

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Volume 13 Number 42

October 12, 2005

ISSN 1068–2341

Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness¹

Linda Darling-Hammond, Deborah J. Holtzman,
Su Jin Gatlin, and Julian Vasquez Heilig
Stanford University

Citation: Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, 13(42). Retrieved [date] from <http://epaa.asu.edu/epaa/v13n42/>.

Abstract

Recent debates about the utility of teacher education have raised questions about whether certified teachers are, in general, more effective than those who have not met the testing and training requirements for certification, and whether some candidates with strong liberal arts backgrounds might be at least as effective as teacher education graduates. This study examines these questions with a large student-level data set from Houston, Texas that links student characteristics and achievement with data about their teachers' certification status, experience, and degree levels from 1995–2002. The data set also allows an examination of whether

¹The authors would like to thank members of the Research and Accountability division of Houston Independent School District for their assistance in assembling the data set. We also thank researchers who provided methodological advice during the study, including Ed Haertel and Tony Bryk of Stanford University; those who read and gave comments on drafts of the report, including David Berliner and Gene Glass of Arizona State University and Peter Youngs of Michigan State University; four anonymous reviewers and *EPAA* editor, Sherman Dorn, who gave the article careful scrutiny and provided useful feedback prior to publication. We benefited from their comments and take responsibility for any shortcomings that remain.



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by H.W. Wilson & Co. Accepted under the editorship of Sherman Dorn. Send commentary to Casey Cobb (casey.cobb@uconn.edu) and errata notes to Sherman Dorn (epaa-editor@shermamdorn.com).

Teach for America (TFA) candidates—recruits from selective universities who receive a few weeks of training before they begin teaching—are as effective as similarly experienced certified teachers. In a series of regression analyses looking at 4th and 5th grade student achievement gains on six different reading and mathematics tests over a six-year period, we find that certified teachers consistently produce stronger student achievement gains than do uncertified teachers. These findings hold for TFA recruits as well as others. Controlling for teacher experience, degrees, and student characteristics, uncertified TFA recruits are less effective than certified teachers, and perform about as well as other uncertified teachers. TFA recruits who become certified after 2 or 3 years do about as well as other certified teachers in supporting student achievement gains; however, nearly all of them leave within three years. Teachers' effectiveness appears strongly related to the preparation they have received for teaching.

Keywords: teacher education; teacher certification; teacher effectiveness.

Introduction

The relationship between teacher education and teacher effectiveness has been hotly debated in recent years in both research and policy circles (see, for example, Ballou & Podgursky, 2000; Darling-Hammond, 2000a; Darling-Hammond & Youngs, 2002; U.S. Department of Education, 2002). On the one hand, advocates of stronger preparation—especially for teachers in schools serving low-income students and students of color—have argued that teachers need to understand how children learn and how to make material accessible to a wide range of students to be successful (National Commission on Teaching and America's Future, 1996; Shulman, 1987). Studies finding positive effects of teacher education and certification on student achievement seem to support this perspective (Betts, Rueben, & Dannenberg, 2000; Darling-Hammond, 2000b; Ferguson, 1991; Fetler, 1999; Goe, 2002; Goldhaber & Brewer, 2000; Hawk, Coble, & Swanson, 1985; Monk, 1994; Strauss & Sawyer, 1986; Wenglinsky, 2000; Wilson, Floden, & Ferrini-Mundy, 2001).

On the other hand, opponents of teacher education and certification have argued that teacher effectiveness may be as much a function of general academic ability or strong subject matter knowledge as it is related to any specialized training in how to teach (Ballou & Podgursky, 2000; Finn, 1999; US Department of Education, 2002). Representing this view, the Secretary of Education argued in his 2002 report on teacher quality for the dismantling of teacher certification systems and the redefinition of teacher qualifications to emphasize higher standards for verbal ability and content knowledge and to de-emphasize education training, making student teaching and education coursework optional (U.S. Department of Education, 2002, p.19). From this perspective, the courses and other expectations that make up "the bulk of current teacher certification regimes" impose "burdensome requirements" (p. 8) that keep talented individuals out of teaching.

The policy implications of these debates are far-reaching, affecting teacher education and certification policies as well as policies regarding school funding and educational rights. As teacher demand has increased and funding inequities have grown over the past 15 years, many urban and poor rural districts have hired a growing number of individuals on emergency permits or waivers who lack formal preparation for teaching. These individuals typically teach low-income and minority students in the most disadvantaged schools (National Commission on Teaching and America's Future, 1996; Shields et al., 2003). Such inequalities—and related disparities in funding and basic education materials—have spawned lawsuits in more than a dozen states arguing that all students have the right to the resources needed to learn to state standards, including fully qualified teachers.

They also sparked the “highly qualified teacher” requirements of the federal No Child Left Behind Act. However, if courts agree that no special training is needed for teaching, as defendants in many of these lawsuits claim, the legal levers for redressing these inequalities would vanish as certification standards are diluted or ignored and students’ recourse is removed.

Cited in the Secretary’s report and at the center of many of these debates has been the Teach for America (TFA) program, which seeks to recruit academically able new college graduates, many of them from selective universities, into two-year teaching commitments in hard-to-staff districts. Following a summer program that provides several weeks of student teaching and basic coursework, recruits are placed in urban and poor rural schools on emergency teaching permits. Although in the early years of the program recruits often taught without any further training, states have increasingly required that they enter a teacher education program upon hiring and pursue coursework with supervision while they teach. Despite the increasing preparation the recruits receive both from TFA and from the formal teacher education programs most enter upon hiring, the program is often seen as an existence proof for the argument that bright, committed individuals can teach successfully without formal teacher education training. For example, Raymond, Luque, and Fletcher (2002) suggest:

TFA corps members are an admittedly select group of college graduates, culled from the finest universities and often performing near the top of their class.... It’s possible that traditional certification programs and pedagogical training are less necessary for them than they are for the typical teacher. (p. 68)

Several studies have sought to examine the effectiveness of TFA recruits, but none has explicitly compared the effectiveness of differently prepared or certified recruits using appropriate controls for students’ prior learning. Two studies have found evidence that TFA recruits’ students achieve comparable or better gains in student learning when compared to other similarly experienced teachers in similar schools (Raymond, Fletcher, & Luque, 2001; Decker, Mayer, & Glazerman, 2004), but in both of these studies the comparison group teachers were also disproportionately untrained and uncertified teachers. Neither of these studies explicitly compared TFA teachers to teachers with standard training and certification, controlling for other student, teacher, and school variables. A study that examined the relative effectiveness of Teach for America teachers as compared to other new teachers with different levels of qualifications in Arizona found that the students of uncertified teachers, including TFA teachers, did less well on academic tests than those of comparably experienced certified teachers on mathematics, reading, and language arts tests (Laczko-Kerr & Berliner, 2002). However, the study did not use controls for prior achievement of students.

This study examines the question of how teacher preparation and certification influence teacher effectiveness for both TFA and other teachers. We use a newly constructed data set from Houston, Texas that allows us to link detailed certification data on teachers to background and achievement data on students, classrooms, and schools for 132,071 students who were in fourth and/or fifth grade from the 1996–1997 school year through the 2001–2002 school, and their 4,408 teachers. In this article, we report on the results for these students and teachers on several different achievement tests: the TAAS, the SAT-9, and the Aprenda.

Methods

This study substantially replicates the results of an earlier study of TFA recruits in Houston conducted for the Hoover Institution’s CREDO center by Raymond, Fletcher, and Luque (2001). We reconstruct and then go beyond their analyses to examine a wider range of achievement

measures over a greater number of years with additional controls, and we include examination of teacher certification pathways more generally. The CREDO study examined the effect of TFA teachers on student achievement gains on the TAAS reading and mathematics tests from grades 3 through 8 between 1996 and 2000.

The CREDO analyses pooled the data across these years and found that, in most estimates, after controlling for teacher experience (a consistently strong predictor of student achievement), along with individual student, classroom, and school demographics and students' prior achievement scores, TFA recruits were about as effective as other teachers of comparable experience working in similar teaching settings. The study found statistically significant positive coefficients for TFA recruits in 2 of 10 estimates: when the students of TFA recruits were compared to those of beginning teachers with 0–1 years of experience on the TAAS mathematics test in 4th and 5th grades and when the students of TFA teachers were compared to those of other teachers on the TAAS mathematics test in 6th through 8th grades. (Two other estimates, both in reading, were significant at the .10 level.) The effect sizes were relatively small: In most cases the differences represented between 1 and 5% of a standard deviation in the average TAAS test score. In the most positive case, TFA teachers accounted for 14% of a standard deviation difference in the average TAAS test score.

The CREDO study did not examine whether TFA teachers were differentially effective when compared to traditionally prepared and certified teachers in Houston, although the researchers noted that certification status was one of the variables in their data set. This question is an important one for interpreting whether and how the findings of the study may generalize to other settings with different teaching pools, because TFA teachers in Houston were compared to an extraordinarily under-qualified pool of teachers. In 1999–2000, the last year of the CREDO study, about 50% of Houston's new teachers (and one-third of all teachers) were uncertified, and the researchers reported that 35% of new hires lacked even a bachelor's degree.² Furthermore, TFA teachers were placed in schools serving high percentages of low-income and minority students, where most under-qualified teachers in the district are placed, and where, the study found, students lose ground in achievement from year to year. The study's controls for teacher experience and student characteristics at the individual, classroom, and school levels thus had the effect of drawing the comparisons largely among inexperienced and uncertified teachers.

The Data Set

With the assistance of the Houston Independent School District (HISD), we assembled a similar data set.³ The data set consists of information on all HISD teachers and students in grades 3 and higher from the 1995–1996 school year through the 2001–2002 school year (a total of 271,015 students and 15,344 teachers). We created a merged longitudinal data file from several files containing student-level data (demographic characteristics and test scores on three sets of tests in reading and mathematics), teacher data (years of teaching experience, highest degree completed, certification information, and Teach for America participation), school data (student demographic information by school), and identifier data linking students with teachers by school year.

Students in the elementary grades were typically linked with a single teacher, presumably a teacher of a self-contained classroom teaching mathematics, reading, and other subjects. Students in secondary grades were linked with several different teachers both within and across subject areas;

² In a later paper, the lead author indicated that this statistic on degree status was incorrect and that the actual proportion of non-degreed teachers in Houston was likely lower.

³ We could not access the CREDO data set for re-analysis as, we were told, it was a “proprietary” data set.

some had as many as 24 different teachers within a given year. Unfortunately, our data set did not allow us to match all teachers with the subjects they taught or to evaluate why there might be so many links for some students. The CREDO analysis eliminated many middle school students from their analysis because of these difficulties and created a “TFA intensity” ratio for the remaining students who had a TFA teacher as one of several teachers. We had a number of concerns about this methodology; consequently, we decided not to pursue an analysis of teacher effectiveness for grades 6 and above, and we limited our analysis to evaluating individual student gain scores linked to teacher characteristics in grades 3–5. Our analyses measured gains from spring of 3rd to spring of 4th grade and from 4th to 5th grade, looking at effects associated with students’ 4th and 5th grade teachers.

We had a total of 223,086 records on students who were in grades 4 or 5 from 1996–97 through 2001–2002.⁴ (A student is represented by a distinct record for each year he or she is in the data base.) Links to teachers were available for 212,724 of the records.⁵ Most of the students without teacher links were coded as “no show” or “withdrawn” in the district records. In each year from 1996–97 to 2001–02, about 35,000 students were linked with teacher records.⁶

Variables

Outcome Variables

As measures of student achievement, we used student test scores in mathematics and reading on three separate standardized tests administered by Houston during the period studied: the Texas Assessment of Academic Skills (TAAS), the Stanford Achievement Test, 9th Edition (SAT-9), and the Aprenda.

The TAAS is a state-mandated, criterion-referenced test that was administered statewide each spring from 1994 through 2002. The examination was given in grades 3–8 and 10. A Spanish version was available for grades 3–6. For the TAAS, the data provided by HISD contained only two metrics: (1) whether or not students met minimum expectations, and (2) the Texas Learning Index (TLI), a derived continuous score that allows for cross-year and cross-grade comparisons.⁷ Since we were interested in score gains, we used the TLI index. The TLI, however, was available only for the

⁴ This is a subset of the 406,036 records we had on students who were in grades 3 through 5 from 1995-96 through 2001-2002. Third graders were part of the final analyses only if they continued in HISD through 4th grade, in which case their 3rd scores were used as controls in the regressions predicting 4th grade achievement.

⁵ Of the 10,362 records without teacher links, 8115 (78%) were coded with an enrollment status of “no show” or “withdraw.” In contrast, of the 212,724 records with teacher links, only 9.2% had a “no show” or “withdraw” enrollment status code.

⁶ The numbers of grades 4 and 5 students with links to teachers were as follows: 35,667 in 1996-97; 35,566 in 1997-98; 33,914 in 1998-99; 34,498 in 1999-2000; 35,996 in 2000-01; 37,083 in 2001-02.

⁷ The TLI range was approximately between 0 and 100 but differed by subject area and by grade each year. For example, for grade 5 math in spring 2000, the top score was 93, but for grade 5 reading in spring 2000, the top score was 101. More information on the TLI can be found on the Texas Education Agency website at <http://www.tea.state.tx.us/student.assessment/resources/techdig02/index.html> (see Chapter 10, “Scaling”) and <http://www.tea.state.tx.us/student.assessment/reporting/freq/index.html>.

English TAAS and not for the Spanish TAAS.⁸ Thus, our TAAS / TLI analyses apply only to students who took the English TAAS.⁹

Because of concerns raised by other researchers regarding potential score distortions on the high-stakes TAAS examinations (Klein, Hamilton, McCaffrey, & Stecher, 2000), we were interested in alternative measures of student achievement as well. Houston began to administer national norm-referenced tests in 1997–1998. The SAT-9 was administered to 1st through 11th graders who received reading and language arts instruction in English. A Spanish-language test, the Aprenda, was administered to 1st through 9th graders who received instruction in Spanish. The 1997–1998 administration was in the fall; administrations in subsequent years were in the spring. Because we wanted to look at growth over a single school year (e.g., spring to spring), we began our SAT-9 and Aprenda analyses with the first spring administration, in 1998–1999. We used normal curve equivalent (NCE) scores to measure annual changes in student performance.¹⁰

Control Variables

A variety of individual, classroom, and school factors can affect student achievement, and we attempted to control for as many of these factors as possible in testing the influence of teacher certification and Teach for America status. To the control variables included in the CREDO study we added students' English proficiency status, teacher degree levels, and a proxy for class size. The full set included the following:

Student prior achievement. We controlled for prior achievement by including in our regression models each student's prior-year test score. Because our data set begins with third grade students, we looked at student performance for students in fourth and fifth grades, controlling for each student's achievement on the same test a year earlier. The inclusion of the prior year score variable also means that our analyses begin with the second year of data for each achievement measure (1996–1997 for the TLI, and 1999–2000 for the SAT-9 and Aprenda), as the first year of achievement data is used as a control.

