# Supplying Disadvantaged Schools with Effective Teachers: Experimental Evidence on Secondary Math Teachers from Teach For America

*By* HANLEY S. CHIANG, MELISSA A. CLARK, AND SHEENA MCCONNELL*

MATHEMATICA POLICY RESEARCH

OCTOBER 2013

DRAFT: PLEASE DO NOT CITE OR CIRCULATE WITHOUT AUTHORS' PERMISSION

***Abstract***

*Teach For America (TFA) is an important but controversial source of teachers for hard-to-staff subjects in high-poverty U.S. schools. We present findings from the first large-scale experimental study of secondary math teachers from TFA. We find that TFA teachers are more effective than other math teachers in the same schools, increasing student math achievement by an average of 0.07 standard deviations over the course of one school year. Addressing concerns about the fact that TFA requires only a two-year commitment, we find that novice TFA teachers are more effective than more experienced non-TFA teachers in the same schools.*

* Chiang: Mathematica Policy Research, 955 Massachusetts Ave., Suite 801, Cambridge, MA 02139 (e-mail: hchiang@mathematica-mpr.com); Clark: Mathematica Policy Research, PO Box 2393, Princeton, NJ 08543 (e-mail: mclark@mathematica-mpr.com); McConnell: Mathematica Policy Research, 1100 1st Street, NE, 12th Floor, Washington, DC 20002 (e-mail: smcconnell@mathematica-mpr.com). This paper is based on a larger study sponsored by the Institute of Education Sciences (IES) to examine the effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs (Clark et al. 2013). We are grateful to IES project officers Elizabeth Warner and Stefanie Schmidt for their support throughout this study. Tim Silva, Kathy Sonnenfeld, Eric Zeidman, Nancy Duda, Mary Grider, Alison Wellington, Daniel Player, Phil Gleason, and Michael Puma made important contributions to the study. We also thank the study's technical working group, Julie Greenberg, Paul Holland, Tim Sass, Jeff Smith, Suzanne Wilson, and Jim Wyckoff, for valuable input on the study design and analysis. Finally, we thank the many school staff who allowed us to implement the study in their schools. The authors are responsible for all errors.

The U.S. education system faces growing concerns about widening disparities in academic achievement and subsequent life outcomes between poor and non-poor students (Duncan and Murnane 2011). In policy debates over how to improve the outcomes of disadvantaged students, ensuring a supply of effective teachers to high-poverty schools has been a central focus of attention. A key impetus has come from the accumulating body of empirical evidence demonstrating that teacher effectiveness is critical to students' academic and life outcomes.[1] Despite the importance of teacher quality to student success, school districts across the United States struggle with obtaining high-quality teachers for schools serving low-income students (Monk 2007; Jacob 2007). These challenges are more serious in particular academic subjects, especially math and science at the secondary level (Ingersoll and Perda 2009; Ingersoll and May 2012).

Views differ widely on how to increase the supply of effective teachers to high-poverty schools. One prominent view is that increasing the amount of formal education and preparation a teacher receives *before* entering the classroom will help ensure effective teaching (Darling-Hammond 2000). Critics of this view contend that the traditional preparation offered by schools of education adds little value to teachers' effectiveness in the classroom and, instead, imposes substantial costs that can deter talented individuals from entering teaching (Hess 2001). In response, many states have lowered the barriers to entering teaching by allowing teachers to participate in alternative certification programs, which allow people to start teaching before completing the certification-related coursework and student teaching that constitute the traditional route into teaching. However, most of these alternative certification programs, like most traditional certification programs, admit nearly all applicants (Mayer et al. 2003; Walsh and Jacobs 2007), so they raise the quantity of teachers without necessarily ensuring quality. In fact, evidence indicates that teachers from less selective alternative certification programs are no more effective than traditionally certified teachers at the elementary level (Constantine et al. 2009).

Teach For America (TFA) represents an innovative approach to supplying teachers to disadvantaged schools that differs from both the traditional and less selective alternative certification programs. Founded in 1989, the TFA program model comprises: (1) investing heavily in recruiting and screening, (2) providing a short but intensive teacher training program,

[1] For example, see Rockoff (2004), Hanushek et al. (2005), Rivkin, Hanushek, and Kain (2005), Aaronson, Barrow, and Sander (2007), and Chetty, Friedman, and Rockoff (2011).

and (3) providing additional support to new teachers. Like other alternative certification programs, it recruits people who typically do not have an education degree or other formal training in education. However, it is much more selective than typical alternative or traditional certification programs.[2] It recruits high-achieving college graduates who, through an intensive application and screening process, demonstrate characteristics that TFA believes are correlated with success in the classroom. It also differs from other certification programs in that it requires its participants, known as "corps members," to commit to only two years of teaching (although they can choose to remain longer). This increases the pool of potential recruits by including people who do not want to commit to a long-term career in teaching. Corps members participate in an intensive five-week training program before beginning their first teaching job. TFA provides ongoing mentoring and other support to the corps members throughout their two-year commitment. TFA encourages corps members who complete their two-year commitment (known as TFA alumni) to continue to work to address educational inequities, whether by continuing to teach or by assuming educational and other leadership positions.

