

Exploratory Study of the UTeach STEM Preparation Program and the Effectiveness of UTeach Teachers

Whitney Cade, Feng Liu, Michael Vaden-Kiernan, Melissa Dodson - American Institutes for Research
May 2019

What We Studied

According to the Department of Education, there are more than 2,000 teacher preparation programs (TPPs) in the United States¹, but despite these numbers, concerns remain about the quality and quantity of STEM teachers in the workforce. For instance, Augustine (2007) finds that 61% of chemistry teachers and 67% of physics teachers did not major in and/or receive certification in their topic area (Augustine, 2007), and thus may not have the same content knowledge as a teacher who specialized in that topic. Furthermore, 20-30% of schools reported difficulties finding and employing STEM teachers between 2000 and 2012 (Cowan et al., 2016).

The UTeach teacher preparation program (TPP), founded at the University of Texas at Austin in 1997, is specifically designed to address both of these concerns by recruiting students directly from STEM majors and offering student the opportunity to receive a secondary STEM teaching certification alongside their STEM degree with no additional time in college. UTeach offers early, tuition-free teaching experiences so that prospective teachers can try teaching before seriously pursuing certification, which allows them to attract a larger pool of prospective teachers. According to UTeach administrators, the adoption of the UTeach program has led to a dramatic increase in the number of STEM majors enrolling in education courses and graduating with teacher certification in STEM disciplines—they estimate that the 46 UTeach programs across the country will graduate over 8,000 teachers by 2023.

While UTeach has undoubtedly increased the number of teachers produced at universities that have adopted the program, questions remain as to the quality of those teachers when compared to other teachers in the workforce. UTeach has never had a formal, third-party evaluation of its teachers in the field; this study intended to fill this gap by investigating the test scores of the students of UTeach-produced teachers.

However, if a difference was found between UTeach teachers and non-UTeach teachers, this may not point to why there is a difference. Very little is known about the malleable factors (i.e. things programs can change) that may influence the effectiveness of the teachers TPPs produce. While many studies have failed to find differences among TPPs (which are often single-university programs; Goldhaber, Liddle, & Theobald, 2013; von Hippel, Bellows, Osborne, Lincove, & Mills, 2016), this study sought to compare the qualities of the Texas UTeach programs to other university-based TPPs that produce the bulk of the teachers in Texas.

¹ <https://title2.ed.gov/Public/Home.aspx>

This study pursued the following research questions:

1. How do students taught by UTeach teachers compare to students taught by non-UTeach teachers in terms of their math and science state test scores?
 - a. Is there a “UTeach effect” when...
 - Breaking UTeach out into its flagship location (Austin) and its replication sites in Texas?
 - Accounting for the schools teachers teach in?
 - Accounting for the selectivity of the teacher’s university?
2. Are there differences in student test scores between UTeach and non-UTeach teachers for different subgroups of students (i.e., does the UTeach effect hold true for all students)?
3. Are there certain malleable factors that may explain differences between TPPs?

How We Analyzed the Data

To answer these research questions, the research team focused on Texas UTeach programs, teachers, and students, using data available at the Texas Education Research Center (ERC). Although the ERC has relevant data going back over a decade, the ability to link students to teachers has only been available since the 2011-12 school year. Therefore, this study only considers data between the 2011-12 and 2015-16 school years. Researchers assembled data on student demographics, course enrollment, and test scores and linked this data with teacher data (courses taught and demographics). This analysis focused on students who took courses that have an associated end-of-course (EOC) test or grade levels with associated end-of-grade (EOG) tests (such as 8th grade math) in math or science, and linked these students with the teacher or teachers who taught this class². Rather than focus on outcomes for single tests, this project conducted analyses on all high school science tests, high school math tests, and middle school math tests (middle school science resulted in too few scores to make analyses feasible).

This project was not supplied with a list of all former UTeach students, but instead had to use data in the ERC to determine likely candidates. Teachers were identified as graduating from a UTeach program if they met the following criteria: received a BS from the University of Texas at Austin, the University of Houston, the University of Texas at Dallas, the University of Texas at Tyler, the University of North Texas, the University of Texas at Arlington, or the University of Texas Rio Grande Valley; graduated from one of these schools after UTeach was implemented at the university; and holds a secondary STEM certification, granted by the same institution that they graduated from.

What We Discovered

Overall, the students of teachers trained by UTeach have higher scores on state math and science tests than the students of teachers trained through other types of programs.