Student demographic characteristics: HISD provided data on the following student-level variables: race/ethnicity (American Indian, Asian/Pacific Islander, African American, Hispanic, and white), eligibility for free/reduced price lunch,¹¹ and limited English proficiency (LEP). For LEP status, the HISD data set contained several different codes representing different levels of eligibility.

⁸ In 1995-1996 and 1996-1997, large numbers of students who took the Spanish TAAS were, apparently incorrectly, assigned a TLI of 0. These students were not included in the analysis; we selected only students who took the English version.

⁹ The data also contained a "score code" to indicate such things as exemptions due to absenteeism, disability, and LEP; a code of "S" meant that the score was suitable for inclusion in calculations. Because many of the scores for students with non-S codes were apparently invalid (e.g., they were often coded as the minimum TLI value for a given test in a given grade in a given year), and because of concern about how scores of non-S coded records could be interpreted, we included in our analyses only the scores of students who had a score code of "S."

¹⁰ We would have preferred to use scaled scores for the SAT-9 and Aprenda analyses, but scaled scores were not included in the data we received from HISD.

¹¹ Students with all codes other than "paid" were classified as being eligible for free/reduced price lunch; students with a "paid" code or with no code at all were classified as not being eligible for free/reduced price lunch.

After we conducted exploratory analyses with these finer categories and found that they behaved similarly, we collapsed LEP status into a binary yes/no variable.¹²

Teacher's years of experience and highest degree completed: We used teachers' total years of teaching experience as a continuous variable.¹³ We used two dummy variables to represent highest degree completed: bachelors degree or lower and masters degree or higher.

Classroom level variables: The data obtained from HISD did not contain classroom-level information, but we were able to create classroom-level variables by aggregating up from the individual student-level data, using the teacher identifier to group students. Among the classroom-level variables we included in our analyses is the average prior year score (for all of the grades 3–5 students in the class with prior year scores). This provides an indication of the teaching context and accounts in part for the influence of peers in the learning environment. We also included the number of students in grades 3–5 in the class as a proxy for class size.¹⁴ We considered a variable characterizing the socioeconomic make up of the classroom (the proportion of free/reduced price lunch students), but we found that this variable was highly collinear with both the individual free/reduced price lunch variable ($r=.65$) and the school free/reduced price lunch variable ($r=.83$), as well as being strongly related to the prior year achievement scores on each test. Thus, we kept the individual and school-level free/reduced price lunch variables ($r=.54$), but not the classroom level variable.

School level demographics: The HISD data contained some school-level student demographic variables, including the school's percentages of African American students, Hispanic students, and those eligible for free/reduced price lunch. We used these three variables to capture features of the school that may be relevant both to teaching context and to community characteristics.

¹² In exploratory analyses, we examined students whose LEP code indicated “former” or “tested but did not qualify” in addition to those coded “LEP” or “not LEP,” but in the final analyses, all of these students were classified as being LEP since the variables behaved similarly to the “LEP” code in regressions.

¹³ In our early analyses, we also examined a set of dummy variables for experience as well as an “experience squared” term that takes into account the possible non-linear nature of the experience effect. Neither of these approaches changed our results regarding teacher certification effects, but the “experience squared” term exhibited collinearity problems and made it more difficult to accurately estimate the effects of experience. We therefore opted for the more straightforward continuous experience term (see Fox, 1997).

¹⁴ There were some “classes” of students associated with a given teacher that appeared to have extraordinarily small sizes, as few as one or two students. These could have been individual students placed in classrooms that largely served other grade levels (hence the number of 4th or 5th grade students' records tied to a teacher's record was very small), classes that were primarily tutorial situations for students with particular needs (e.g. one-on-one reading tutorials), or classes that were very small special education classes. Because these would result in teacher effectiveness being evaluated on tiny samples, we restricted the “classes” we included in the estimates to those with at least 15 3rd-5th grade students attached to a given teacher. In addition, because we included a term for the class average previous test scores, intended to account for contextual classroom influences, as well as a term for each individual student's prior test score, we did not want to approach a situation in which an individual student's test score would be weighted disproportionately as a predictor variable (as would be the case if a single student – or a very small number of students - were attached to a given teacher and evaluated as his or her “class”). The number of students excluded by this rule was very small: for example, 699 student records were excluded from the TLI math analyses, accounting for 0.66% of the 106,210 total student records.

Teacher Preparation Pathway and Certification

In exploring the influence of preparation and certification on teacher effectiveness, we had a complex set of variables to examine, represented by more than 100 certification and license codes in use in Houston. These represented both certification categories and subject matter and specialty areas in which teachers held a license. Although we would have liked to have had direct measures of teacher preparation, such as coursework and program measures, we could use these codes to identify many aspects of the pathways teachers pursued into teaching—for example, whether teachers began teaching with a credential or entered as an emergency certified teacher before attaining a standard credential, whether they entered through an alternate certification program, whether they had certification in multiple areas (including specialty areas like bilingual education or reading), and whether they entered teaching through Teach for America. We also examined the interaction between TFA status and certification, since all of the TFA teachers in our sample became certified during their tenure.

Teach for America status is a straightforward yes/no variable. Certification status, on the other hand, is considerably more complex. Houston provided us with two different files containing data on teacher certification, each of which had a different certification coding scheme. One file contained eight different certification types, and the other contained 13 distinct certification codes. (Any given teacher was in one of these files only.¹⁵) These different classifications were related to changes in the state and local certification systems over the study years. In addition to the types of certification, the files included the areas of certification (e.g. elementary, bilingual, reading, music, counselor, etc.). We found that, for elementary teachers, these areas of certification were very similar across teachers with standard certification (for example, most teachers secured a reading endorsement along with an elementary teaching certificate), and after exploring the file, we concluded that adding these additional details would not contribute to our explanation of differences in teacher effectiveness.

Finally, our data included a certification date attached to the records of more than 75% of the teachers in grades 3–5. The certification date was frequently not the year of initial employment, since many teachers enter Houston schools without certification and secure some form of training that leads to a certification or permit later. Since we were concerned with the amount and kind of training a teacher might have in a specific year, we created a year-by-year certification code for each teacher. After investigating the data, we learned that most teachers' certification dates were in the spring and summer months prior to a given school year. We coded teachers who were not certified by the start of a given school year (defined as August 30th) as being uncertified for that year; thus, a teacher's certification code does not “kick in” until the school year following his or her certification date. If this decision were to bias the results, it would be on the side of understating the differences in effectiveness between certified and uncertified (or “not yet certified”) teachers, rather than overstating them.¹⁶

¹⁵ An additional 293 teachers of grades 3–5 were not included in either file. No certification data was available for these teachers, so their certification is categorized as “unknown.”

¹⁶ We adopted a conservative rule for coding certification to be sure we did not artificially overstate the differences between certified and uncertified teachers. This rule meant that teachers who were certified during the course of a school year (after August 30) were coded as uncertified for that year. If those who completed their preparation and/or passed their tests were more effective than others, that greater level of effectiveness would therefore be counted in the uncertified teacher category for that year, thus increasing the potential effectiveness measure for uncertified teachers and reducing the apparent differences between uncertified and certified teachers.

We coded the 1558 teachers in grades 3–5 who had a certification code but no certification date as possessing their certification for all of the years in which they are in the data set. We created a dummy variable, “certification date unknown,” to control for the fact that for this group of teachers, we do not know whether they actually possessed their certification in any given year.

This variable takes a value of 1 for teachers without a certification date and 0 for teachers with a certification date.

After researching Texas’ credentialing system and evaluating our data, we collapsed the certification types into the following categories

Standard includes the standard, provisional, professional, and out-of-state certificates, all of which require the completion of an approved educator preparation program and passage of the appropriate certification examinations.¹⁷

Alternative includes the Texas Alternative Certification Program (ACP) and probationary certificates, which are issued to individuals who have a bachelor’s degree, have passed subject matter tests, and are accepted into approved alternative certification programs in Texas.¹⁸ These alternative and probationary certificates are renewable for up to 3 years while the individual completes the requirements for a standard certificate.

Emergency/temporary includes the categories of emergency permit uncertified, permit (teacher aide), temporary certificate, and recognition. These categories, used either by the state or the Houston district (which has some of its own categories of permits), are issued on a temporary basis to individuals who have not undertaken teacher education and who are supposed to receive mentoring and training while they are teaching.

Certified out-of-field includes individuals who are “emergency permit certified”—that is, already certified but teaching out of their field of certification—and those who hold a temporary classroom permit to teach in a field other than their field of preparation.

Certified, no-test includes individuals holding the school district permit or non-renewal permit granted to those who have completed preparation but not passed the state test.

Uncertified is the designation we gave later-certified individuals during the years before they secured some kind of state or local certification or permit.¹⁹

Certification code missing includes individuals for whom our files held no certification code. In some analyses, we further collapsed these categories into standard certification, alternative or other nonstandard certification, and uncertified.

¹⁷ The provisional and professional certificates were the Texas lifetime certificates granted before 1999 to those who had graduated from an approved teacher education program and passed the certification tests. The professional certificate was granted to individuals with a postbaccalaureate degree. After 1999, these were replaced by the standard certificate, granted for a 5 year renewable term. Out-of-state certificates are granted to individuals who hold the equivalent standard certificate in other states. The specific requirements for each credential can be found in the Texas Administrative Code, Title 19, Part 7, Chapter 230 and Chapter 232, available at [http://info.sos.state.tx.us/pls/pub/readtac\\$ext](http://info.sos.state.tx.us/pls/pub/readtac$ext).

¹⁸ In Houston, about half of ACP holders are enrolled in the HISD intern program. The others attend programs at universities (e.g. Prairie View A&M, University of Houston), through the Regional office, or are enrolled in something called a “deficiency plan,” an individualized program to make up specific needs for the credential.

¹⁹ We actually created six separate “certified later, but not yet” categories, one for each of the certification codes, such that we could tell not only that a teacher was certified later, but specifically which category of certification they later received. In analyses using these variables, we found that the categories all behaved similarly. Thus, in the analyses published here, all of the “certified later, but not yet” teachers were grouped into a single category of “uncertified.”

Although the files included up to eight separate certification codes for each teacher, an analysis of the codes indicated that, for most teachers with multiple codes, all of the codes were in the same category (within the seven classifications listed above).²⁰ Therefore, we based our certification variable on the first certification code assigned to each teacher.

TFA teachers fell into the categories of uncertified, alternatively certified, and standard certified teachers in different years of their teaching careers. Despite the fact that virtually all TFA entrants are placed into the same alternative certification program upon entry, HISD coded only 6 of the 190 4th and 5th grade TFA teacher records as “alternatively certified.” The others were classified as uncertified until they completed their program. All TFA recruits were coded by Houston as “standard certified” when they completed their programs. We performed initial analyses with the separate “uncertified” and “alternatively certified” categories and found that the results were the same for these two groups. We therefore combined all of the TFA candidates prior to receipt of certification into a single category: “uncertified / alternatively certified.” Because our goal was to look at the interaction between TFA status and certification, we examined TFA teachers by their certification category in comparison to non-TFA teachers in each category of certification.

Univariate statistics for all of the variables in the regressions (based on the student records actually included in the regressions) are included in Table A–1 in the Appendix.

Analyses

After merging our several data files and cleaning the resulting data set, we ran a series of descriptive analyses of the characteristics of students and teachers and examined the distribution of teachers to students of different kinds. Then we developed ordinary least squares (OLS) regression estimates of the predictors of six sets of student test scores (three tests with two subject areas for each), both with data pooled across all of our study years and for each year individually.²¹ In each set of estimates, conducted at the individual student level, we controlled for prior year test scores, student race/ethnicity, poverty, and language status; teacher years of experience and degree level; class size and class average previous year’s test score; and school demographics (the percentages of students who were African American, Hispanic, and those eligible for free/reduced price lunch). Key independent variables were teacher certification status and TFA status.

These analyses were helpful for comparing our results to those of the earlier CREDO study, which used similar methods. Because we were also concerned about the influences of classroom and

²⁰ Of the 15,344 total teacher cases in our file, 8756 had only a single certification code. Of the remaining 6588, 2084 had the same certification code for their second certification area as for their first and 4504 had codes that were different only because of the change in state terminology for the standard certificate before and after 1999. The “standard,” “provisional,” and “professional” codes are the different terms given to certificates for fully prepared teachers before and after a major certification reform in 1999. Of the remaining 460, many had non-standard codes that were also equivalent versions of different state and local terms used before and after 1999 (e.g. 67 teachers coded as emergency uncertified/school district permit and 61 coded as ACP/probationary). Thus, we used only a single certification code for each teacher. Among 2104 teachers of grades 3-5 who had first code of “provisional,” 1488 had a second certification code. For 781 of these teachers, the second code is also “provisional”; for 694, the second code is “professional.” In line with the state rules, we categorize all of these as “standard” certificates.

²¹ The numbers of records for 4th and 5th grade student taking – and teachers administering—each of these tests included more than 100,000 student records and 4,000 teacher records for those participating in the TAAS tests over 6 years, more than 60,000 student records and 2,000 teacher records for those participating in the SAT-9 tests over 3 years, and more than 11,000 student records and 750 teacher records for those participating in the Aprenda tests over 3 years.

school contextual effects in a nested data set like this one, we conducted a set of preliminary analyses using hierarchical linear modeling (HLM) techniques, which are presented in Appendix C. They confirm the general findings of the OLS analyses.

Results

Descriptive Statistics

Our data set closely reflects both published data about the Houston Independent School District and the CREDO data. We summarize here data on the 4th and 5th grade students and teachers who are the focus of our analysis. Table 1 shows 4th and 5th grade student enrollments by race/ethnicity, language background, and socioeconomic status for each of the years in our study. Hispanic students in Houston comprise a majority of the population, followed by African American students, who comprise about a third of the student body. About three-quarters of the students are eligible for free or reduced price lunch. Before 1998, about 12% of students were classified as limited English proficient, but the proportion jumped to over one-third from 1998–99 on. Our conversations with Houston staff suggested that this was due to a change in classification practices between 1997–98 and 1998–99, rather than a change in student demographics.