TFA is a growing and important source of teachers in low-income schools. Since it placed its first cohort of 489 corps members in the 1990–91 school year, TFA has expanded considerably, and in the 2011–2012 school year, more than 9,000 first- and second-year TFA corps members were teaching in 43 urban and rural regions across the country. The program's growth is expected to continue—in 2010, TFA received a $50 million Investing in Innovation (i3) Scale-Up grant from the U.S. Department of Education to increase the size of its teacher corps by 80 percent, and to expand to up to 54 regions, by the 2014–15 school year.

Despite its growth as a source of teachers for high-poverty schools, TFA is highly controversial. One strand of criticism is that TFA teachers—and teachers from alternative certification programs more generally—are underprepared for teaching relative to teachers who have completed a traditional university-based teacher education program (Darling-Hammond 1990, 2000; Darling-Hammond et al. 2005). An additional criticism, not applicable to other alternative certification programs, is that, because TFA asks its teachers to make only a two-year commitment to teaching, its teachers are more likely to be inexperienced and therefore ineffective (Heilig and Jez 2010).

[2] The only other large, highly-selective alternative certification programs are the Teaching Fellows programs affiliated with the organization TNTP. These programs are similar to TFA except that they focus more on recruiting mid-career professionals and expect their teachers to make a long-term career in teaching rather than just a two-year commitment.

This paper presents the findings of a large-scale random assignment study of the effectiveness of math TFA teachers in middle and high schools. The study includes over 4,500 students and 136 teachers in 45 schools across 11 school districts in 8 states. Students were randomly assigned to a math class taught by a TFA teacher or to a math class taught by a teacher from some other program. We compare the end-of-year math achievement of students taught by TFA and non-TFA teachers to estimate the effectiveness of TFA teachers relative to non-TFA teachers. We find that math teachers from TFA are more effective than other teachers in the same schools, increasing student math achievement by an average of 0.07 standard deviations over the course of a school year. Addressing the concern about the limited experience of TFA teachers, we also find that novice TFA math teachers (those in their first or second year of teaching) are more effective than experienced non-TFA math teachers (those with five or more years of teaching experience) in the same schools. In essence, our findings show that the TFA program model can simultaneously boost the quantity and quality of teachers in hard-to-staff subjects within high-poverty schools.

The study focuses on TFA teachers teaching math in grades 6-12 (secondary math) because school districts report greater difficulties filling vacancies in secondary math (as well as science) than in other subjects (Ingersoll and Perda 2009). For high-poverty schools, this challenge is compounded by a net tendency for math teachers to transfer from high- to low-poverty schools (Ingersoll and May 2012). Moreover, a substantial number of TFA corps members—about 23 percent in the 2010–2011 school year—teach secondary math.

Despite the controversy surrounding TFA, rigorous evidence on the effectiveness of teachers from this program has been sparse. The only prior experimental study of the effectiveness of TFA teachers focuses on elementary schools. Decker, Mayer, and Glazerman (2004) randomly assigned nearly 1,800 students to either TFA teachers or teachers who received their certification through other routes within 17 elementary schools across 7 districts in 5 states. The results show that students of TFA teachers perform as well as students of non-TFA teachers in reading and score statistically better in math by approximately 0.15 standard deviations.

No experimental study of TFA teachers has been conducted at the secondary school level. Some recent nonexperimental studies have compared the achievement of secondary school students taught by TFA and non-TFA teachers based on a variety of analytic methods to control for possible nonrandom sorting of students to teachers. Using data from New York City, Boyd et al. (2006) find that students of middle school TFA teachers in their first year of teaching score higher in math, but lower in reading, than those of traditionally certified teachers in their first year of teaching, after controlling for students' prior scores, demographic covariates, and school fixed effects. With similar methods, Kane, Rockoff, and Staiger (2008) also use data from New York City and find that students of middle school TFA teachers score higher in math than students of traditionally certified teachers, and neither higher nor lower in reading. Using data from high schools in North Carolina, Xu, Hannaway, and Taylor (2011) exploit within-student, cross-subject variation in the certification route of students' teachers and find that TFA teachers raise student achievement relative to non-TFA teachers, especially achievement in science.

Our study contributes to this existing body of literature on the effectiveness of TFA teachers in three ways. First, whereas most prior work is potentially subject to bias from the sorting of students to teachers based on unobserved factors, our study is free of this threat due to its experimental design. Second, previous experimental evidence at the elementary level does not specifically address hard-to-staff subject areas; this study focuses on a subject area, secondary math, that high-poverty schools find particularly challenging to staff. Third, unlike all previous studies of TFA teachers at the secondary level, this study combines evidence from districts in multiple states.

The rest of the paper is organized as follows. Section I provides more details on TFA. In Section II, we describe our research design and data collection. Section III describes the schools, teachers, and students in the sample. We describe our estimation methods in Section IV. In Section V, we present the experimental findings on the effectiveness of TFA teachers. In Section VI, we explore whether easily observed credentials can explain the difference in effectiveness between TFA and non-TFA teachers, and we provide some conclusions in Section VII.

## I. Teach For America

The goal of TFA's recruitment process is to enroll people with characteristics that TFA believes are correlated with their becoming effective corps members: demonstrated leadership and achievement, perseverance, critical thinking skills, organizational ability, interpersonal skills, a strong dedication to TFA's mission, and respect for individuals' diverse experiences and ability to work effectively with people from diverse backgrounds. In addition, all corps members must be U.S. citizens or permanent residents, have an undergraduate grade point average (GPA) of 2.5 (although the average GPA of admitted corps members is about 3.6), and have a bachelor's degree from an accredited college or university prior to beginning the TFA summer training program.