Students of UTeach teachers score higher on both math and science tests than students of other types of teachers. These effects on student test scores are quite sizeable; when considering a 9-month school year, students with a UTeach teacher taking a high school EOC showed the equivalent of 7.7 months more learning, and the smallest effect, the middle school EOC math, is equivalent to 2.7 months³.

This project also attempted to dive deeper into this effect by separating out the UTeach teachers who entered UTeach with the intention to teach and those who had originally entered without the intention to teach but had completed the UTeach program and become a teacher anyway (as measured by early surveys administered to students in the UTeach program). By offering free courses that give early experiences with teaching, UTeach lets STEM majors “try out” teaching, thus tapping into a pool of people that are untapped in other programs where gaining admittance to a TPP is more purposeful and requires several prerequisite classes. Upon integrating the survey data necessary to assess the

² Only test scores from the first time a student took a test are considered in this study.

³ All learning-months estimates use Lipsey et al.’s (2012) estimations of annual achievement gains for math and science.

teacher's early-program intentions towards teaching, there were not enough matches between the surveys and the existing ERC data to make analyses on this front feasible (see Cade, Liu, Vaden-Kiernan, and Dodson, 2019).

This “UTeach effect” is not driven solely by the University of Austin; teachers trained by the other Texas UTeach sites see a similar effect with their students.

UTeach has been implemented longest at its flagship school, UT Austin (20 years), while the other Texas UTeach universities (called “replication sites”) have only adopted UTeach in the last 10 years. Therefore, it is possible that it is UT Austin and not the replication sites are driving these large learning gains. However, when comparing UT Austin teachers and replication sites separately to all other teachers in Texas, a similar pattern arises. For high school math and science, the students of teachers from both UT Austin and the replication site outperform students of non-UTeach teachers (with gains ranging from 3.5 to 7.8 additional months of learning). For middle school, though, only students of UT Austin teachers outperform students of non-UTeach teachers (6.3 months). When comparing Austin teachers to replication site teachers directly, significant differences in performance only arise for middle school math, where UT Austin students outperform the students of all other UTeach replication sites (a difference of 5 months of learning).

The UTeach effect is not a result of the schools teachers are placed in.

It is possible that UTeach teachers are placed in higher-performing schools and thus teach higher performing students, which could account for differences between students of UTeach and non-UTeach teachers. After factoring in school characteristics (like the average school math and English test scores from the prior year, percent of minority students, and the percent of students who receive a free/reduced price lunch), there is very little change to the effects reported in the first Key Finding. Using stricter controls for schools (like actual school codes and statistical models that nest teachers and students under schools) does reduce the UTeach effect somewhat, although there still remains a statistically significant difference between UTeach and non-UTeach teachers. Therefore, it seems unlikely that UTeach teachers appear more effective than their non-UTeach counterparts because they teach at higher performing schools.

The UTeach effect could be due to the highly selective nature of the universities that offer UTeach, but the evidence for that is mixed.

It is possible that the UTeach effect is due to the highly selective nature of the universities which have adopted it, given that they are among some of the most prominent universities in Texas. This would mean that, rather than having an exceptional preparation program, these universities could simply have access to a pool of higher-performing college students who would have gone on to become effective teachers regardless of where they went to college. This hypothesis was tested in three different ways: 1) by accounting for each university's minimum required math and reading SAT scores for entrance and the teacher's performance on their STEM certification test, 2) by comparing student who graduated from UTeach universities *before* UTeach was implemented to all other teachers, and 3) by comparing English teachers from UTeach universities to English teachers from all other universities (since UTeach is STEM only, these students are trained by other programs at the same universities as UTeach teachers). Among all three methods, the only method with evidence for a selection effect was with the English teachers – it was found that students of English teachers from UTeach universities had higher English exam scores than students of English teachers from other institutions (with results ranging from 1.2 years to 2.4 months). However, given the failure of the other methods to show evidence of a university selection effect and the differing subject matter and training programs for English teachers, it is difficult to conclude that the UTeach effect is due solely to the highly selective nature of the universities they attended.

UTeach teachers have some minor strengths and weaknesses when it comes to teaching certain types of learners.

This project also investigated the whether UTeach teachers were as effective at teaching all groups of students, specifically focusing on economically disadvantaged students, English language learners, females, Hispanic students, and African American students.

This project found that:

- UTeach teachers are no better or worse at teaching economically disadvantaged students than other teachers in Texas.
- English language learners taught by UTeach teachers have lower high school science test scores compared with the students of non-UTeach teachers, but this effect was weak.
- Female students taught by UTeach teachers have higher high school math test scores, but this effect was weak.
- Hispanic students taught by UTeach teachers have lower middle school math test scores, but this effect was modest.
- African American students taught by UTeach teachers have lower high school math test scores, but this effect was weak.