Table 1
Student Demographic Characteristics, Houston ISD, Grades 4–5

Population Group	1996–97	1997–98	1998–99	1999–00	2000–01	2001–02
American Indian	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Asian/Pacific Islander	2.6%	2.7%	2.8%	2.9%	2.9%	2.8%
African American	33.8%	33.7%	33.8%	33.7%	32.8%	32.6%
Hispanic	51.1%	51.9%	52.4%	52.1%	53.3%	54.8%
White	12.4%	11.7%	10.9%	11.3%	10.9%	9.7%
Free/Reduced Lunch	74.7%	72.7%	76.7%	71.1%	72.5%	75.0%
Limited English Proficient	11.0%	12.2%	34.4%	37.5%	40.3%	42.1%
Total (N)	37,396	37,536	34,589	36,622	38,015	38,928

Houston's teaching force for these years was less experienced than most. (See Table 2.) Nationally, the teaching force averaged about 15 years of experience during these years, and about 5 percent of all teachers were brand new to teaching. In Houston, beginning teachers (with less than 2 years of experience) were a large and growing share of the teaching force, increasing from 14% to 23% of teachers from 1996 to 2002. Teachers with six or more years of experience decreased from 64% to 57% during that time. Most teachers held bachelors degrees, with the proportion of such teachers increasing from 68% to 75% from 1996 to 2002 and the share with masters degrees declining at about the same rate, from 31% in 1996 to 23% in 2002.

Table 2
Teacher Experience and Degrees, Houston ISD, Grades 4–5

Trait	1996–97	1997–98	1998–99	1999–00	2000–01	2001–02
Experience						
0–1 year	14.2%	15.0%	19.0%	18.4%	19.2%	22.7%
2–5 years	21.6%	19.8%	16.8%	18.8%	19.9%	20.5%
6–10 years	19.0%	18.0%	18.4%	17.7%	16.5%	15.3%
11+ years	45.2%	47.2%	45.8%	45.1%	44.4%	41.5%
Highest Degree						
< Bachelors	0.5%	0.2%	3.5%	3.8%	5.5%	0.6%
Bachelors	67.6%	69.2%	66.4%	68.7%	67.4%	75.0%
Masters	30.9%	29.9%	28.9%	26.4%	25.5%	23.2%
Doctoral	0.8%	0.7%	1.1%	1.1%	1.6%	1.2%
Total N	1,841	1,802	1,617	1,770	1,805	1,794

Note: Includes all teachers who taught any 4th or 5th grade students.

We also looked at the certification status of TFA and non-TFA teachers for each year of the study. (See Figure 1, below, and Table A–2 in the Appendix.)

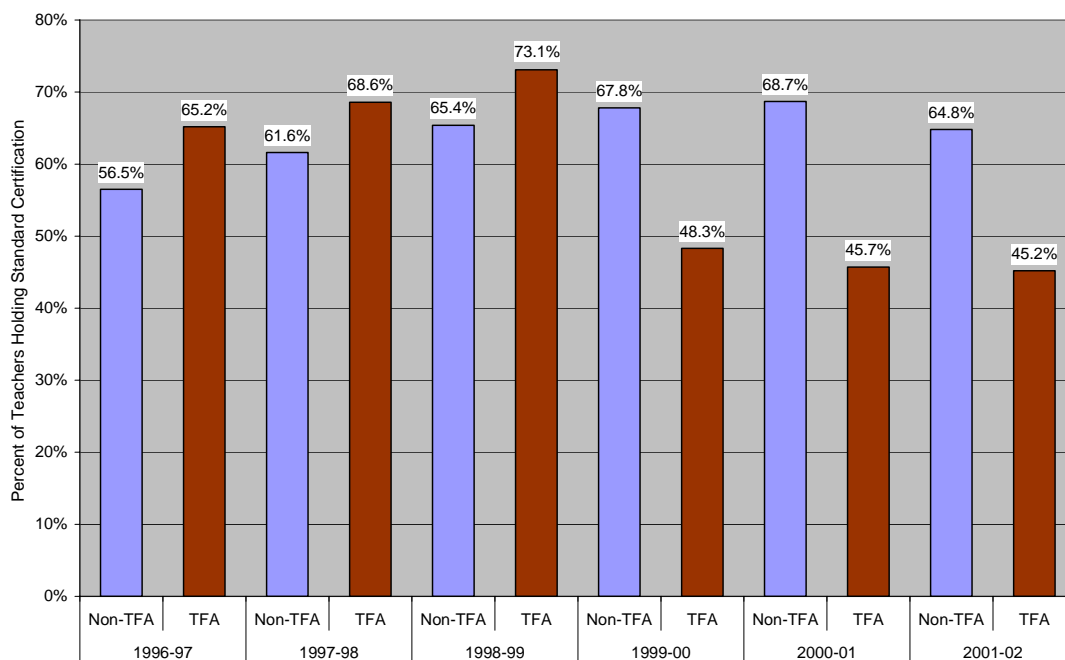


Figure 1
Certification Status of HISD Teachers 1996-2002

Strikingly, in each year of the study, the proportion of teachers in these grade levels teaching without standard certification ranged from one-third to nearly half. In the early years of the study (1996–1998), Teach for America teachers appeared to be as likely as other teachers in these grade levels to hold standard certification.²² From 1999–2001, Teach for America teachers were noticeably

²² We note that about 16% of teacher records with certification codes lacked certification dates in the personnel file for 1996-97, and about 13% lacked certification dates in 1997-98, which makes us unwilling to

less likely to hold standard certification than other teachers. As we describe later, the dramatic shift in the relative qualifications of TFA teachers and other Houston teachers between 1998–99 and 1999–00 may be related to changes in their relative effectiveness as reflected in our regression estimates for different years. These certification patterns are related in part to the experience levels of TFA recruits, who make a two-year commitment to teaching. Most TFA recruits are placed in a teacher education program upon arrival (with few exceptions, this is the Houston alternative certification program), and most are certified after one or two years. In our data set, TFA recruits were certified by their second or third year of teaching.

The Houston program is designed to provide beginning teachers with weekly training sessions run by the district and a mentor at their school, as well as release time to observe other teachers once a month. The recruits take teacher education courses at a local university; in recent years the University of St. Thomas has offered up to six courses in the certification program and an expanded master's degree at a discounted rate for those who want to pursue additional study. The program is designed to be finished in a year; however, many TFA recruits in our data base did not become certified until their third year of teaching. This may be a function of taking additional time to complete courses or pass the requisite certification tests, or because some recruits did not complete the program.

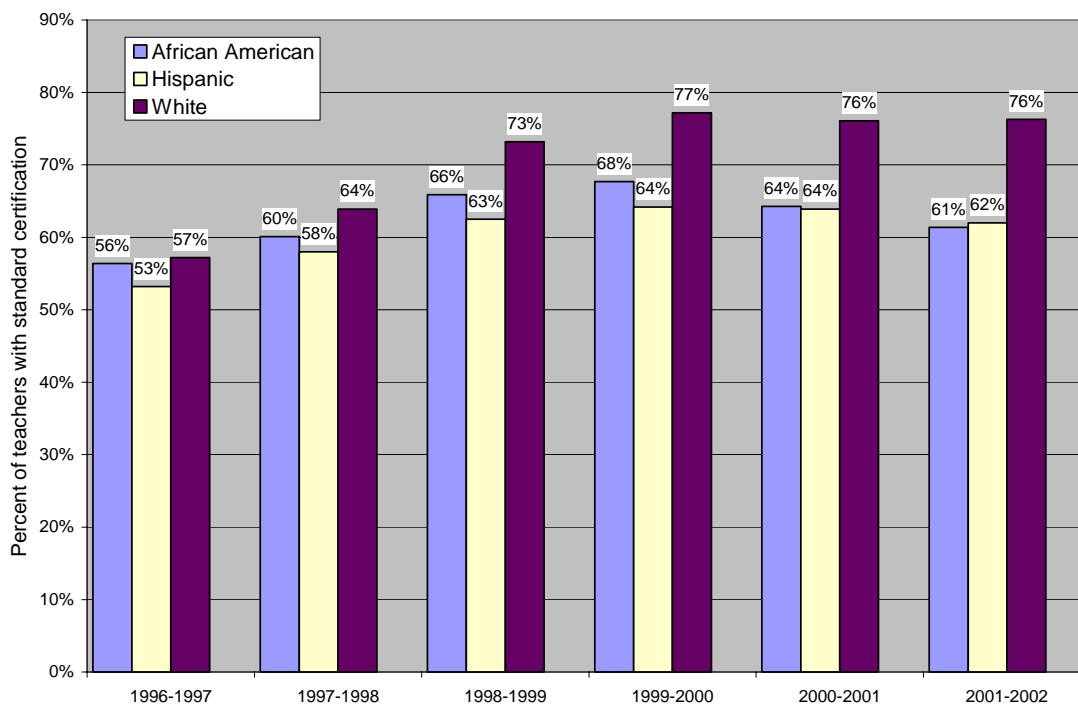


Figure 2
Proportion of HISD Students Taught by Teachers with Standard Certification

We found that teachers without standard certification, including TFA teachers, were disproportionately likely to be teaching African American and Latino students and low-income

draw strong conclusions about when teachers acquired their certification in these years. In subsequent years, the numbers of files with certification codes but without certification dates dropped sharply, totaling only 6.5% in 1998-99, 4.5% in 1999-00, 2.9% in 2000-01, and 1.8% in 2001-02. We have greater confidence in the certification dates for these years.

students. Although the percentages of Houston students being taught by standard-certified teachers rose substantially over the years covered by this study, the racial/ethnic and economic disparities associated with students' access to certified teachers also increased substantially. In 1996–1997, for example, 56% of black students and 57% of white students were taught by standard certified teachers, a difference of less than 1%. By 2001–2002, 76% of white students had standard-certified teachers, while only 61% of black students did—a difference of 15%. (See Figure 2.) Similarly, in 1996–97, 54% of low-income students (those eligible for free or reduced price lunch) had teachers who held standard certification, as compared to 57% of students not eligible for free or reduced price lunch, whereby 2001–02, the proportions were 61% and 72%, respectively. This suggests that as Houston hired and retained greater numbers of certified teachers, these teachers were disproportionately distributed to higher-income students and white students.

Like the CREDO study, we found high rates of attrition for TFA teachers, which were higher than those for other HISD beginning teachers. Raymond and colleagues reported that across the four years of their study, TFA teachers had left teaching in Houston by their third year at rates of between 60% and 100%. We found similar attrition rates: Between 57% and 90% of TFA recruits had left teaching in Houston after their second year, and between 72% and 100% of recruits had left after their third year. (See Table 3.)

Table 3

Attrition Rates of Beginning Teachers, Houston ISD, 1996–97—1998–99 cohorts

Grade level and cohort	After two years		After three years	
	TFA	Non-TFA	TFA	Non-TFA
Grades 4–5				
1996–97 entrants	90.0%	42.6%	100.0%	48.5%
1997–98 entrants	64.3%	30.9%	78.6%	31.6%
1998–99 entrants	57.1%	44.8%	100.0%	54.6%
All Grades				
1996–97 entrants	80.8%	36.4%	96.2%	44.5%
1997–98 entrants	64.0%	23.2%	72.0%	35.3%
1998–99 entrants	57.7%	51.2%	84.6%	54.8%

Thus, although a substantial proportion of TFA recruits became certified within two or three years, few stayed in the district after they had completed their initial preparation for teaching. Generally rates of attrition for TFA teachers were about twice as high as for non-TFA teachers. Attrition rates for newly hired non-TFA teachers ranged between 32% and 55% after three years. We note that all beginning teachers had a somewhat higher attrition rate in 2000–01 (the third year of teaching for 1998–99 entrants), which may have been a function of either reductions in force in Houston or particularly difficult teaching conditions that resulted in many new teachers' leaving.

Analyses of Teacher Effectiveness

The General TFA Effect

Although our data set differed in minor ways from that used by the CREDO research team, and we controlled for several additional variables often found to influence achievement (students' English proficiency, teacher degree levels, and class size), we achieved similar results: When we

examined the effects of TFA status on student achievement gains on the TAAS tests for the pooled years 1996–97 to 2001–02—controlling for prior student achievement, student demographic characteristics, and teacher characteristics—we found that TFA teachers exerted a positive effect on achievement on the TAAS/ TLI in math and a non-significant effect on the TAAS/ TLI in reading. However, when we looked at other test measures and for individual years, the results were quite different. On the SAT-9 and Apenda, TFA teachers had a negative effect on student scores in both math and reading. (See summary in Table 4. The full model is presented in Appendix B–1.)

Table 4

Teach for America effects in multiple regressions, Houston ISD, by Test (pooling across years)

Statistic	TLI	TLI	SAT-9	SAT-9	Apenda	Apenda
	Math	Reading	Math	Reading	Math	Reading
TFA coefficient (Unstandardized)	0.689***	-0.039	-0.882**	-0.642*	-2.87**	-2.58**
T-value	(3.98)	(-0.184)	(-2.74)	(-2.08)	(-2.88)	(-2.61)
Effect Size	0.066	-0.003	-0.046	-0.03	-0.174	-0.156
R ²	.43	.39	.62	.68	.42	.43
N	105,511	103,122	60,488	60,607	11,437	11,436

*p<.05; **p<.01; ***p<.001

T-values are in parentheses. Equations control for student's previous year's test score, student race/ethnicity, free/reduced price lunch status, and LEP status; teacher years of experience, degree level, certification date unknown; class number of students, class average previous year's test score, school demographics (racial/ethnic composition and poverty).

We found that TFA effects varied by year of the tests. (See Table 5.)

Table 5

Teach for America effects in multiple regressions, Houston ISD, by Test and Year

Year	TLI	TLI	SAT-9	SAT-9	Apenda	Apenda
	Math	Reading	Math	Reading	Math	Reading
1996–97	1.48** (2.77)	-0.075 (-0.118)	—	—	—	—
1997–98	1.34** (2.85)	0.300 (0.569)	—	—	—	—
1998–99	2.83*** (5.27)	2.42*** (3.76)	—	—	—	—
1999–00	-0.727~ (-1.62)	-1.16* (-2.08)	-0.769 (-1.17)	-1.34* (-2.12)	-1.04 (-0.752)	1.04 (0.739)
2000–01	0.286 (0.957)	-0.690 (-1.62)	0.152 (0.270)	0.426 (0.804)	-3.99* (-2.38)	-2.73 (-1.64)
2001–02	-0.353 (-1.47)	-0.972** (-2.70)	-1.61** (-3.33)	-0.904~ (-1.91)	-6.21** (-3.26)	-8.71*** (-4.06)

~ p < .10; *p < .05; **p < .01; ***p < .001

T-values are in parentheses. Equations control for: student's previous year's test score, student race/ethnicity, free/reduced price lunch status, and LEP status; teacher years of experience, degree level, certification date unknown; class number of students, class average previous year's test score, school demographics (racial/ethnic composition and income).