TFA's admission process is intensive and highly selective. The process has three stages: (1) an online application, (2) a 25- to 45-minute telephone interview, and (3) a full-day in-person "final interview" in which the applicant participates in a one-on-one interview and is observed presenting a lesson and participating in a group discussion. To determine who should be screened out at each stage of the application process, TFA relies heavily, although not completely, on a regression model of achievement growth of the students of current and previous corps members. The explanatory variables in the model comprise over 20 characteristics of the corps members collected during the application process. TFA does not make public the parameters of this model, which is re-estimated annually. Of those applicants who submitted an online application in recent years, only about 12 percent were offered places in the program. Of those offered places, about 80 percent accepted.

Before teaching, corps members must complete TFA's pre-service training program. The core of this training is a five-week full-time "summer institute." At this institute, corps members attend courses on lesson planning, content delivery, classroom management, student assessment, literacy, and effective interactions with diverse populations. Corps members also lead small-group or whole-class instruction in classes at a local school district's summer school program under the supervision of a regular classroom teacher. Training also involves self-directed assignments before the summer institute and several-day meetings before and after the summer institute in the region in which the corps member will teach.

Although the process by which corps members are placed in schools varies by region, in all regions TFA plays an active role. TFA assigns corps members to a region based on the corps members' preferences, the needs of the region, and region-specific requirements for teachers, such as any special requirements for math teachers. TFA staff direct corps members to schools at which to interview and may discuss with district officials and principals how best to assign the corps members to schools.

Once corps members begin teaching, TFA continues to provide training and support for two years. In the 10 regions in our study, TFA provides an average of just over 40 hours of formal training to each corps member after he or she begins teaching. In addition, TFA assigns each corps member to a TFA staff person who observes the corps member teaching and then meets one-on-one with the corps member to provide feedback. TFA also schedules group meetings with corps members to provide additional guidance.

TFA typically does not provide teacher certification, so most TFA corps members need to enroll in a state-authorized alternative certification program operated by another organization such as a local university or school district. (In a few regions, TFA is a state-authorized certification provider and certifies its own corps members.) These programs may require corps members to participate in coursework prior to entering the classroom, although this is typically not intensive; virtually all programs require coursework during the first year of teaching, and some extend into the following summer or the second year of teaching.

Corps members are paid the same salary as other new teachers, but may receive additional financial support. As well as covering the costs of room and board during the summer institute and other meetings, TFA offers needs-based no-interest loans and grants to cover training, relocation, and testing and certification fees. Most TFA corps members at the time of our study were also eligible for AmeriCorps education awards of about $5,400 per year.

## II. Research Design and Data Collection

### *A. Experimental Design*

We conducted the study in two school years—the 2009-10 and 2010-11 school years—on separate cross sections of teachers and students. Before each study school year, we identified schools in which TFA teachers and teachers from other certification routes were teaching different classes (or "sections") covering the same math course. Just prior to the start of the school year in each participating school, we randomly assigned students who signed up for a particular math course to a class taught by a TFA teacher or a class taught by a comparison teacher who entered teaching through a traditional education or alternative certification program. Students who were assigned to a TFA teacher constitute the treatment group, and those who were assigned to a comparison teacher constitute the control group. The set of classes between which students were assigned formed a randomization block for the experimental design. Classes in the same randomization block covered the same course at the same level (for instance, honors Algebra I or remedial 6th grade math) in the same school and were typically taught during the same class period to facilitate random assignment.

All secondary math teachers who entered teaching through TFA were potentially eligible to be included in the study sample. This included teachers who were still fulfilling their two-year commitment to the program (TFA corps members) and those who remained in teaching after completing their two-year commitment (TFA alumni). The comparison teachers could have entered teaching through a traditional route to certification or through an alternative route that was not highly selective in its admissions—this allowed the sample to reflect the typical mix of non-TFA math teachers in the study schools. We excluded from the sample teachers who entered through a handful of alternative certification programs that we identified as being similar to TFA in terms of their selectivity.[3]

We did not impose any restrictions on the amount of prior teaching experience that teachers in the study could possess. Therefore, TFA and comparisons teachers who were compared in the study could (and did) have different experience levels. TFA teachers in the study had an average of two years of teaching experience, compared with an average of 10 years among the comparison teachers, consistent with the fact that TFA requires its teachers to make only a two-year commitment. Because we imposed no restrictions on teacher experience, the sample reflects differences in teaching experience of TFA and comparison teachers in the study schools. Therefore, the study design mimics the choice that a school administrator faces when selecting the type of teacher to fill a teaching position over the long run, given that relying on the group with higher expected turnover—TFA teachers—would imply that in steady state the position will be held by a less experienced teacher than otherwise would have occurred. Through this design, we can directly examine the common criticism that TFA teachers tend to be less effective than their counterparts from other programs due to their relative inexperience.

