replication pre	0.03**	0.03**	0.02*	0.01	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel 1: EOC Science						
Austin	0.13***	0.14***	0.14***	0.06***	0.06***	0.07**
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Other UTeach	0.02	0.06	0.06	-0.02	-0.02	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)
Replication pre	-0.04***	-0.04***	-0.04***	-0.02***	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. "EOG M" denotes teachers of EOG math (Grades 6–8), "EOC M" teachers of EOC math (Algebra I, Geometry, and Algebra II), and "EOC S" denotes teachers of EOC science (Biology, Chemistry, and Physics).

Table 13: Reading Results for non-STEM Graduates of Campuses With UTeach

	1	2	3	4	5
Panel 1: EOG Reading					
Austin	0.05***	0.05***	0.03***	0.02*	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Other UTeach	-0.01	0.01	-0.01	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Other UTeach Pre	0.00	0.00	-0.00	-0.00	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Panel 1: EOC Reading					
Austin	0.09***	0.11***	0.03**	0.02**	0.03**
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Other UTeach	0.02	0.05***	0.02*	0.01	0.02
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Other UTeach Pre	0.02***	0.02**	0.01**	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Student chars	X	X	X	X	X
Teacher chars		X	X	X	X
Fixed effect			School	School-track	Student

Notes: See notes from Table 5. In any given year, only one EOC at a given level is administered (e.g., English I or Reading I). Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. "EOG R" denotes teachers of EOG reading (Grades 6_8) and "EOC R" teachers of EOC reading (Reading II, Reading II, English I, and English II).

Table 14: Novice (in First Through Third Year of Teaching) Teachers Only

	1	2	3	4	5	6
Panel 1: EOG Math						
Austin	0.01	0.01	-0.00	0.05	0.03	-0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.09)
Other UTeach	0.07***	0.08***	0.06***	0.03	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Panel 1: EOC Math						
Austin	0.13***	0.13***	0.12***	0.01	-0.00	-0.05
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)
Other UTeach	0.08***	0.08***	0.06**	0.00	-0.00	0.02
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
Panel 1: EOC Science						
Austin	0.16***	0.16***	0.16***	0.10***	0.09***	-0.04
	(0.05)	(0.05)	(0.05)	(0.04)	(0.03)	(0.05)
Other UTeach	0.04	0.06	0.06	-0.00	-0.02	0.06
	(0.04)	(0.04)	(0.04)	(0.05)	(0.02)	(0.05)
Student chars	X	X	X	X	X	X
Teacher chars		X	X	X	X	X
STEM certified			X			
Fixed effect				School	School-track	Student

Notes: See notes from Table 5. Replication sites include University of Houston, University of North Texas, UT Dallas, UT Arlington, and UT Tyler. “EOG M” denotes teachers of EOG math (Grades 6–8), “EOC M” teachers of EOC math (Algebra I, Geometry, and Algebra II), and “EOC S” denotes teachers of EOC science (Biology, Chemistry, and Physics).

Appendix Table 1: Full Regression Coefficients From Most Commonly Taken Subjects

	EOG M	Algebra I	Biology
	(1)	(2)	(3)
UTeach	0.08***	0.10***	0.12***
	(0.02)	(0.02)	(0.03)
Male	0.02***	0.00	0.02***
	(0.00)	(0.00)	(0.00)
Black	-0.13***	-0.10***	-0.06***
	(0.00)	(0.01)	0.00
Hispanic	-0.08***	-0.05***	-0.08***
	(0.00)	(0.01)	(0.00)
Asian	0.35***	0.15***	0.27***
	(0.01)	(0.01)	(0.01)
Other	-0.01**	(0.00)	0.04***
	(0.00)	(0.01)	(0.00)
Prior Math	0.70***	0.42***	0.14***
	(0.00)	(0.00)	(0.00)
Prior reading	0.17***	0.16***	0.28***
	(0.00)	(0.00)	(0.00)
Prior science			0.44***
			(0.00)
1-2 years exp	0.05***	0.07***	0.05***
	(0.00)	(0.01)	(0.01)
3-5 years exp	0.09***	0.07***	0.08***
	(0.00)	(0.01)	(0.01)
6-9 years exp	0.10***	0.04***	0.07***
	(0.00)	(0.01)	(0.01)
10+ years exp	0.09***	0.04***	0.04***
	(0.00)	(0.01)	(0.01)
Missing exp	-0.02*	-0.09***	-0.04**
	(0.01)	(0.03)	(0.01)
Student chars	X	X	X
Teacher chars	X	X	X

Notes: Additional controls include squared and cubic terms in prior test scores, limited English proficiency, FRL eligibility, special education status, and gifted status. Standard errors clustered at the school-cohort level. The regressions represented here are identical to those that generated column 2 of Tables 5 and 6.

Appendix Table 2: Program Estimates

	EOG M		EOC M		EOC S	
	1	2	3	4	5	6
Arlington	-0.06**	-0.08**	0.09***	-0.07	0.15	0.14
	(0.02)	(0.04)	(0.03)	(0.13)	(0.16)	(0.10)
Austin	0.06**	0.03	0.11***	0.01	0.14***	0.06***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Tyler	0.05	0.07	-0.22	-0.16		
	(0.05)	(0.06)	(0.21)	(0.26)		
Dallas	0.23***	0.17***	0.12***	0.04	0.29	0.21
	(0.09)	(0.06)	(0.05)	(0.05)	(0.19)	(0.18)
RioGrande	-0.10	-0.06	0.19	0.15	-0.07**	-0.04
	(0.07)	(0.06)	(0.19)	(0.14)	(0.03)	(0.03)
UNT	0.05*	0.06**	0.06	0.00	0.01	0.01
	(0.03)	(0.03)	(0.04)	(0.02)	(0.05)	(0.05)
Houston	0.09***	0.03	0.05	-0.02	0.02	-0.10***
	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)	(0.03)
Student chars	X	X	X	X	X	X
Teacher chars	X	X	X	X	X	X
School FEs		x		x		X

Notes: “EOG M” denotes teachers of EOG math (Grades 6–8), “EOC M” teachers of EOC math (Algebra I, Geometry, and Algebra II), and “EOC S” denotes teachers of EOC science (Biology, Chemistry, and Physics).