On the TAAS math tests, the positive effects found in the overall analysis by both CREDO and us held only for the years 1996–97 through 1998–99. Starting in 1999–2000, the point at which TFA teachers were noticeably less likely to be certified than other teachers in these grade levels, the TFA coefficient in math became non-significant. In reading on the TAAS, where the TFA coefficient in the pooled years analysis had been non-significant, the TFA coefficients were significant and positive in 1998–99, the year in which TFA recruits were much more likely to be certified than other Houston teachers. (See Table A–2 in the Appendix.) The coefficients in reading were significant and negative in two of the three years between 1999–00 and 2001–02, when TFA recruits were much less likely than other Houston teachers to be certified. On the SAT-9 and Aprenda, which were given only in the years after 1998–99, TFA coefficients were non-significant or negative. By 2001–02, the coefficients for TFA were negative across the board.

Thus, examined across measures and individual years, the TFA coefficient is positive on the TAAS in mathematics only in the first three years of the six years of data we analyzed and on the TAAS in reading only in 1998–99, the year when TFA recruits were better qualified than other Houston teachers.²³ In this year, 73% of TFA teachers had standard certification, as compared to only 65% of other Houston teachers. This is probably because that year's TFA cohort included a greater share of more experienced members who had completed their preparation. In the following year (1999–2000), however, only 48% of TFA recruits held standard certification as compared to 68% of other Houston teachers. At this point and thereafter the observed effect of TFA recruits relative to other teachers shifted to non-significant or negative on each test.

These year-by-year analyses point out how the TFA effect varies according to the relative qualifications of TFA candidates and others in Houston schools. We pursue this further in the analysis below that examines TFA status by certification status.

The differences we observed in the influences of TFA status across tests may have been a function of differences in the tests. For example, the SAT-9 is generally considered a more rigorous test more focused on higher level thinking skills than the TAAS, and because it was not a high-stakes test in Houston, it may have been less subject to distortions caused by teaching to the test. (We note also that our equations predicted much more of the variance in the SAT-9 tests than the TAAS, which may also be a function of the different positions of the two tests in the accountability system in Houston.) It is also possible, though, that the differences in outcomes were a function of the years in which the tests were offered, since the SAT-9 and Aprenda were administered and analyzed in the years after 1998–99, when TFA recruits were, as a group, less well qualified, and were also found to be less effective than other teachers in producing student gains on the TAAS.

The Effects of Certification

Using the same basic models, we also examined the relationship between teacher certification and teacher effectiveness. First we looked at the effects of different certification categories on teacher effectiveness, irrespective of TFA status. Then we examined the interaction of certification and TFA status.

We found that, relative to teachers with standard certification, uncertified teachers and those in most other non-standard certification categories generally had negative effects on student achievement, after controlling for student characteristics and prior achievement, as well as teacher experience and degrees. Uncertified teachers showed significant negative effects across five of the

²³ Because the sample sizes for TFA candidates are relatively small in any given year (about 30 recruits), we did not calculate effect sizes for these coefficients.

six tests. Teachers with missing certification codes (who are likely to be uncertified) also showed significant negative effects on student achievement on four of six tests. Similarly, teachers who were certified without passing the state teacher certification tests (a special permit category in HISD: such teachers would be classified as uncertified in other districts that do not offer such a permit) showed negative effects across four of six tests, plus one more at the .10 level of significance. (See Table 6 for a summary and Table B–2 in the Appendix for the full model.)

Table 6

Teacher Certification Status and Student Achievement Gains, Houston ISD Grades 4–5, by Test (pooled across years)

Certification status	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
Uncertified	-0.525*** (-8.03) [-0.05]	-0.580*** (-7.23) [-0.05]	-0.414** (-3.30) [-0.02]	-0.516*** (-4.31) [-0.03]	-1.41*** (-4.61) [-0.09]	-0.066 (-0.216) [-0.003]
Alternative certification	-0.897* (-2.02) [-0.09]	-0.818 (-1.49) [-0.07]	-2.31** (-3.29) [0.12]	-1.40* (-2.09) [-0.07]	-0.491 (-0.454) [-0.03]	3.13** (2.91) [0.19]
Emergency/temporary certification	-0.701*** (-3.85) [-0.07]	-0.690** (-3.07) [-0.06]	-0.636~ (-1.68) [-0.03]	0.780* (2.16) [0.04]	-1.40 (-1.15) [-0.08]	-0.754 (-0.619) [-0.05]
Certified, but out-of-field	0.667* (2.52) [0.06]	0.902** (2.77) [0.07]	-1.80*** (-4.36) [-0.09]	0.216 (0.548) [0.01]	-3.96** (-2.64) [-0.24]	3.64* (2.45) [0.22]
Certified, no test	-0.187 (-0.655) [-0.02]	-0.655~ (-1.87) [-0.05]	-4.49*** (-7.06) [-0.23]	-2.34*** (-3.83) [-0.12]	-3.52*** (-3.66) [-0.21]	-4.76*** (-5.01) [-0.29]
Certification code missing	0.113 (0.549) [0.01]	-0.642* (-2.53) [-0.05]	-2.26*** (-4.63) [-0.12]	0.121 (0.260) [0.01]	-4.78*** (-3.84) [-0.29]	-4.37*** (-3.52) [-0.26]
R ²	.43	.40	.62	.68	.42	.43
N	105,511	103,122	60,488	60,607	11,437	11,436

~p<.10, * p < .05, ** p < .01, ***p< .001. T-values are in parentheses; effect sizes are in brackets. The reference group is standard-certified teachers. Equations control for student's previous year's test score, student race/ethnicity, free/reduced price lunch status, and LEP status; teacher years of experience, degree level, certification date unknown; class number of students, class average previous year's test score, school demographics (racial/ethnic composition and poverty).

The influences were a bit more mixed in the other categories. Alternatively certified teachers had negative effects on achievement on three tests. On the Aprenda in reading, alternatively certified teachers had a significant positive effect. Since the Houston alternative certification program enrolls a substantial number of Hispanic teachers, it may be that more of these teachers are Spanish-speaking and able to support the literacy progress of Spanish-speaking students who take the Aprenda. Teachers on emergency or temporary certificates showed negative effects on student achievement on three tests; however, they showed a positive effect on the SAT-9 in reading. Interestingly, teachers who were already certified but credentialed to teach out of field had significant positive effects on two out of three reading tests and one mathematics test (the TAAS/TLI in reading and math and the Aprenda in reading, with a positive coefficient on the SAT-

9 in reading as well), but significant negative effects on the other two math tests (the SAT-9 and Aprenda).

Overall, teachers without certification or with non-standard certification were found to be less effective in raising student test scores than teachers with standard certification in 22 of 36 estimates ($p < .10$). In general, relative to teachers with standard certification, teachers lacking full certification slowed student progress over the course of a year by about $\frac{1}{2}$ to 1 month in grade equivalent terms on most achievement tests. However, some categories of teachers with substandard certification (those who had not passed the certification tests or who had no record of being certified) had an even larger negative effect on the Spanish-speaking students who took the Aprenda, slowing their progress by 2 to 3 months within a year in comparison to the progress they would be expected to make with a fully certified teacher. The effects of certification status were generally much stronger than the effects of teacher experience. For example, on the SAT-9 and Aprenda tests, the positive effect of an additional year of teacher experience was about one-tenth the size of the effect of having a fully certified teacher.

The Combined Effects of TFA and Certification Status

In Table 7 we display the results for Teach for America teachers within different certification categories considered in relation to non-TFA standard certified teachers and other differently certified non-TFA teachers. (The full model is presented in Table B-3 in the Appendix.) Again we found that uncertified teachers and those with less than standard certification—whether TFA or non-TFA—exert negative effects on student achievement relative to teachers with standard certification.²⁴ Uncertified TFA teachers showed significant negative effects on student achievement in five of six estimates (and the sixth also has a negative coefficient.) The same was true for uncertified teachers who were not members of Teach for America. For non-TFA teachers, those with nonstandard certifications (alternative, emergency, temporary, or certified without having passed the test) also showed negative influences on achievement that are significant at the .10 level or below in four of six estimates.

On all tests but one (the TAAS math test), the negative effect of having an uncertified TFA teacher was greater than the negative effect of having an uncertified or nonstandard certified teacher who was not recruited through TFA, depressing student achievement by between one-half month to 3 months annually compared to a fully certified teacher. For other categories of less than fully certified teachers the negative effects generally ranged from about 0.2 month to 1.5 months, depending on the test. Because certification status may be correlated with experience, especially in the first year or two, we also conducted estimates using dummy variables for experience, including a control for teachers with 0 to 1 years of experience, and got similar results with similar effect sizes.²⁵

²⁴ All TFA teachers had one of three certification types: uncertified or alternatively certified before they completed their certification program and standard certified after they had completed the program. We combined the six TFA teachers who were coded by HISD as “alternatively certified” with the others who were not yet certified for the year(s) before they received standard certification. We grouped all other teachers into the following categories: uncertified, alternative plus all other nonstandard certification categories, and standard certified (including certified, out-of-field).

²⁵ We used experience categories 0-1 years, 2-5 years, 6-10 years, and 11 or more years. In all cases, teachers with 0 to 1 year of experience had a strong negative effect on student achievement. After controlling for this, uncertified TFA teachers had a negative effect in 5 estimates (one of these was marginal at $p < .10$) and a non-significant effect in one estimate relative to non-TFA standard certified teachers. Uncertified non-TFA teachers also had a negative effect in 5 estimates and a non-significant effect in one.

Table 7
Teacher Certification Status, TFA Status, and Student Achievement Gains, Houston ISD Grades 4–5, by Test (pooled across years)

Certification and TFA Status	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprinda Math	Aprinda Reading
TFA teacher, uncertified ^a	-0.288 (-1.14) [-0.028]	-1.22*** (-3.93) [-0.099]	-1.89*** (-4.41) [-0.099]	-1.09** (-2.67) [-0.055]	-4.06** (-3.01) [-0.246]	-5.09*** (-3.81) [-0.301]
TFA teacher, standard certified	1.18*** (4.93) [0.113]	0.437 (1.49) [0.035]	-0.008 (0.017) [-0.0004]	-0.458 (-0.994) [-0.023]	-2.55~ (-1.73) [-0.154]	0.839 (0.572) [0.051]
Non-TFA teacher, uncertified	-0.488*** (-7.36) [-0.047]	-0.550*** (-6.76) [-0.045]	-0.341** (-2.68) [-0.018]	-0.519*** (-4.25) [-0.026]	-1.33*** (-4.30) [-0.080]	-0.014 (-0.045) [-0.001]
Non-TFA teacher, nonstandard certified	-0.284 (-1.63) [-0.027]	-0.605** (-2.81) [-0.049]	-1.54*** (-4.60) [-0.081]	-0.051 (-0.159) [-0.003]	-2.45*** (-3.53) [-0.148]	-1.26~ (-1.83) [-0.076]
R ²	.43	.40	.62	.68	.42	.43
N	105,511	103,122	60,488	60,607	11,437	11,436

~p<.10, * p < .05, ** p < .01, ***p< .001. T-values are in parentheses; effect-sizes are in brackets.

^aIncludes the six TFA teachers coded in the alternative certification category. All of these teachers were coded by HISD and by us as standard certified when they completed their certification program. Reference group is non-TFA standard-certified teachers. Equations control for student's previous year's test score, student race/ethnicity, free/reduced price lunch status, and LEP status; teacher years of experience, degree level, certification date unknown; class number of students, class average previous year's test score, school demographics (racial/ethnic composition and poverty).

Relative to other teachers with standard certification, TFA teachers with standard certification did about as well, with only two statistically significant differences. As signaled in the CREDO analyses and ours, certified TFA teachers did better than other standard certified teachers in supporting student achievement on the TAAS test in mathematics, increasing student achievement by just over 1 month in grade equivalent terms during the course of a year. On the other hand, TFA teachers' students did marginally worse than students of other standard certified teachers (p<.10) on the Aprinda in mathematics, lagging in achievement by about 1.5 months over a year's time.

In another analysis of these data, summarized in Appendix C, hierarchical linear modeling (HLM) techniques were used to take into account the nesting of students within classrooms and of classrooms within schools. The findings of this analysis were similar to the OLS findings: 4th and 5th grade students taught by uncertified TFA teachers performed less well on three tests (TAAS reading, SAT-9 reading, and Aprinda math) than the students of non-TFA standard certified teachers (controlling for student prior test scores, teacher experience and degrees and student / school characteristics); coefficients on the other three tests were also negative but not significant. Non-TFA uncertified teachers had negative effects, relative to non-TFA standard certified teachers, on four of the six tests and insignificant effects on the other two. Students of certified TFA teachers performed comparably to students of other certified teachers on all six tests, with no statistically significant differences.

Overall, then, teachers' abilities to support student achievement appear to depend, both for TFA teachers and others, substantially on the level of preparation these teachers have had, as reflected in their certification status.

Discussion

Although a number of studies have found that students taught by fully certified teachers appear to achieve at higher levels, few have been able to examine individual student-level data over multiple years on multiple measures with appropriate controls. Previous studies of Teach for America, as a specific pathway into teaching, have either failed to control for certification status or for students' prior achievement in examining the outcomes of this program on teachers' effectiveness. Our ability to look at these questions using a large data set that represents these teacher variables and a range of student, classroom, and school controls has provided a unique opportunity to evaluate how teacher education and pathways into teaching may influence teacher effectiveness.

Of course, certification is only a proxy for the real variables of interest that pertain to teachers' knowledge and skills. These include knowledge of the subject matter content to be taught and knowledge of how to teach that content to a wide range of learners, as well as the ability to manage a classroom, design and implement instruction, and work skillfully with students, parents, and other professionals. In Texas, teachers who have achieved standard certification are required to have passed tests of core academic skills in communications and mathematics, tests of specialized subject matter knowledge, and tests of pedagogical knowledge. They also have completed an approved teacher education program which includes specified courses in the content area(s) to be taught as well as coursework in teaching and learning; instructional methods and strategies; classroom management; curriculum; measurement and evaluation of student learning; human growth and development; multicultural education; the education of special needs students; legal and ethical aspects of teaching; organization of schools; technology; and the teaching of reading (Texas Administrative Code, Title 19, Part 7, Rule 230.191, 2004).

This array of requirements, in combination, appears to make a difference in teacher effectiveness. Like other studies cited earlier, we find that 4th and 5th grade teachers in Houston who hold full certification—the professional or standard certificate Texas awards to recruits who have graduated from an approved teacher education program—are more effective than other teachers in stimulating student achievement gains in both reading and mathematics on three different test batteries over a multi-year period. This relationship holds whether the teachers are recruited through Teach for America or through other pathways. Those who have completed the training that leads to certification are more effective than those who have not.