A total of 5,790 students were randomized in 111 randomization blocks with 136 math teachers in 45 schools, 11 districts, and 8 states. To obtain this sample, we recruited school districts with large concentrations of secondary math TFA teachers and then, within those districts, we contacted schools prior to each study year to determine their eligibility for the study and willingness to participate. Eligible schools were those with sets of TFA and comparison

[3] Most notably, teachers from the TNTP Teaching Fellows programs were excluded from the sample. As noted earlier, the Teaching Fellows programs are large, highly selective alternative route programs that follow a similar model to TFA; Teaching Fellows teachers were evaluated in a separate analysis in the Institute of Education Sciences-funded evaluation on which this paper is based. In practice, there was little overlap in the schools in which the TFA and Teaching Fellows teachers in the evaluation taught, and only one or two Teaching Fellows teachers who could have been included in a randomization block with a TFA teacher were excluded from this analysis. The other programs that were excluded from the study due to their similarity to TFA (and the Teaching Fellows programs) were small in number and size; we excluded five other programs that, collectively, trained only 14 math teachers nationwide in 2007 (Clark et al. 2008). By excluding these routes, we ensured that the teachers being compared in the study entered teaching through meaningfully different routes.

teachers teaching math classes that could form a randomization block for the study. Math courses eligible for inclusion included 6th, 7th, and 8th grade math; general high school math; Algebra I; Algebra II; and Geometry.

Before the start of each new school year, schools sent us lists of students whom they wanted placed into one of the classes in an identified randomization block, and we randomly assigned these students to classes. Because in most cases the classes were in the same period, the random assignment did not affect the students' class assignment or schedules for any other class. Schools could request specific assignments for a small number of students (for instance, students with disabilities whose Individualized Education Plans (IEPs) required them to be placed with particular teachers), in which case the students were excluded from the sample. In practice, this was rare, with only 26 students who were enrolled at the start of the school year exempted from random assignment. After school began, schools were asked to call a toll-free number to request immediate random assignment for any late enrolling students, up through at least the first month of school.

In general, we randomly assigned students between classes in a randomization block with equal probability, with a few exceptions. First, in randomization blocks in which a student had been exempted from random assignment and nonrandomly placed in a particular class, we randomly assigned the remaining students between the remaining available slots in the block—so they had a slightly lower probability of assignment to the class in which the exempted student had been placed. Second, after school began, if class sizes were imbalanced, we randomly assigned late enrolling students with slightly higher probability to the smaller classes, with the goal of ensuring that final class sizes were roughly equivalent within blocks (both to accommodate schools' preferences for balanced class sizes and to ensure comparability for the analysis). We adjusted for unequal probabilities of assignment within blocks through the use of sample weights, discussed further below.

To monitor movement in and out of the study classes, we asked the schools to send us updated class lists—essentially, enrollment snapshots—for the study classes at three times during the school year. From these lists, we were able to track study students moving out of the study classes and non-study students moving into the study classes.

## *B. Data Collection*

We measured student math achievement, the main outcome for the study, using scores from math assessments administered at the end of the school year in which the students were randomly assigned. For students in grades 6 to 8, we obtained scores on state-required assessments. For students in grades 9 to 12, we administered end-of-course computer adaptive math assessments developed by the Northwest Evaluation Association (NWEA) in the subject in which the student was enrolled (general high school math, Algebra I, Algebra II, or Geometry). The state assessments were expected to align closely with state curriculum standards, while the NWEA assessments were expected to be less prone to floor or ceiling effects for high- and low-achieving students, and less vulnerable to concerns about teaching to the test. We attempted to collect test data on all students in the study sample unless they moved out of the school district, including students who moved to a different class within the school and those who moved to a different school within the district. For comparability across tests, all scores were converted to $z$-scores. For middle school grades, the $z$-score was based on the statewide mean and standard deviation of scores in the grade level and year in which the assessment was administered; for the high school grades, the $z$-score was based on the national mean and standard deviation of scores for the NWEA assessments.

In addition to student math scores from the end of the study school year, we collected baseline reading and math scores (also converted to $z$-scores) from prior state assessments and demographic characteristics on all students from district records. Baseline scores were drawn from the most recent prior grade at which end-of-grade state assessments were administered.[4]

To collect information on the teachers in the study, we attempted to administer a survey to all 136 teachers in the study in the spring of each of the study school years; surveys were administered to 127 teachers (a response rate of 93 percent). We also collected teachers' scores from either the Praxis II Mathematics Content Knowledge Test (taken by the high school teachers in the sample, along with a few middle school teachers in states that allowed or required middle school teachers to take this test) or the Praxis II Middle School Mathematics Test (taken by the remaining middle school teachers in the sample). We administered the Praxis test to

[4] Baseline scores were drawn from the same grade level for all students in a randomization block. In randomization blocks that included students from more than one grade (for instance, a Geometry course that included both 10th and 11th graders), we identified the highest grade level that at least 90 percent of students in the block had reached, and then drew baseline scores from the highest prior grade in which end-of-grade state assessments were administered.

teachers who had not taken it previously and gathered existing scores from those who had, obtaining scores for 115 (84 percent) of the 136 teachers in the study.

### *C. Student Mobility and Attrition after Randomization*

The greatest threat to the internal validity of the study is non-random attrition from the original randomization sample. Attrition occurred whenever we could not obtain the end-of-year math score of a student in the randomization sample. This occurred in our study for four reasons: (1) parents did not provide consent for us to obtain state assessment scores (in middle schools) or administer the end-of-course test (in high schools); (2) students left the participating school district; (3) we were unable to administer the test to some high school students, mainly because they were absent from class and did not show up for a make-up test; and (4) school districts did not have state assessment data on some students.

We obtained end-of-year scores for 4,573 students out of the 5,790 students who were randomly assigned—about 79 percent of all students in the randomization sample (Table 1). Reassuringly, rates of mobility and nonmissing outcome data are very similar between treatment and control students, suggesting that student attrition from the study is unlikely to have been related to treatment status. The overall percentage of students with nonmissing outcome data is slightly higher in the treatment group (80 percent) than in the control group (79 percent).