All other results for other tests or groups of students which were not mentioned above were equal between UTeach teachers and non-UTeach teachers. While there is clearly some room for UTeach to improve in terms of equity, the differences between UTeach and non-UTeach teachers are generally small and may only be significant due to the vast number of teachers in the sample.

These results can be replicated using different analytic techniques.

This project analyzed all results using two different approaches: hierarchical linear models and value-added modeling. Both approaches yielded very similar results (reported above). To read a paper summarizing the results of the value-added modeling, please see Backes, Goldhaber, Cade, Sullivan, and Dodson (2018) or the working paper here: <https://caldercenter.org/sites/default/files/WP%20173.pdf>

More work is needed to determine which program features relate to teacher quality.

Examining the malleable factors of a TPP which could influence the quality of its teachers is a large undertaking. The research team reviewed the literature and determined that no databases of program features exist outside of private organizations, and very few instruments collect data at a program feature-level. The instruments that exist are typically home-grown and do not have strong psychometric properties. As part of an effort to gather information on malleable factors in the face of these limitations, the research team developed a rubric to capture publicly-available data on teacher preparation features from the websites of the 50 largest universities in Texas. An institution's documents and websites usually reflect the basic requirements and program details. The research team employed a strict protocol for gathering the information. Two researchers followed the protocol and the team conducted numerous checks to ensure inter-rater reliability. Despite the limitations of such an approach (such as missing or outdated information posted to websites), the team was able to find information for the majority of major institutions in Texas. However, analyses of these collected features did not paint a clear picture of which features are important in TPPs when it comes to producing quality teachers. More research is needed to assemble a high-quality feature list which captures the unique traits of each TPP.

Policy Recommendations/Implications

These results indicate that there is some evidence that UTeach may be an effective teacher preparation program when compared to other kinds of programs. Students of these teachers have higher test scores in high school math and science and middle school math. This project tried to explore theories to explain this effect, such as placement in high performing schools and university selectivity, but overall, it appears that the UTeach effect cannot be fully explained by these hypotheses. There is some evidence that UTeach teachers are not universally effective with all kinds of students, indicating that there are some areas of growth and exploration for the program.

If these results highlight a truly better approach to preparing STEM teachers, this has two major implications. First and foremost, it would indicate that the UTeach model is worth replicating widely and would need further investigation to

isolate exactly what makes the program so effective. Second, this would mean that it is possible to assess the differences between TPPs, though perhaps only with programs which have several replicating sites. Several studies looking at the differences between all universities have struggled to find many differences between them, but perhaps, if a program has several sites, these investigations are more viable.

To better understand the UTeach effect, a deeper examination of the qualities of UTeach programs as compared to the qualities of other TPPs must occur, and collecting this information will require a number of steps: extensive discussions with key stakeholders to develop a sound framework of features, a multi-method approach to data collection, and a rigorous research collection and analytic plan.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150156 to the Southwest Educational Development Laboratory. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- Augustine, N. R. (2007). *Is America falling off the flat earth?* National Academies Press.
- Backes, B., Goldhaber, D., Cade, W., Sullivan, K., & Dodson, M. (2018). Can UTeach? Assessing the Relative Effectiveness of STEM Teachers. *Economic of Education Review*, 64, 184–198.
- Cade, W., Liu, F., Vaden-Kiernan, M., & Dodson, M. (2019). *An Examination of Recruitment and Retention of UTeach Program Candidates*. Report prepared for the National Math and Science Initiative. Austin, TX: American Institutes for Research.
- Cowan, J., Goldhaber, D., Hayes, K., & Theobald, R. (2016). Missing elements in the discussion of teacher shortages. *Educational Researcher*, 0013189X16679145.
- 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.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *"Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms."* National Center for Special Education Research.
- 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.

The University of Texas at Austin ERC is a research center and P-20/Workforce Repository site which provides access to longitudinal, student-level data for scientific inquiry and policymaking purposes. Since its inception in 2008, the Texas ERC's goal is to bridge the gap between theory and policy by providing a cooperative research environment for study by both scholars and policy makers. As part of its mission, the ERC works with researchers, practitioners, state and federal agencies, and other policymakers to help inform upon critical issues relating to education today. The views expressed are those of the authors and should not be attributed to The University of Texas at Austin or any of the funders or supporting organizations mentioned herein including the State of Texas. Any errors are attributable to the authors.