Although some have suggested that perhaps bright college graduates like those who join TFA may not require professional preparation for teaching, we found no instance where uncertified Teach for America teachers performed as well as standard certified teachers of comparable experience levels teaching in similar settings. In the OLS estimates, on 5 of 6 tests, uncertified TFA teachers showed a significant negative effect on student achievement gains relative to standard certified teachers. (The sixth coefficient was also negative but non-significant.) The effect sizes are noticeable: Over the course of a year, students taught by uncertified TFA teachers could be expected to achieve at levels that are, in grade equivalent terms, one-half month to 3 months lower than students taught by teachers with standard certification. Those taught by other teachers who are uncertified or who hold nonstandard certification generally achieve at levels 0.2 to 1.5 months behind their counterparts taught by standard certified teachers. Students in the most impacted

schools, who have a steady parade of such teachers each year, would generally lose 1 to 2 years of ground in grade equivalent terms between kindergarten and 6th grade, assuming the effects we found for 4th and 5th grades generalize to other grade levels.

At the same time, Teach for America teachers who had achieved standard certification generally performed on a par with other certified teachers, after controlling for degrees and experience, as well as a variety of student and school factors. They were more effective than other certified teachers on one of the six measures we examined (the TAAS mathematics test) and marginally less effective ($p < .10$) on another (the Aprenda mathematics test).

We were able to confirm these general findings with an HLM analysis that essentially replicated our analyses of certification and TFA status on student achievement using data pooled across years. However, it would be useful in future work to use hierarchical linear modeling in a longitudinal framework that tracks teachers' effectiveness across years or to explore functional data analyses that allow consideration of teachers' developmental career paths as they unfold over time. This would allow consideration of a number of other questions: Do some categories of teachers become more effective than others as they gain education and experience? How much of the apparently stronger performance of groups of teachers as they achieve certification or gain experience is a function of enhanced effectiveness and how much is actually a selection effect caused by weaker candidates dropping out of the data set as they leave teaching after a period of time? What are the cumulative effects for students of having different kinds of teachers with different configurations of training and experience over multiple years?

The Successes and Limitations of TFA and Other Pathways in the Houston Context

This study, in combination with the findings of the CREDO study, suggests some successes and limitations of both the Teach for America program and other alternative programs in Houston. Across the country, Teach for America operates only in districts that, for a variety of reasons, hire many uncertified teachers. During the years studied, Houston was such a district, although there were improvements in the recruitment of certified teachers over the years studied (from about 56% of teachers in 1996 to 67% in 2001).

Our analyses suggest that in contexts where many teachers have little preparation for teaching and where there is high turnover, TFA may make a positive contribution. The Teach for America organization often notes that its goal is to bring stability for at least one or two years to classrooms in poor and minority schools that might otherwise have a parade of substitute teachers, and argues that its recruits do as well as other teachers these students might have. Given the likelihood that these students would otherwise have equally inexperienced and uncertified teachers, this claim seems to be at least partially supported by our data. Entering TFA teachers appear to perform about as well as other uncertified teachers in Houston on at least some tests, after controlling for experience, degree status, and student characteristics. Most of them stay for two years, which may provide a modest degree of stability to schools that might otherwise experience an even more quickly revolving door for teachers in and out of classrooms.

It might also be argued that the reputedly strong liberal arts background of TFA teachers may contribute to their students' relatively better showing on the TAAS mathematics tests. On this one test—though not on the other two mathematics tests used in Houston (the SAT-9 and the Aprenda)—the students of fully certified TFA recruits performed significantly better than the students of other certified teachers. Given the longstanding concerns about the mathematics background of many elementary school teachers, it would be plausible that candidates who have attended relatively selective colleges would have a stronger basic mathematics background in high school and college than the average elementary teaching candidate. That this effect did not hold up

on the SAT-9 and the Aprenda may be a function of differences in what the tests measure or of the quality of the TFA cohorts in the later years of our study, when the SAT-9 and Aprenda were administered.

The strength of TFA cohorts may differ from year to year, as the program's recruitment and training practices fluctuate. The strongest positive TFA effects were in 1998–99, when TFA candidates were much more likely to be certified than the average Houston teacher. The most negative year was 2001–02, when TFA recruits had negative effects on student achievement on 5 of 6 tests. In addition to the fact that TFA teachers were much less likely than other Houston teachers to be certified in this year, there may have been other selection or training effects operating. The TFA program has expanded rapidly in recent years and may have been less able to be highly selective in recent years as its numbers have grown.

Indeed, as the proportion of fully certified teachers in Houston grew over the years in our study, and as the proportion of TFA teachers who were certified declined, TFA cohorts went from being better qualified on average to less well-qualified on average than other HISD teachers. We found that interpreting the relative influence of TFA teachers on student achievement depends on knowing about the characteristics of the comparison group of teachers. When compared to less well qualified teachers, TFA teachers appeared to have a neutral or positive effect. When compared to a pool of teachers who were on average better qualified, TFA teachers appeared to have a negative effect. Thus, one could anticipate that the relative effectiveness of this or any other group of teachers must be evaluated in a specific context at a particular point in time, with close understanding of the qualifications and characteristics of the comparison groups of teachers employed in that same setting at that point in time.

The limitations of teachers without preparation are illuminated by this study. It is clear that, across the board, Houston students achieved stronger achievement gains in both reading and mathematics when they were taught by standard certified teachers rather than uncertified teachers. Uncertified teachers, both TFA and non-TFA, did particularly poorly with Spanish-speaking students who took the Aprenda tests. This might be a function of the specialized knowledge needed to teach English language learners that may be more consistently acquired in pre-service teacher education programs. These programs may have more time to teach not only the basics of classroom management and lesson planning, but also the strategies for teaching content to students who have specific language needs.

Alternatively certified teachers, although they were generally less effective than standard certified teachers, had a strong positive effect on student achievement for Spanish-speaking students on the Aprenda in reading. This may be because Houston's alternative certification program recruits a large number of Hispanic candidates, whose Spanish language skills may help them teach reading in Spanish more effectively than many other teachers.

Finally, although students taught by TFA recruits (and other uncertified teachers) were slowed in their academic progress in the first year of the recruits' teaching efforts, one of the accomplishments of Teach for America and HISD was the development of means for enabling recruits to participate in preparation and become certified for teaching in their second or third year of teaching. Houston students did not reap long-term benefits from these efforts, however, as the vast majority of the TFA recruits left after their second or third year of teaching. High turnover of beginning teachers is both extremely costly for school districts (Benner, 2000) and counterproductive for students, as teacher effectiveness typically increases markedly after about the second year of teaching (Hanushek, Kain, & Rivkin, 1998).

Implications for the Recruitment and Preparation of Teachers

These findings raise several considerations for districts seeking to design programs and pathways for recruiting and preparing teachers. First, the results suggest that, in the best of circumstances, there are benefits to recruiting fully prepared teachers who can launch their careers at a higher level of effectiveness. However, absent a set of policies that can provide a fully prepared teaching force, the study findings also suggest that candidates who enter alternative programs like the one offered by HISD can become more effective as they complete the program and become certified over a period of two or three years.

We noted that recruits who enter teaching through alternative programs and pathways may bring different potential strengths with them. We speculated, for example, that TFA teachers' potentially stronger grounding in mathematics may have contributed to their greater effectiveness in supporting student achievement in mathematics in some years and that HISD alternative program recruits may have been more effective in supporting Spanish-speaking students in reading because many are Spanish-speaking teachers. However, these benefits did not hold across different tests and areas. This suggests that the personal characteristics sought by alternative programs are only a starting point for training. Programs should seek to build on the skills candidates bring with them while ensuring that other areas of knowledge and skill are explicitly addressed and developed.

For students to benefit from their teachers each year, alternative programs and other nontraditional pathways into teaching should ensure that new recruits-in-training can practice under the close supervision of expert veterans, so that their students—from their very first days in the classroom—have the benefit of a classroom informed by a prepared teacher's advice and counsel, rather than being put at a disadvantage. Even if such teachers improve their effectiveness with training over several years, the students have only one opportunity to experience second grade, for example, and cannot afford to lose ground in acquiring basic skills even for a single year. And students in high-need schools in cities like Houston are likely to experience many beginning teachers who are not yet prepared, creating cumulative negative effects on their achievement, if they are not offset or supported by teachers with more expertise.

Finally, as alternative pathway teachers become more expert, students and schools gain benefits only if the teachers stay in the schools that have invested in their training. Cost-effective recruitment programs that increase the share of prepared teachers students ultimately encounter will likely need to recruit candidates with the expectation of a longer teaching commitment than the two years TFA candidates currently pledge. The North Carolina Teaching Fellows program, for example, recruits high-ability students into teacher education by providing service scholarships that cover the full costs of high quality pre-service training, repaid by at least 4 years of service in public schools. An evaluation found more than 75% still teaching after seven years, and many of the remainder were still in public schools as administrators (National Commission on Teaching and America's Future, 1996).

A broader challenge for states, school districts, and teacher preparers is how to develop and expand the reach of strong, efficient, and affordable preparation routes that enable teachers to be competent when they enter teaching and that retain teachers as they become more effective. In addition to successful service scholarships like those offered in North Carolina and some other states, the literature includes examples of urban teacher education programs that have strong records of preparing capable teachers who stay in the city schools (see for example, Darling-Hammond, Chung, & Frelow, 2002; Darling-Hammond & MacDonald, 2000; Koppich, 2000; Snyder, 2000; Zeichner, 2000). Increasing the availability of such programs could help stem turnover, as several recent studies have found that teacher attrition is strongly related to the extent of preparation

teachers have had upon entry (Henke, Chen, Geis, & Knepper, 2000; National Commission on Teaching and America's Future, 2003).

However, analyses of urban districts that have resolved teacher shortages indicate that additional state and local policies are needed to create the labor market conditions required to hire and retain an adequate supply of prepared teachers. These typically include aggressive outreach and streamlined hiring systems, training subsidies and partnerships with local universities, and recruitment incentives, as well as competitive salaries, reasonable working conditions, and supportive administrators (Darling-Hammond & Sykes, 2003; Murnane, Singer, Willett, Kemple, & Olsen, 1991; National Commission on Teaching and America's Future, 2003). Thus, improvements in teachers' preparedness are likely to rest on a policy strategy that also includes the right mix of incentives for recruiting and retaining qualified teachers.

References

- Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record*, 102(1), pp. 1–27.
- Benner, A. D. (2000). *The cost of teacher turnover*. Austin, TX: Texas Center for Educational Research.
- Betts, J.R., Rueben, K.S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California*. San Francisco: Public Policy Institute of California.
- Darling-Hammond, L. (2000a). Reforming teacher preparation and licensing: Debating the evidence. *Teachers College Record*, 102(1), pp. 28–56.
- Darling-Hammond, L. (2000b). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved October 3, 2005, from <http://epaa.asu.edu/epaa/v8n1>.
- Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Variation in teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education*, 53(4): 286–302.
- Darling-Hammond, L., & MacDonald, M. (2000). Where there is learning there is hope: The preparation of teachers at the Bank Street College of Education. In L. Darling-Hammond (Ed.), *Studies of excellence in teacher education: Preparation at the graduate level* (pp. 1–95). Washington, DC: American Association of Colleges for Teacher Education.
- Darling-Hammond, L. & Sykes, G. (2003). Wanted: A national teacher supply policy for education: The right way to meet the ‘highly qualified teacher’ challenge. *Educational Policy Analysis Archives*, 11(33), retrieved October 3, 2005, from <http://epaa.asu.edu/epaa/v11n33>.
- Darling-Hammond, L. & Youngs, P. (2002). Defining “highly qualified teachers:” What does “scientifically-based research” actually tell us? *Educational Researcher*, 31(9), 13–25.
- Decker, P.T., Mayer, D.P., & Glazerman, S. (2004). *The Effects of Teach For America on Students: Findings from a National Evaluation*. Princeton, NJ: Mathematica Policy Research, Inc.
- Ferguson, R.F. (1991, Summer). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28(2), 465–498.
- Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, 7(9), retrieved October 3, 2005, from <http://epaa.asu.edu/epaa/v7n9>.

- Finn, C.E. (1999). Foreword. In M. Kanstoroom & C. E. Finn, Jr. (eds.), *Better teachers, better schools* (pp. v–vii). Washington, DC: The Thomas B. Fordham Foundation.
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: Sage Publications.
- Goe, L. (2002). Legislating equity: The distribution of emergency permit teachers in California. *Educational Policy Analysis Archives*, 10(42), retrieved October 3, 2005, from <http://epaa.asu.edu/epaa/v10n42>.
- Goldhaber, D.D., & Brewer, D.J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129–146.
- Gomez, D. L., and Grobe, R. P. (1990). *Three Years of Alternative Certification in Dallas: Where are we?* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.
- Hanushek, E., Kain, J., & Rivkin, S. (1998). *Teachers, schools, and academic achievement*. (Working Paper 6691). Cambridge, MA: National Bureau of Economic Research.
- Hawk, P.P., Coble, C.R., & Swanson, M. (1985). Certification: It does matter. *Journal of Teacher Education*, 36(3), 13–15.
- Henke, R.R., Chen, X., Geis, S., & Knepper, P. (2000). *Progress through the teacher pipeline: 1992–93 college graduates and elementary/secondary school teaching as of 1997*. NCES 2000–152. Washington, DC: National Center for Education Statistics.
- International Reading Association (2003). *Prepared to make a difference: Research evidence on how some of America's best college programs prepare teachers of reading*. Newark, NJ: International Reading Association.
- Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stetcher, B.M. (2000). *What do test scores in Texas tell us?* Santa Monica: The RAND Corporation.
- Koppich, J. (2000). Trinity University: Preparing teachers for tomorrow's schools. In L. Darling-Hammond, (Ed.), *Studies of excellence in teacher education: Preparation in a five-year program* (pp. 1–48). Washington, DC: American Association of Colleges for Teacher Education.
- Laczko-Kerr, I., & Berliner, D. (2002). The effectiveness of Teach for America and other under-certified teachers on student academic achievement: A case of harmful public policy. *Educational Policy Analysis Archives*, 10(37), retrieved October 3, 2005, from <http://epaa.asu.edu/epaa/v10n37>.
- Monk, D.H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.