Table 1 also highlights the fact that some students left their originally assigned classes during the school year. Slightly over three-fourths of students in the randomization sample were, as of the end of the study school year, still in the set of study classrooms and with their originally assigned type of teacher (TFA or non-TFA). Only 2 percent of students had switched to a study classroom taught by the "opposite" type of teacher—that is, cases in which students who were assigned to a TFA teacher switched to a non-TFA study teacher or students who were assigned to a non-TFA teacher switched to a TFA study teacher. The remaining sample members had transferred to a non-study classroom in the same school or left their original school. As shown in Table 1, specific types of mobility occurred with strikingly similar frequencies in the treatment and control groups; moreover, within each of those mobility groups, similar percentages of treatment and control students have nonmissing outcome data.

If assigned treatment status is truly random and attrition is truly unrelated to treatment status, then treatment and control students in the final analysis sample should have similar average values of baseline covariates. Table 2 shows that this is indeed the case. For 13 covariates measuring students' baseline achievement and demographic characteristics, none of the differences between treatment and control students are substantively meaningful or statistically significant at the 5 percent level. Taken together, the descriptive statistics for mobility rates, prevalence of nonmissing outcome data, and baseline covariate values strongly suggest that random assignment was properly implemented and attrition poses little threat to estimating the causal effects of TFA teachers. Later, in our analysis, we show that the maximal amount of selection bias that could have been introduced by attrition is not large enough to alter our main findings.

## III. Characteristics of Schools, Teachers, and Students in the Sample

### *A. Characteristics of Schools in the Sample*

Even though study schools were not randomly selected from the full set of secondary schools employing TFA teachers nationwide, the study schools are similar to secondary schools employing TFA teachers nationwide along many dimensions, and both the study schools and all secondary schools with TFA teachers are considerably more disadvantaged than the typical secondary school nationwide. For instance, both schools in the sample and secondary schools employing TFA teachers nationwide serve predominantly students from racial and ethnic minority groups—57 percent of students in both sets of schools are Black, and approximately 32 percent are Hispanic. Close to 80 percent of students at both types of schools are eligible for free or reduced-price lunch (compared with 51 percent at the typical secondary school nationwide).

The few differences between study schools and all TFA schools nationwide are likely due to study eligibility requirements. For instance, the average study school has significantly more students per grade than the average secondary school employing TFA teachers (240 versus 184 students per grade), consistent with the fact that schools with more students per grade were more likely to have multiple classes per subject taught during the same period to form randomization blocks. Similarly, although 23 percent of secondary schools with TFA placements nationwide are charter schools, there are no charter schools in the study sample. Charter schools are typically smaller than average and therefore less likely to have eligible randomization blocks.

*B. Characteristics of Teachers in the Sample*

The TFA teachers in the sample differ from the comparison teachers in many ways (Table 3). For instance, relative to comparison teachers, TFA teachers are younger (average age of 25 versus 38) and less likely to be members of racial or ethnic minorities (89 percent of TFA teachers are White and non-Hispanic, compared with only 30 percent of comparison teachers). TFA teachers are also considerably more likely to have graduated from a selective college or university (81 versus 23 percent).[5]

The TFA teachers display greater math content knowledge but are less likely to have majored in math (Table 3). TFA teachers who took the Praxis II Mathematics Content Knowledge Test outperformed comparison teachers by 22 points (or 0.93 standard deviations), and those who took the Praxis II Middle School Mathematics Test outperformed comparison teachers by 22 points (or 1.19 standard deviations). TFA teachers in the sample are less likely than comparison teachers to have majored in math (8 versus 26 percent) or secondary math education (0 versus 16 percent), but more likely to have majored in some other math-related subject (statistics, engineering, computer science, finance, economics, physics, or astrophysics) (27 versus 12 percent).

Not surprisingly given the fact that TFA asks its corps members to make only a two-year commitment to teaching, TFA teachers in the study have less teaching experience than comparison teachers (Table 3). As noted above, on average TFA teachers in our sample have an average of two years of experience compared with an average experience of 10 years among the non-TFA teachers. Eighty-three percent of the TFA teachers are in their first or second year of teaching, compared with 10 percent of comparison teachers. Seventy percent of the comparison teachers have been teaching more than five years, while none of the TFA teachers have been teaching this long. Consistent with the fact that they are more likely to be in their first or second year of teaching and thus likely still fulfilling coursework requirements for certification, TFA teachers are more likely than comparison teachers to have taken coursework during the study year (50 versus 21 percent).

As noted above, comparison teachers could have entered teaching through a traditional route to certification or an alternative certification program that was less selective than TFA. Fifty-nine

[5] Selective colleges are those ranked by *Barron's Profiles of American Colleges* as very competitive, highly competitive, or most competitive.

percent of comparison teachers are from traditional education programs, while 41 percent are from alternative certification programs.

### *C. Characteristics of Students in the Sample*

Consistent with TFA's goal of serving disadvantaged students, students in the study face multiple academic and socioeconomic disadvantages, as highlighted in Table 2. Students in the analysis sample had baseline achievement levels that are far below the average for their peers statewide: both treatment and control group students scored, on average, about half a standard deviation below the mean achievement in their states in both reading and math prior to the study period. Mirroring the demographic characteristics of their schools, students in the analysis sample are predominantly nonwhite and eligible for subsidized school meals.

## IV. Estimation Methods

### *A. Main Estimation Model*

To estimate the impacts of TFA teachers relative to comparison teachers, we estimate a regression model of the following form:

(1) $$y_{ijk} = \alpha_k + \beta_1 T_{ijk} + \mathrm{X}_{ijk}\beta_2 + \varepsilon_{ijk}$$

where $y_{ijk}$ is the end-of-year math test score of student *i* assigned to teacher *j* in randomization block *k*, $\alpha_k$ is a randomization block fixed effect, $T_{ijk}$ is a dummy variable for being randomly assigned to a TFA teacher, and $\mathrm{X}_{ijk}$ is a vector of student-level covariates. We use Huber-White standard errors that are robust to clustering at the teacher level.

We refer to the parameter of interest, $\beta_1$, as the impact of TFA teachers relative to comparison teachers—or, synonymously, the difference in effectiveness between TFA teachers and comparison teachers. This parameter is an intent-to-treat (ITT) effect, capturing the expected net difference in end-of-year math achievement from assigning a student to a TFA teacher rather than a comparison teacher at the beginning of the school year.

Although the covariates ($\mathrm{X}_{ijk}$) are not necessary to ensure unbiased impact estimates within our experimental design, we include them into equation (1) to improve precision. The covariates include all variables shown in Table 2.[6] Missing values of covariates are replaced with block-specific means, and we also include a vector of dummy variables (one for each covariate) indicating that we replaced the missing value with the block-specific mean for the covariate.

To ensure unbiased estimates of $\beta_1$, it is necessary to account explicitly for within-block differences among students in the probability of being assigned to the treatment group. As discussed previously, late enrollees to the study classrooms typically had different probabilities of assignment to the treatment group than early enrollees did. Without any correction, differences in assignment probabilities can lead to the overrepresentation of particular types of students in the treatment group relative to the control group. We eliminate this threat to causal validity by weighting students according to the inverse of their probability of assignment to the treatment group. Horvitz and Thompson (1952) show that this method recovers unbiased estimates. We scale the weights so that, within each combination of treatment status and block, the weights sum to one-half of the total number of students in the block. In our sensitivity analysis, we show that the presence or absence of weights does not discernibly influence the estimated effect.

We conduct tests of statistical significance using a conservative approach that guards against the tendency for Huber-White standard errors to inflate type I errors in finite samples (Donald and Lang 2007; Angrist and Pischke 2009). Specifically, our tests use a *t*-distribution with degrees of freedom equal to the number of teachers minus the number of covariates varying only at the teacher level—namely the treatment dummy and the randomization block dummies.[7]

### *B. Alternative Parameters of Interest*

Our ITT analysis attributes to each teacher the scores of all students assigned to his or her class at the beginning of the year. Because not all students stayed in their originally assigned classes—as documented by Table 1—the ITT impacts are not equivalent to the impacts of being taught by a TFA teacher for a full school year.[8] In this paper, we focus primarily on the ITT estimates for

[6] We also include four dummy variables denoting that a specified number of years (1, 2, 3, and 4 or more) have elapsed between the baseline math test and the outcome test.

[7] Teachers are not always fully nested within randomization blocks; some teachers span multiple blocks if they taught more than one class in the study. For these degrees-of-freedom calculations only, we amalgamate blocks that share the same teacher so that the resulting block indicators vary only at the teacher level.

[8] Table 1 documents mobility in the sample of students who were randomly assigned. However, it is the mobility of students in the *analysis* sample—students with outcome data—that determines the degree of discrepancy between the ITT estimate and the estimated effect of being taught by a TFA teacher. Mobility is less prevalent in the analysis sample than in the randomization sample, because students who remained in

two main reasons. First, from the perspective of a school administrator, it is easy to assign a student to a given teacher at the beginning of the school year, but whether the student stays with that teacher is influenced, in part, by factors beyond the administrator's control. The ITT impact realistically reflects the potential for mobility to dilute the effects of a student's initially assigned teacher, so it can be considered the most relevant parameter to inform a school administrator's choice between hiring different types of teachers. Second, as we discuss below, we have only imperfect measures of the amount of time for which a student was actually taught by a specified teacher, whereas a student's initial assignment is known with certainty.

Despite our primary focus on the ITT impact, we also explore the estimation of an alternative parameter: the effect of a student's actual duration of being taught by a TFA math teacher on his or her math achievement. This parameter more faithfully captures the instructional ability of TFA teachers relative to comparison teachers, independent of student mobility.

Duration of exposure to a particular type of teacher is potentially endogenous; students may switch classes as a result of preferences by parents, school administrators, and the students themselves in response to unobserved factors correlated with academic achievement. Nevertheless, students' initial, randomly determined assignment to teachers is an exogenous source of variation in actual exposure to TFA teachers. Letting $D_{ijk}$ denote a measure of students' duration of enrollment with a TFA teacher, we can estimate the impact of $D_{ijk}$ on student achievement by applying two-stage least squares to the structural model of interest,

(2) $$y_{ijk} = \alpha_k + \delta_1 D_{ijk} + \mathrm{X}_{ijk}\beta_2 + u_{ijk}\,,$$

using $T_{ijk}$ as an instrument for $D_{ijk}$ in the first-stage equation

(3) $$D_{ijk} = \alpha_k + \pi_1 T_{ijk} + \mathrm{X}_{ijk}\pi_2 + \omega_{ijk}\,.$$

The coefficient $\delta_1$ in equation (2) is a local average treatment effect (LATE), capturing the average effect of duration with a TFA math teacher on students' math achievement within a particular population of students: those whose duration with a TFA teacher was affected by their randomly assigned status (Imbens and Angrist 1994; Angrist and Imbens 1995).[9] These students, known as "compliers," either experienced a longer duration with a TFA teacher by being assigned to the treatment group than they would have experienced if assigned to the control

their originally assigned classrooms are disproportionately more likely to have outcome data. Nevertheless, not all students with outcome data stayed with their originally assigned type of teacher.