- Murnane, R., Singer, J., Willett, J., Kemple, J., & Olsen, R. (1991). *Who will teach? Policies that matter*. Cambridge, MA: Harvard University Press.
- National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York: Author.
- National Commission on Teaching and America's Future. (2003). *No dream denied*. Washington, DC: Author.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Ed.). Thousand Oaks, CA: Sage Publications.
- Raymond, M. & Fletcher, S. H. (2002). The Teach for America evaluation. *Education Next*, 2(1): 62–68.
- Raymond, M., Fletcher, S.H., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: The Hoover Institution, Center for Research on Education Outcomes.
- Shields, P.M., Esch, C.E., Humphrey, D.C., Wechsler, M.E., Chang-Ross, C.M., Gallagher, H.A., Guha, R., Tiffany-Morales, J.D., & Woodworth, K.R. (2003). *The status of the teaching profession 2003*. Santa Cruz, CA: The Center for the Future of Teaching and Learning.
- Snyder, J. (2000). Knowing children—understanding teaching: The developmental teacher education program at the University of California, Berkeley. In L. Darling-Hammond (Ed.), *Studies of excellence in teacher education: Preparation at the graduate level* (pp. 97–172). Washington, D.C.: American Association of Colleges for Teacher Education.
- Strauss, R. P. and Sawyer, E.A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, 5(1): 41–48.
- U.S. Department of Education (2002). *The Secretary's report on teacher quality*. Washington, DC: U.S. Department of Education.
- Wenglinsky, H. (2000). *How teaching matters: Bringing the classroom back into discussions of teacher quality*. Princeton, NJ: Educational Testing Service.
- Wilson, S.M., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. A research report prepared for the U.S. Department of Education. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- Zeichner, K. M. (2000). Ability-based teacher education: Elementary teacher education at Alverno College. In L. Darling-Hammond, (Ed.), *Studies of excellence in teacher*

education: Preparation in the undergraduate years (pp. 1–66). Washington, D.C.: American Association of Colleges for Teacher Education Publications.

About the Authors

Linda Darling-Hammond

Deborah J. Holtzman

Su Jin Gatlin

Julian Vasquez Heilig

Stanford University

Email: ldh@stanford.edu

Linda Darling-Hammond is Charles E. Ducommun Professor of Education at Stanford University. Among her many publications on issues of teacher quality, school reform, and educational equity are *The Right to Learn: A Blueprint for Creating Schools that Work* and *Teaching as the Learning Profession: A Handbook of Policy and Practice*.

Deborah J. Holtzman is a PhD candidate in the Administration and Policy Analysis program at the Stanford University School of Education. She is also a Research Analyst at the American Institutes for Research. Her current research interests include accountability and alignment among instruction, standards, and assessment.

Su Jin Gatlin is a PhD candidate in Administration and Policy Analysis at Stanford University's School of Education. She is also completing a Master's in Economics at Stanford and holds a BA in Statistics from UC Berkeley. Her current research interests include the role of wealth on postsecondary attendance and school finance reform.

Julian Vasquez Heilig is a Ph.D. candidate in Administration and Policy Analysis at Stanford University. His research focuses on accountability policy and urban education.

APPENDIX A

Descriptive Statistics

Table A-1

Univariate Statistics for Variables in the Regression Models¹

	TLI / TAAS		SAT-9		Aprenda	
	(1996-97—2001-02)		(1999-2000—2001-02)		(1999-2000—2001-02)	
	Mean	S.Dev.	Mean	S. Dev.	Mean	S. Dev.
Test Score - Math	81.31	10.45	55.60	19.13	58.52	16.53
Test Score - Reading	85.46	12.34	48.83	19.73	56.06	16.53
Previous Year test score - Math	78.07	12.00	55.85	19.80	57.02	16.03
Previous Year test score - Reading	82.56	12.59	49.07	19.33	56.69	15.48
Free/ Reduced Price Lunch (student)	0.69	0.46	0.72	0.45	0.98	0.15
American Indian (student)	0.001	0.03	0.001	0.03	0.000	0.000
Asian/Pacific Islander (student)	0.03	0.18	0.036	0.19	0.000	0.02
African American (student)	0.43	0.50	0.42	0.49	0.001	0.03
Hispanic (student)	0.38	0.49	0.41	0.41	0.99	0.05
Limited English Proficient (student)	0.16	0.37	0.24	0.43	0.95	0.21
Total years of experience (teacher)	12.54	10.24	12.14	10.46	7.15	7.43
Masters or Doctoral degree (teacher)	0.29	0.46	0.27	0.45	0.15	0.36
Classroom Number of Students (in grades 3-5)	25.62	4.13	26.18	4.37	24.77	4.38
Class average previous year test score - math	77.60	6.95	55.69	11.87	57.42	7.27
Class average previous year test score - reading	82.03	7.28	48.94	12.19	56.99	7.01
Percent African American students in school	37.69	33.23	36.82	33.07	12.93	16.48
Percent Hispanic students in school	46.61	32.46	48.45	32.55	81.42	18.84
Percent free/reduced price lunch students in school	74.85	43.32	75.49	42.97	95.66	20.35
Certification date unknown (teacher)	0.113	0.32	0.054	0.23	0.093	0.290
Uncertified	0.206	0.40	0.224	0.42	0.261	0.439
Alternatively certified	0.003	0.06	0.005	0.07	0.013	0.111

	TLI / TAAS (1996–97—2001–02)		SAT-9 (1999–2000—2001–02)		Aprenda (1999–2000—2001–02)	
	Mean	S.Dev.	Mean	S. Dev.	Mean	S. Dev.
Emergency / temporary certified	0.052	0.22	0.030	0.17	0.039	0.194
Certified Out-of-Field	0.009	0.09	0.014	0.12	0.006	0.080
Certified no-test	0.010	0.10	0.006	0.08	0.018	0.134
Certification missing	0.051	0.22	0.028	0.17	0.052	0.222
TFA Status	0.021	0.14	0.023	0.15	0.015	0.120
TFA teacher, uncertified	0.010	0.09	0.013	0.11	0.008	0.089
TFA teacher, standard certified	0.011	0.11	0.010	0.10	0.007	0.081
Non-TFA teacher, uncertified	0.197	0.40	0.211	0.41	0.255	0.436
Non-TFA teacher, nonstandard certified	0.116	0.32	0.068	0.25	0.119	0.324

¹This table reports statistics for records in the pooled-year regressions. Except where specific to reading scores, the statistics report the means and standard deviations for data used in the math analyses for each test. Although there were small differences in the sample sizes for reading and math on each test, the statistics are identical to the one-tenths place.

Table A-2
Certification Status of 4th and 5th Grade Teachers, by Year

Certification Status	1996-97*		1997-98*		1998-99		1999-00		2000-01		2001-02	
	Non-TFA	TFA	Non-TFA	TFA	Non-TFA	TFA	Non-TFA	TFA	Non-TFA	TFA	Non-TFA	TFA
Standard (N)	56.5% (1028)	65.2% (15)	61.6% (1089)	68.6% (24)	65.4% (1041)	73.1% (19)	67.8% (1181)	48.3% (14)	68.7% (1216)	45.7% (16)	64.8% (1137)	45.2% (19)
Alternative (N)	0.2% (4)	0.0% (0)	0.3% (5)	0.0% (0)	0.3% (5)	3.8% (1)	0.3% (5)	0.0% (0)	0.6% (11)	8.6% (3)	0.7% (12)	4.8% (2)
Emergency/Temporary (N)	9.5% (172)	0.0% (0)	8.7% (154)	0.0% (0)	5.8% (92)	0.0% (0)	3.9% (68)	0.0% (0)	2.7% (48)	0.0% (0)	2.0% (35)	0.0% (0)
Certified Out-of-field (N)	0.1% (2)	0.0% (0)	0.2% (3)	0.0% (0)	0.3% (4)	0.0% (0)	0.3% (5)	0.0% (0)	1.0% (17)	0.0% (0)	2.2% (38)	0.0% (0)
Certified, No test (N)	2.3% (42)	0.0% (0)	1.2% (22)	0.0% (0)	0.6% (9)	0.0% (0)	0.9% (16)	0.0% (0)	0.8% (15)	0.0% (0)	1.0% (18)	0.0% (0)
Not yet certified (N)	20.4% (370)	34.8% (8)	20.1% (355)	31.4% (11)	23.0% (366)	23.1% (6)	23.2% (404)	51.7% (15)	22.1% (391)	45.7% (16)	25.8% (453)	50.0% (21)
Certification missing (N)	11.0% (200)	0.0% (0)	7.9% (139)	0.0% (0)	4.7% (74)	0.0% (0)	3.6% (62)	0.0% (0)	4.1% (72)	0.0% (0)	3.5% (61)	0.0% (0)

* In these years, the HISD data set included a number of candidates with certification codes but not certification dates (16% of the total in 1996-97 and 13% in 1997-98). They are listed in this table in the category indicated by their certification code for each year that they appear in the data set. We are cautious in drawing inferences for these two years about when the certification status went into effect.

Note: Includes all teachers who taught any 4th or 5th grade student

Table A-3
Student Sample Sizes, By Test and Year

Year	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
1996-97	16,635	16,265	—	—	—	—
1997-98	16,738	16,342	—	—	—	—
1998-99	16,954	16,616	—	—	—	—
1999-00	18,067	17,550	19,880	19,934	3,854	3,852
2000-01	18,205	17,723	20,006	20,039	3,820	3,819
2001-02	18,912	18,626	20,602	20,634	3,763	3,765

APPENDIX B

Ordinary Least Squares Regression Estimates

Table B-1

Regression Coefficients, TFA Status and Student Achievement, Houston ISD, Grades 4–5 (pooled years)

Variable	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
Constant	30.35*** (77.87)	31.67*** (66.24)	15.46*** (27.67)	9.96*** (18.57)	24.46*** (6.37)	7.33~ (1.92)
Previous year test score	0.498*** (201.90)	0.526*** (184.37)	0.683*** (223.70)	0.742*** (240.41)	0.670*** (81.37)	0.706*** (83.31)
Free/reduced price lunch status	-0.922*** (-13.77)	-1.41*** (-17.19)	-1.70*** (-12.36)	-1.68*** (-12.76)	0.677 (0.856)	0.991 (1.26)
American Indian	0.161 (0.189)	0.135 (0.129)	1.71 (0.963)	-0.278 (-0.170)	—	—
Asian/Pacific Islander	0.678*** (4.50)	0.541** (2.94)	2.32*** (7.81)	0.616* (2.17)	9.55 (1.34)	9.84 (1.39)
African American	-1.66*** (-17.19)	-2.23*** (-18.94)	-3.03*** (-15.41)	-2.49*** (-13.17)	-3.17 (-0.592)	11.06~ (1.82)
Hispanic	-0.665*** (-6.98)	-1.32*** (-11.32)	-1.22*** (-6.25)	-1.86*** (9.95)	-0.866 (-0.263)	4.45 (1.36)
Limited English Proficient	1.58*** (20.70)	0.664*** (7.09)	0.884*** (6.34)	0.872*** (6.51)	2.08*** (3.71)	1.36* (2.44)
Total teaching experience	-0.012*** (-4.45)	0.002 (0.508)	0.035*** (6.90)	0.054*** (11.24)	0.107*** (6.22)	0.024 (1.39)
Masters degree or higher	-0.224*** (-3.87)	-0.097 (-1.37)	-0.285* (-2.47)	-0.356** (-3.22)	-0.524 (-1.46)	0.236 (0.663)
Classroom number of students	0.148*** (24.94)	0.019** (2.67)	-0.029** (-2.68)	0.020~ (1.87)	0.144*** (5.22)	0.197*** (7.19)
Class average previous year test score	0.117*** (25.86)	0.150*** (28.37)	0.100*** (17.17)	0.123*** (21.20)	-0.049** (-2.66)	-0.061** (-3.22)

Does Teacher Preparation Matter?

Variable	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
% African American students in school	0.010*** (4.41)	-0.003 (-1.15)	0.003 (0.710)	-0.022*** (-4.98)	-0.137*** (-6.36)	0.015 (0.492)
% Hispanic students in schools	0.013*** (5.42)	-0.002 (-0.627)	-0.010* (-2.02)	-0.030*** (-6.54)	-0.125*** (-6.21)	-0.003 (-0.154)
% free/reduced price lunch students in school	-0.003** (-2.74)	0.001 (0.799)	-0.002 (-0.698)	0.005** (2.63)	0.048*** (6.43)	0.005 (0.699)
TFA teacher status	0.689*** (3.98)	-0.039 (-0.184)	-0.882** (-2.74)	-0.642* (-2.08)	-2.87** (-2.88)	-2.58** (-2.61)
R ²	.43	.39	.62	.68	.42	.43
N	105,511	103,122	60,488	60,607	11,437	11,436

T-values are in parentheses.

~ p < .10; *p < .05; **p < .01; ***p < .001

Table B-2

Regression Coefficients, Certification Status and Student Achievement, Houston ISD, Grades 4–5 (pooled years)

Variable	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Apranda Math	Apranda Reading
Constant	30.98*** (79.19)	32.11*** (66.60)	15.59*** (27.94)	10.05*** (18.70)	24.02*** (6.26)	7.82* (2.05)
Previous year test score	0.498*** (202.28)	0.527*** (184.45)	0.683*** (223.85)	0.742*** (240.49)	0.670*** (81.61)	0.706*** (83.54)
Free/reduced price lunch status	-0.912*** (-13.63)	-1.41*** (-17.16)	-1.69*** (-12.29)	-1.68*** (-12.73)	0.656 (0.831)	0.888 (1.14)
American Indian	0.210 (0.247)	0.130 (0.125)	1.87 (1.05)	-0.134 (-0.079)	—	—
Asian/Pacific Islander	0.703*** (4.67)	0.553** (3.01)	2.35*** (7.92)	0.638* (2.24)	9.60 (1.35)	10.36 (1.47)
African American	-1.67*** (-17.31)	-2.23*** (-18.98)	-3.03*** (-15.38)	-2.47*** (-13.11)	-3.27 (-0.613)	10.87~ (1.80)
Hispanic	-0.657*** (-6.91)	-1.32*** (-11.32)	-1.22*** (-6.27)	-1.86*** (9.96)	-0.358 (-0.109)	4.81 (1.48)
Limited English Proficient	1.50*** (19.52)	0.631*** (6.72)	0.902*** (6.47)	0.873*** (6.52)	1.84** (3.26)	1.05~ (1.88)
Total teaching experience	-0.025*** (-9.23)	-0.008* (-2.23)	0.031*** (5.76)	0.050*** (9.72)	0.082*** (4.03)	0.030 (1.61)
Masters degree or higher	-0.169** (-2.88)	-0.077 (-1.08)	-0.252* (-2.18)	-0.329** (-2.97)	-0.393 (-1.05)	0.369 (0.991)
Classroom number of students	0.147*** (24.86)	0.020** (2.76)	-0.028** (-2.52)	0.022* (2.06)	0.145*** (5.24)	0.194*** (7.06)
Class average previous year test score	0.113*** (25.10)	0.147*** (27.83)	0.100*** (17.18)	0.122*** (21.12)	-0.041* (-2.21)	-0.057** (-3.00)
% African American students in school	0.011*** (4.63)	-0.003 (-0.871)	0.003 (0.747)	-0.022*** (-4.95)	-0.132*** (-6.04)	0.003 (0.154)
% Hispanic students in schools	0.014*** (5.65)	-0.001 (-0.424)	-0.010* (-2.12)	-0.030*** (-6.51)	-0.126*** (-6.22)	-0.012 (-0.614)
% free/reduced price lunch students in school	-0.003* (-2.43)	0.001 (0.725)	-0.002 (-0.768)	0.005* (2.55)	0.050*** (6.57)	0.006 (0.758)

Does Teacher Preparation Matter?

Variable	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
Uncertified	-0.525*** (-8.03)	-0.580*** (-7.23)	-0.414** (-3.30)	-0.516*** (-4.31)	-1.41*** (-4.61)	-0.066 (-0.216)
Alternatively certified	-0.897* (-2.02)	-0.818 (-1.49)	-2.31** (-3.29)	-1.40* (-2.09)	-0.491 (-0.454)	3.13** (2.91)
Emergency/temporary certified	-0.701*** (-3.85)	-0.690** (-3.07)	-0.636~ (-1.68)	0.780* (2.16)	-1.40 (-1.15)	-0.754 (-0.619)
Certified out-of-field	0.667* (2.52)	0.902** (2.77)	-1.80*** (-4.36)	0.216 (0.548)	-3.96** (-2.64)	3.64* (2.45)
Certified, no test	-0.187 (-0.655)	-0.655~ (-1.87)	-4.49*** (-7.06)	-2.34*** (-3.83)	-3.52*** (-3.66)	-4.76*** (-5.01)
Certification code missing	0.113 (0.549)	-0.642* (-2.53)	-2.26*** (-4.63)	0.121 (0.260)	-4.78*** (-3.84)	-4.37*** (-3.52)
Certification date unknown	-1.09*** (-6.17)	0.043 (0.199)	1.65*** (4.07)	0.152 (0.393)	4.43*** (3.82)	4.66*** (4.03)
R ²	.43	.39	.62	.68	.42	.43
N	105,511	103,122	60,488	60,607	11,437	11,436

T-values are in parentheses.

~ p <.10; *p<.05; **p<.01; ***p<.001

Table B-3

Regression Coefficients, TFA Status, Certification Status, and Student Achievement, Houston ISD, Grades 4–5 (pooled years)

Variable	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Apranda Math	Apranda Reading
Constant	30.91*** (77.11)	32.12*** (66.74)	15.58*** (27.93)	10.09*** (18.78)	24.79*** (6.46)	7.39~ (1.94)
Previous year test score	0.498*** (201.26)	0.527*** (184.46)	0.684*** (223.77)	0.743*** (240.46)	0.670*** (81.52)	0.706*** (83.42)
Free/reduced price lunch status	-0.917*** (-13.71)	-1.41*** (-17.16)	-1.69*** (-12.32)	-1.68*** (-12.70)	0.703 (0.891)	0.981 (1.25)
American Indian	0.186 (0.218)	0.115 (0.110)	1.64 (0.922)	-0.332 (-0.197)	—	—
Asian/Pacific Islander	0.702*** (4.66)	0.553** (3.01)	2.33*** (7.83)	0.630* (2.21)	9.55 (1.35)	10.21 (1.45)
African American	-1.67*** (-17.36)	-2.23*** (-18.99)	-3.03*** (-15.37)	-2.48*** (-13.14)	-3.15 (-0.590)	11.12~ (1.83)
Hispanic	-0.657*** (-6.91)	-1.32*** (-11.32)	-1.21*** (-6.24)	-1.86*** (9.97)	-0.477 (-0.145)	4.71 (1.44)
Limited English Proficient	1.51*** (19.67)	0.639*** (6.81)	0.883*** (6.33)	0.865*** (6.46)	1.89** (3.37)	1.33* (2.39)
Total teaching experience	-0.022*** (-7.88)	-0.007* (-2.07)	0.029*** (5.34)	0.048*** (9.42)	0.077*** (4.03)	0.027 (1.43)
Masters degree or higher	-0.169** (-2.93)	-0.070 (-0.992)	-0.296* (-2.56)	-0.358** (-3.24)	-0.418 (-1.15)	0.303 (0.838)
Classroom number of students	0.147*** (24.81)	0.020** (2.71)	-0.029** (-2.61)	0.021* (1.99)	0.138*** (5.00)	0.194*** (7.07)
Class average previous year test score	0.114*** (25.33)	0.147*** (27.85)	0.099*** (17.05)	0.122*** (21.04)	-0.048** (-2.61)	-0.058** (-3.02)
% African American students in school	0.010*** (4.39)	-0.003 (-0.921)	0.004 (0.913)	-0.022*** (-4.84)	-0.136*** (-6.27)	0.007 (0.319)
% Hispanic students in schools	0.013*** (5.22)	-0.001 (-0.485)	-0.008~ (-1.78)	-0.029*** (-6.32)	-0.125*** (-6.22)	-0.010 (-0.499)
% free/reduced price lunch students in school	-0.002* (-2.22)	0.001 (0.767)	-0.002 (-0.889)	0.005* (2.46)	0.049*** (6.55)	0.006 (0.809)

Does Teacher Preparation Matter?

Variable	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
TFA teacher, uncertified	-0.288 (-1.14)	-1.22*** (-3.93)	-1.89*** (-4.41)	-1.09** (-2.67)	-4.06** (-3.01)	-5.09*** (-3.81)
TFA teacher, standard certified	1.18*** (4.93)	0.437 (1.49)	-0.008 (0.017)	-0.458 (-0.994)	-2.55~ (-1.73)	0.839 (0.572)
Non-TFA teacher, uncertified	-0.488*** (-7.36)	-0.550*** (-6.76)	-0.341** (-2.68)	-0.519*** (-4.25)	-1.33*** (-4.30)	-0.014 (-0.045)
Non-TFA teacher, nonstandard certified	-0.284 (-1.63)	-0.605** (-2.81)	-1.54*** (-4.60)	-0.051 (-0.159)	-2.45*** (-3.53)	-1.26~ (-1.83)
R ²	.43	.39	.62	.68	.42	.43
N	105,511	103,122	60,488	60,607	11,437	11,436

T-values are in parentheses.

~ p < .10; *p < .05; **p < .01; ***p < .001

APPENDIX C

Analysis with Hierarchical Linear Modeling

This appendix describes a preliminary analysis of the data that was done using Hierarchical Linear Modeling (HLM). This analysis paralleled the OLS regression analysis for Teacher Certification and TFA Status (see Table 7 of the main paper) in every way except that the HLM analysis took into account the nesting of students within classrooms and classrooms within schools. The variables used were identical, as were the records included.

Rationale for the HLM Analysis

We originally used OLS analysis in order to recreate the analyses conducted in the CREDO study using Houston's data. However, as undertaken here, the HLM analysis had two primary advantages over the OLS analysis. First, the OLS analysis treated all students as independent of one another when in fact students were groups within classrooms as well as within schools. Similarly, the OLS analysis did not take into account the grouping of teachers within schools. As such, the standard errors from the OLS results could be estimated to be smaller than appropriate, possibly inflating significance results. The HLM analysis attempted to remedy this problem by taking into account the nesting of students (level 1) within classrooms/teachers (level 2) within schools (level 3).

Second, the HLM analysis allowed for the modeling of *within-school* relationships between the teacher variables and student achievement outcomes (that is, the relationships net of any school membership effects). This was accomplished by centering the teacher/classroom (level 2) variables around their school (level 3) means. (See Raudenbush & Bryk, 2002, pp. 135-139).

Despite its advantages, the HLM analysis reported here has some important limitations. In particular, this analysis (like the OLS analysis) did not take into account the longitudinal nature of the data. Each student was represented by a distinct and separate record for each year he or she was in the data base; this was also true of teachers and schools. Just as with the OLS analysis, a single student may have been included as a fourth grader and, again, as a fifth grader; such a student was, in effect, treated as two independent students. The same was true for teachers, many of whom were in the data base for multiple years but were treated independently for each classroom-year, and schools. Ideally, the data would be modeled as cross-classified random effects of students crossing teachers nested within schools over time; however, this is a complex specification that could not be undertaken for this paper. This approach may be pursued in future work.

Method

The HLM analysis used the exact same data and variables as were used for the OLS analysis. However, the HLM analysis employed a three-level technique where level 1 was students, level 2 was teachers/classrooms, and level 3 was schools. The level-1 predictors were the student control variables: previous-year test score, free/reduced price lunch status, race/ethnicity and LEP status. The level-2 predictors included the four variables of primary interest—(1) TFA teacher, uncertified or nonstandard certified; (2) TFA teacher, standard certified; (3) non-TFA teacher, uncertified; and (4) non-TFA teacher, nonstandard certified—with non-TFA standard-certified teachers serving as the reference group. Also included at level 2 were the teacher/classroom-level controls: years of

teaching experience, degree level (masters/doctoral vs. bachelors or below), an indicator for missing certification date, classroom number of students, and class average previous year test score. The level-3 predictors were the schoolwide percentages of free/reduced price lunch students, African American students, and Hispanic students.

Models were built up from level 1, with predictors added in groups. For each different outcome variable (the six test measures), a series of six models were constructed. The first (Model A) was fully unconditional, with no predictors, thereby allowing computation of the variance explained at each of the three levels. The second model (Model B) added the student-level previous-year achievement variable, grand-mean centered, at level 1. Model C added all of the other student-level predictors, grand-mean centered. Model D added at level 2 all of the teacher/classroom level predictors, group-mean centered, except the classroom previous achievement average. Model E added in the classroom previous achievement average, group-mean centered (and at the same time changed the centering on the level-1 previous achievement variable to group-mean centering). The final model (Model F) added at level 3 the school-level predictors, grand-mean centered.

The models included random effects only on the intercepts (which, due to the centering, represent adjusted class/school achievement means); all slope effects were set as fixed. Accordingly, the level-2 and level-3 variables were used as predictors only in the intercept equations.

The equations for Model F are as follows (i indexes students, j indexes teachers/classrooms, and k indexes schools):

Level 1:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{PREV ACH}_{ijk} - \text{PREV ACH}_{\cdot jk}) + \pi_{2jk}(\text{FRPL}_{ijk} - \text{FRPL}_{\cdot jk}) + \\ + \pi_{3jk}(\text{AMIND}_{ijk} - \text{AMIND}_{\cdot jk}) + \pi_{4jk}(\text{ASIAN}_{ijk} - \text{ASIAN}_{\cdot jk}) + \\ \pi_{5jk}(\text{BLACK}_{ijk} - \text{BLACK}_{\cdot jk}) + \pi_{6jk}(\text{HISPANIC}_{ijk} - \text{HISPANIC}_{\cdot jk}) + \\ \pi_{7jk}(\text{LEP}_{ijk} - \text{LEP}_{\cdot jk}) + e_{ijk}$$

Level 2:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{TFANONSTD}_{jk} - \text{TFANONSTD}_{\cdot k}) + \beta_{02k}(\text{TFASTD}_{jk} - \text{TFASTD}_{\cdot k}) + \beta_{03k}(\text{NONTFAUNCERT}_{jk} - \text{NONTFAUNCERT}_{\cdot k}) + \\ \beta_{04k}(\text{NONTFAOHCERT}_{jk} - \text{NONTFAOHCERT}_{\cdot k}) + \beta_{05k}(\text{EXPER}_{jk} - \text{EXPER}_{\cdot k}) + \beta_{06k}(\text{MADOC}_{jk} - \text{MADOC}_{\cdot k}) + \beta_{07k}(\text{NOCERTDATE}_{jk} - \text{NOCERTDATE}_{\cdot k}) + \beta_{08k}(\text{CLSNUMSTU}_{jk} - \text{CLSNUMSTU}_{\cdot k}) + \\ \beta_{09k}(\text{CLSAVGPREVACH}_{jk} - \text{CLSAVGPREVACH}_{\cdot k}) + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k}$$

$$\pi_{2jk} = \beta_{20k}$$

...

$$\pi_{7jk} = \beta_{70k}$$

Level 3:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(\text{BLACKPCT}'_k - \text{BLACKPCT}') + \gamma_{002}(\text{HISPPCT}'_k - \text{HISPPCT}') + \gamma_{003}(\text{FRPLPCT}'_k - \text{FRPLPCT}') + u_{00k}$$

$$\beta_{01k} = \gamma_{010}$$

$$\beta_{02k} = \gamma_{020}$$

...

$$\beta_{09k} = \gamma_{090}$$

$$\beta_{10k} = \gamma_{100}$$

$$\begin{array}{l} \beta_{20k} = \gamma_{200} \\ \dots \\ \beta_{70k} = \gamma_{700} \end{array}$$

Results

Table C-1 shows the Model F results (fixed effects and variance components) from all six outcome measures. Tables C-2 and C-3 show the results from Models A-F for TAAS/TLI math and TAAS/TLI reading, mainly as an illustration of the full model-building process; results from the full process for the other four outcome measures are available upon request.

The TFA/Certification status results are generally similar to those obtained by the OLS analyses, although, as to be expected, significance levels tend to be lower. Students taught by uncertified TFA teachers made significantly less progress on three of the six tests (TAAS/TLI reading, SAT-9 reading, and Aprenda math) than the students of non-TFA standard-certified teachers; TFA coefficients on the other three tests were also negative but not significant. Thus, the HLM results support the OLS finding that in no instance do the students of uncertified TFA teachers perform better than the students of standard-certified teachers.

Non-TFA uncertified teachers had significant negative effects, relative to non-TFA standard-certified teachers, on four of the six tests, and non-significant effects on the other two. Students of certified TFA teachers performed comparably to students of other certified teachers on all six tests (no statistically significant differences). Again, these results are similar to the OLS results.