[9] More specifically, as shown by Angrist and Imbens (1995), the LATE is a weighted average of impacts across all of the possible increments in exposure that were induced by the experiment, with weights proportional to the number of compliers exhibiting each increment in exposure.

group, or experienced a shorter duration with a TFA teacher by being assigned to the control group than they would have experienced if assigned to the treatment group.

One limitation in estimating equation (2) is that we collected a snapshot of students' enrollment in math classes at only three points during the school year: (1) in the fall, about two to four weeks after random assignment, (2) at the beginning of the spring semester, and (3) toward the end of the spring semester. Given these data, we define $D_{ijk}$ as the fraction of enrollment snapshots in which a student is observed to be taught by a TFA teacher; the variable can take on the values of 0, 1/3, 2/3, and 1. Therefore, the LATE coefficient, $\delta_1$, represents the expected difference in math achievement from being taught by a TFA teacher at all enrollment snapshots (loosely interpretable as a full school year) rather than at no enrollment snapshots.

To construct $D_{ijk}$, it is important to account for missing enrollment information. For students who left the set of study classrooms before the end of the school year—either by transferring to a non-study classroom or leaving the school entirely—we do not know what types of teachers they had after their departure, even if we know their spring test scores. Twelve percent of students in the analysis sample are missing information from at least one snapshot. Nevertheless, we can make extreme assumptions about the teacher assignments that mobile students had after leaving the study classrooms, which imply upper and lower bounds for the degree to which students complied with their assigned treatment status throughout the school year. First, we assume that all departing students moved to a class taught by the same type of teacher as that to which they were originally assigned (that is, departing students in the treatment group were subsequently taught by a TFA teacher and departing students in the control group were subsequently taught by a non-TFA teacher), leading to an upper bound for $\pi_1$ in equation (3). Second, we assume that departing students were subsequently taught by the "opposite" type of teacher to their original assignment (that is, departing students in the treatment group were subsequently taught by a non-TFA teacher and departing students in the control group were subsequently taught by a TFA teacher), leading to a lower bound for $\pi_1$. Since $\delta_1 = \beta_1 / \pi_1$, the upper and lower bounds for $\pi_1$ lead, respectively, to lower and upper bounds for the LATE.

## V. Experimental Findings

### *A. Main Estimates*

Table 4 shows the difference in effectiveness teaching secondary math between TFA teachers and comparison teachers, estimated from the experimental design. The ITT estimate in column 1 represents our main estimate for the impact of TFA teachers. On average, TFA teachers are more effective than comparison teachers teaching the same math courses in the same schools. Students assigned to TFA teachers score 0.07 standard deviations higher on end-of-year math assessments than students assigned to comparison teachers.

The magnitude of the difference in effectiveness between TFA and comparison teachers can be interpreted in several ways. First, the effect size can be expressed as a change in percentiles of achievement within the statewide or national reference populations that took the same math assessment. If assigned to a comparison teacher, the average student in the study would have had a *z*-score of -0.60, equivalent to the 27th percentile of achievement in the reference population based on a normal distribution for test scores. If assigned to a TFA teacher, this student would, instead, have had a *z*-score of -0.52—equivalent to the 30th percentile. Thus, the average student in the study gains 3 percentile points from being assigned to a TFA teacher rather than a comparison teacher.

Alternatively, the effect size can be compared with educationally relevant benchmarks. An illustrative benchmark is the average one-year gain in achievement exhibited by students on nationally normed assessments in grades 6 through 11, which Hill et al. (2008) calculates to be 0.27 standard deviations. On the basis of this benchmark, TFA teachers' effect of 0.07 standard deviations on math scores amounts to 26 percent of an average year of learning by students nationwide, or 2.6 months of learning in a 10-month school year.

The remaining columns of Table 4 explore different ways of scaling up the ITT estimate into a LATE estimate, capturing the impact of enrollment duration with a TFA math teacher on the math achievement of compliers—those students whose duration was affected by their assigned treatment status. As discussed in Section IV, the different approaches correspond to different assumptions about which types of teachers taught mobile students after they left the study classrooms. Under assumptions that imply the maximal level of compliance with assigned treatment status, the first-stage coefficient is 0.96; that is, assignment to the treatment group, instead of the control group, increased by 96 percentage points the fraction of enrollment snapshots at which students were taught by a TFA teacher. With this high level of compliance, there is no material difference between the LATE estimate and the ITT estimate. The alternative assumptions that imply a lower bound for compliance yield a first-stage coefficient of 0.80 (column 4) and a resulting LATE of 0.09 standard deviations. That is, among compliers, being taught by TFA teachers at all enrollment snapshots raises math achievement by 0.09 standard deviations compared with being taught by comparison teachers at all snapshots. Whatever the assumptions regarding the types of teachers who taught leavers, compliance with assigned treatment status is high, leading to little discrepancy between the ITT and LATE. Therefore, for the remainder of this paper, we focus on the ITT estimates, given that compliance is high and those estimates do not rely on assumptions about the enrollment behavior of students who left the study classrooms during the school year.