Table C-1
*HLM Parameter Estimates, Teacher Certification, TFA Status, and Student Achievement Gains,
 Houston ISD, Grades 4–5 (pooled years)*

Variables	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
Fixed Effects, Adjusted Achievement Means						
Intercept	80.35*** (608.77)	84.13*** (590.63)	53.40*** (217.33)	45.97*** (194.79)	57.94*** (134.64)	55.69*** (151.96)
Student-level variables ^a						
Previous year test score	0.500*** (94.55)	0.528*** (94.53)	0.686*** (161.87)	0.744*** (166.59)	0.666*** (73.73)	0.701*** (68.25)
Free/reduced price lunch status	-0.709*** (-11.49)	-1.14*** (-15.17)	-1.05*** (-8.39)	-1.26*** (-9.68)	0.347 (0.493)	0.583 (0.845)
American Indian	0.110 (0.239)	-0.37 (-0.497)	1.69 (1.28)	-0.550 (-0.354)	N/A	N/A
Asian/Pacific Islander	0.810*** (8.35)	0.379** (2.84)	2.26*** (8.06)	0.636* (2.062)	9.958 (1.132)	10.93** (2.68)
African American	-1.62*** (-16.44)	-2.14*** (-19.33)	-2.80*** (-14.05)	-2.42*** (-12.74)	-4.760 (-0.897)	8.57 (1.34)
Hispanic	-0.536*** (-6.27)	-1.23*** (-11.37)	-1.23*** (-6.31)	-1.75*** (-9.27)	-0.976 (-0.192)	6.26** (2.73)
Limited English Proficient	0.804*** (11.00)	0.620*** (6.18)	1.07*** (7.07)	0.883*** (6.19)	1.42* (2.08)	1.73** (2.70)
Classroom-level variables ^b						
TFA teacher, uncertified or nonstandard certified	-0.445 (-0.768)	-1.34* (-2.20)	-1.52 (-1.57)	-1.54* (-2.41)	-6.75** (-2.69)	-2.60 (-0.927)
TFA teacher, standard certified	0.901 (1.54)	0.974 (1.31)	0.093 (0.074)	0.497 (0.493)	-3.03 (-0.646)	-2.88 (-1.07)
Non-TFA teacher, uncertified	-0.574*** (-3.81)	-0.583** (-3.58)	-0.552~ (-1.94)	-0.540* (-2.06)	-0.566 (-0.613)	0.685 (0.813)
Non-TFA teacher, nonstandard certified	-0.387 (-0.797)	-0.587 (-1.11)	-1.82* (-2.00)	-0.567 (-0.758)	0.968 (0.555)	-2.04 (-0.796)
Total teaching experience	-0.021** (-3.01)	-0.000 (-0.040)	0.024~ (1.87)	0.044*** (3.95)	0.022 (0.443)	0.003 (0.060)
Masters degree or higher	-0.022 (-0.177)	0.065 (0.461)	0.014 (0.060)	-0.022 (-0.102)	-0.483 (-0.482)	1.50~ (1.66)
Certification Date unknown	-0.279 (-0.555)	0.165 (0.297)	1.85~ (1.79)	0.967 (1.13)	0.429 (0.247)	4.74~ (1.82)
Classroom number of students	0.181*** (9.33)	0.013 (0.643)	0.013 (0.373)	0.067* (2.19)	-0.157 (-1.46)	0.293** (2.99)

Variables	TLI Math	TLI Reading	SAT-9 Math	SAT-9 Reading	Aprenda Math	Aprenda Reading
Class average previous year test score	0.584*** (44.71)	0.627*** (47.61)	0.787*** (62.52)	0.869*** (78.23)	0.556*** (9.84)	0.501*** (9.56)
School-level variables ^c						
% African American students in school	-0.065*** (-7.05)	-0.082*** (-8.61)	-0.213*** (217.33)	-0.262*** (194.79)	-0.111 (-1.46)	0.014 (0.249)
% Hispanic students in school	-0.056*** (-5.70)	-0.089*** (-8.77)	-0.214*** (-9.30)	-0.282*** (-12.917)	-0.140* (-2.01)	-0.042 (-0.812)
% free/reduced price lunch students in school	-0.004 (-0.869)	-0.006 (-1.15)	-0.025* (-2.11)	-0.034** (-3.03)	0.029 (1.06)	0.008 (0.448)
Variance Components						
Level 1 (student)	43.44	72.13	108.40	104.88	110.56	127.96
Level 2 (classroom)	13.37	14.00	22.35	16.57	30.82	34.64
Level 3 (school)	15.08	17.37	27.41	25.94	34.74	16.77
Level 1 N	105511	103122	60488	60607	11437	11436
Level 2 N	6530	6519	3287	3287	763	762
Level 3 N	1105	1105	559	559	310	310

Coefficients (and Robust T) and Variance Components from Final Models (Modeled adjusted mean achievement only; all slope coefficients set as fixed)

~p<.10, * p<.05, ** p<.01, ***p<.001

^a All student-level variables grand-mean centered except the previous year test score, which was group-mean centered.

^b All classroom/teacher-level variables group-mean centered.

^c All school-level variables grand-mean centered.

Table C-2
HLM Parameter Estimates, Teacher Certification, TFA Status, and Student Achievement Gains in TLI Math only, Houston ISD, Grades 4–5, by model (pooled years)

Variables	Model A	Model B	Model C	Model D	Model E	Model F
Fixed Effects, Adjusted Achievement Means						
Intercept	80.52*** (571.35)	80.99*** (764.46)	81.02*** (790.99)	81.00*** (789.19)	80.38*** (584.21)	80.35*** (608.77)
Student-level variables ^a						
Previous year test score		0.515*** (97.20)	0.505*** (96.70)	0.505*** (96.49)	0.499*** (94.61)	0.500*** (94.55)
Free/reduced price lunch status			-0.727*** (-11.95)	-0.731*** (-12.03)	-0.749*** (-12.20)	-0.709*** (-11.49)
American Indian			0.076 (0.164)	0.079 (0.171)	0.070 (0.152)	0.110 (0.239)
Asian/Pacific Islander			0.784*** (8.18)	0.790*** (8.24)	0.800*** (8.22)	0.810*** (8.35)
African American			-1.64*** (-17.09)	-1.65*** (-17.18)	-1.70*** (-17.49)	-1.62*** (-16.44)
Hispanic			-0.544*** (-6.50)	-0.550*** (-6.58)	-0.579*** (-6.84)	-0.536*** (-6.27)
Limited English Proficient			0.822*** (11.32)	0.813*** (11.22)	0.798*** (10.90)	0.804*** (11.00)
Classroom-level variables ^b						
TFA teacher, uncertified or nonstandard certified				-0.470 (-0.876)	-0.447 (-0.772)	-0.445 (-0.768)
TFA teacher, standard certified				0.992~ (1.69)	0.897 (1.53)	0.901 (1.54)
Non-TFA teacher, uncertified				-0.669*** (-4.54)	-0.574*** (-3.81)	-0.574*** (-3.81)
Non-TFA teacher, nonstandard certified				-0.457 (-0.938)	-0.388 (-0.798)	-0.387 (-0.797)
Total teaching experience				-0.015* (-2.15)	-0.021** (-3.02)	-0.021** (-3.01)
Masters degree or higher				-0.106 (-0.847)	-0.023 (-0.184)	-0.022 (-0.177)
Certification Date unknown				-0.265 (-0.525)	-0.276 (-0.549)	-0.279 (-0.555)
Classroom number of students				0.187*** (9.59)	0.182*** (9.34)	0.181*** (9.33)

Variables	Model A	Model B	Model C	Model D	Model E	Model F
Class average previous year test score					0.583*** (44.67)	0.584*** (44.71)
School-level variables ^c						
% African American students in school						-0.065*** (-7.05)
% Hispanic students in school						-0.056*** (-5.70)
% free/reduced price lunch students in school						-0.004 (-0.869)
Variance Components						
Level 1 (student)	71.05	43.86	43.43	43.43	43.44	43.44
Level 2 (classroom)	24.44	13.98	13.79	13.26	13.34	13.37
Level 3 (school)	15.91	9.48	8.64	8.76	16.95	15.08

Coefficients (and Robust T) and Variance Components from Final Models (Modeled adjusted mean achievement only; all slope coefficients set as fixed)

~p<.10, * p<.05, ** p<.01, ***p<.001

Reference Group for TFA/certification variables is Non-TFA teacher, standard certified

Reference Group for student ethnicity variables is White students

^a All student-level variables grand-mean centered, except, in Models E and F only, the previous year test score, which was group-mean centered in these two models.

^b All classroom/teacher-level variables group-mean centered.

^c All school-level variables grand-mean centered.

Table C-3
HLM Parameter Estimates, Teacher Certification, TFA Status, and Student Achievement Gains in TLI Reading only, Houston ISD, Grades 4–5, by model (pooled years)

Variables	Model A	Model B	Model C	Model D	Model E	Model F
Fixed Effects, Adjusted Achievement Means						
Intercept	84.41*** (539.01)	85.08*** (771.44)	85.18*** (817.49)	85.17*** (816.57)	84.19*** (552.47)	84.13*** (590.63)
Student-level variables ^a						
Previous year test score		.547*** (97.62)	.537*** (97.47)	.537*** (97.31)	.527*** (94.63)	.528*** (94.53)
Free/reduced price lunch status			-1.20*** (-16.12)	-1.20*** (-16.12)	-1.22*** (-16.20)	-1.14*** (-15.17)
American Indian			-.412 (-.549)	-.419 (-.559)	-.443 (-.590)	-.37 (-.497)
Asian/Pacific Islander			.369** (2.77)	.369** (2.78)	.357** (2.66)	.379** (2.84)
African American			-2.21*** (-20.84)	-2.21*** (-20.84)	-2.24*** (-20.62)	-2.14*** (-19.33)
Hispanic			-1.29*** (-12.30)	-1.29*** (-12.29)	-1.34*** (-12.55)	-1.23*** (-11.37)
Limited English Proficient			.589*** (5.89)	.583*** (5.84)	.612*** (6.10)	.620*** (6.18)
Classroom-level variables ^b						
TFA teacher, uncertified or nonstandard certified				-1.52** (-2.70)	-1.35* (-2.21)	-1.34* (-2.20)
TFA teacher, standard certified				.984 (1.31)	.959 (1.29)	.974 (1.31)
Non-TFA teacher, uncertified				-.722*** (-4.44)	-.584** (-3.59)	-.583** (-3.58)
Non-TFA teacher, nonstandard certified				-.751 (-1.40)	-.595 (-1.12)	-.587 (-1.11)
Total teaching experience				.005 (.667)	-.000 (-.054)	-.000 (-.040)
Masters degree or higher				-.004 (-.031)	.063 (.447)	.065 (.461)
Certification Date unknown				.212 (.377)	.177 (.319)	.165 (.297)
Classroom number of students				.030 (1.41)	.014 (.658)	.013 (.643)
Class average					.626***	.627***

Variables	Model A	Model B	Model C	Model D	Model E	Model F
previous year test score					(47.47)	(47.61)
School-level variables ^c						
% African American students in school						-.082*** (-8.61)
% Hispanic students in school						-.089*** (-8.77)
% free/reduced price lunch students in school						-.006 (-1.15)
Variance Components						
Level 1 (student)	106.32	72.71	72.12	72.12	72.13	72.13
Level 2 (classroom)	27.91	14.65	14.30	14.14	13.94	14.00
Level 3 (school)	19.77	10.01	8.81	8.84	20.88	17.37

Coefficients (and Robust T) and Variance Components from Final Models (Modeled adjusted mean achievement only; all slope coefficients set as fixed)

~p<.10, * p<.05, ** p<.01, ***p<.001

Reference Group for TFA/certification variables is Non-TFA teacher, standard certified

Reference Group for student ethnicity variables is White students

^a All student-level variables grand-mean centered, except, in Models E and F only, the previous year test score, which was group-mean centered in these two models.

^b All classroom/teacher-level variables group-mean centered.

^c All school-level variables grand-mean centered.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermamdorn.com.

Editorial Board

Michael W. Apple

University of Wisconsin

Greg Camilli

Rutgers University

Linda Darling-Hammond

Stanford University

Gustavo E. Fischman

Arizona State University

Gene V Glass

Arizona State University

Aimee Howley

Ohio University

William Hunter

University of Ontario Institute of Technology

Daniel Kallós

Umeå University

Thomas Mauhs-Pugh

Green Mountain College

Heinrich Mintrop

University of California, Berkeley

Anthony G. Rud Jr.

Purdue University

Terrence G. Wiley

Arizona State University

David C. Berliner

Arizona State University

Casey Cobb

University of Connecticut

Mark E. Fetler

California Commission on Teacher Credentialing

Richard Garlikov

Birmingham, Alabama

Thomas F. Green

Syracuse University

Craig B. Howley

Appalachia Educational Laboratory

Patricia Fey Jarvis

Seattle, Washington

Benjamin Levin

University of Manitoba

Les McLean

University of Toronto

Michele Moses

Arizona State University

Michael Scriven

Western Michigan University

John Willinsky

University of British Columbia

EDUCATION POLICY ANALYSIS ARCHIVES
English-language Graduate-Student Editorial Board

Noga Admon
New York University

Jessica Allen
University of Colorado

Cheryl Aman
University of British Columbia

Anne Black
University of Connecticut

Marisa Burian-Fitzgerald
Michigan State University

Chad d'Entremont
Teachers College Columbia University

Carol Da Silva
Harvard University

Tara Donahue
Michigan State University

Camille Farrington
University of Illinois Chicago

Chris Frey
Indiana University

Amy Garrett Dikkers
University of Minnesota

Misty Ginicola
Yale University

Jake Gross
Indiana University

Hee Kyung Hong
Loyola University Chicago

Jennifer Lloyd
University of British Columbia

Heather Lord
Yale University

Shereza Mohammed
Florida Atlantic University

Ben Superfine
University of Michigan

John Weathers
University of Pennsylvania

Kyo Yamashiro
University of California Los Angeles

Archivos Analíticos de Políticas Educativas

Associate Editors

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

Editorial Board

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco

Dalila Andrade de Oliveira

Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil

Alejandro Canales

Universidad Nacional Autónoma
de México

Erwin Epstein

Loyola University, Chicago,
Illinois

Rollin Kent

Universidad Autónoma de
Puebla. Puebla, México

Daniel C. Levy

University at Albany, SUNY,
Albany, New York

María Loreto Egaña

Programa Interdisciplinario de
Investigación en Educación

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez Gómez

Universidad de Málaga

Diana Rhoten

Social Science Research Council,
New York, New York

Susan Street

Centro de Investigaciones y
Estudios Superiores en
Antropología Social Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona Lisboa,

Adrián Acosta

Universidad de Guadalajara
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Ursula Casanova

Arizona State University,
Tempe, Arizona

Mariano Fernández

Enguita Universidad de
Salamanca. España

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Belo Horizonte

Mariano Narodowski

Universidad Torcuato Di
Tella, Argentina

Vanilda Paiva

Universidade Estadual Do
Rio De Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, Los Angeles,
California

Carlos A. Torres

UCLA

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación, Chile

Teresa Bracho

Centro de Investigación y
Docencia Económica-CIDE

Sigfredo Chiroque

Instituto de Pedagogía Popular,
Perú

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
Sacramento, California

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

Romualdo Portella do

Oliveira

Universidade de São Paulo

Daniel Schugurensky

Ontario Institute for Studies in
Education, Canada

Daniel Suarez

Laboratorio de Políticas
Públicas-Universidad de
Buenos Aires, Argentina

Jurjo Torres Santomé

Universidad de la Coruña,
España