### *B. Sensitivity Analyses*

Table 5 shows robustness checks for estimating the effects of TFA teachers. The key conclusion from the main estimates—that TFA teachers are more effective than comparison teachers—is robust to several changes in the specification of the estimation model or sample.

First, we explore two simple changes to our estimation approach: excluding all covariates except randomization block dummies (row 1 of Table 5) and omitting the analysis weights that account for treatment assignment probabilities (row 2 of Table 5). The estimated effects do not discernibly change with these modifications to the estimation approach, providing yet another indication that covariates are balanced between the treatment and control groups and assignment probabilities do not vary sufficiently to alter the estimates in any material way.

Next, we consider the threat to internal validity posed by the presence of students in the study classrooms who were not randomly assigned. Students in the study classrooms who had not been randomly assigned were excluded from the analysis, but their presence could, in theory, have affected the achievement of randomly assigned students via peer effects. We were largely successful in minimizing nonrandom placements into the study classrooms during the random assignment period, which lasted through the first month of school. Of students who enrolled in the study classes during this time, only 2 percent were nonrandomly placed into those classes, usually as a result of schools' requests to exempt students from random assignment or the schools' failure to request assignments for late-enrolling students. However, after the first month of school, schools were free to place newly enrolling students without random assignment. At the final enrollment snapshot, 20 percent of the students enrolled in the study classes had not been randomly assigned to those classes, with identical percentages for treatment and control classes.

Although treatment and control classes had similar proportions of nonrandomly placed students, there is still the possibility of unobserved differences in the *types* of students who were nonrandomly placed into those classes, which could threaten the internal validity of estimated effects on the randomization sample. Therefore, we conduct a robustness check to drop randomization blocks with the greatest potential for this threat based on high rates of nonrandom placement. Specifically, we drop blocks in which students who entered the study classes through a method other than random assignment constituted more than 10 percent of students enrolling before the end of the first month of school or more than 25 percent of students on the final enrollment snapshot. This criterion results in the exclusion of 30 percent of the blocks in the sample. Nevertheless, in the remaining blocks, the estimated effect of TFA teachers, 0.06 standard deviations, is similar to the full-sample estimate (row 3 of Table 5).

Another complicating factor in the study design is the presence of supplemental math classes—separate from the regular math classes included in this study—that some schools offered to reinforce material taught in the regular classes. Schools' criteria for deciding which students to assign to supplemental math classes varied. This study purposely did not include any schools that made such decisions *after* the start of the school year, because assignments to supplemental classes could then be an endogenous response to compensate for poor teaching by students' regular math teachers. However, 48 percent of the randomization blocks in the study included at least some students who were assigned by their schools to supplemental math classes before we conducted random assignment to the main classes. In most cases, we did not require students to take supplemental classes with the same type of teacher (TFA or non-TFA) as the teacher who taught their main classes, so the presence of supplemental classes generally diluted the treatment-control contrast in the types of math teachers to whom students were exposed. Therefore, as a robustness check, we remove all randomization blocks with supplemental math instruction. The estimated effect of TFA teachers rises to 0.11 standard deviations (row 4 of Table 5).

Our final set of sensitivity analyses assesses the degree to which attrition may have introduced selection bias into the main estimate. Using the approach developed by Lee (2009), we estimate lower and upper bounds for the estimated effect that account for the maximal possible amount of selection bias. Recall, from Table 1, that outcome data were obtained for a slightly higher percentage of the randomization sample in the treatment group (79.5 percent) than in the control group (78.5 percent). Thus, the *analysis* sample in the treatment group may have a slightly different mix of students than the analysis sample in the control group. Following the monotonicity assumption in Lee (2009), we assume that any student who would have outcome data if assigned to the control group would also have outcome data if assigned to the treatment group. This implies that the treatment analysis sample includes all individuals who would have been in the analysis sample if assigned to the control group, plus an excess set of individuals who were newly induced to have outcome data from being assigned to the treatment group. These excess individuals constitute 1.2 percent [= (79.5-78.5)/79.5] of the treatment analysis sample.

Removing the excess individuals from the treatment analysis sample would restore treatment-control equivalence in the mix of individuals examined, but it is not possible to identify the excess individuals. However, by making extreme assumptions for the rankings of the excess individuals' outcomes within the treatment analysis sample, we estimate bounds for the effects of TFA teachers. Specifically, we employ a two-step method. First, we estimate equation (1) on the full analysis sample and obtain the residuals. We then use two alternative ways to trim the treatment analysis sample: removing either students whose residuals are in the top 1.2 percent or those whose residuals are in the bottom 1.2 percent of the treatment analysis sample. Second, we re-estimate equation (1) on the two trimmed samples, obtaining lower and upper bounds for TFA teachers' average effect.[10]

Bounds for the estimated effect of TFA teachers yield the same qualitative conclusion as the main estimate: TFA teachers are more effective than comparison teachers even after accounting for attrition-induced bias. At worst, TFA teachers raise secondary students' math achievement by an average of 0.05 standard deviations relative to comparison teachers (row 5 of Table 5); at best, TFA teachers raise achievement by an average of 0.11 standard deviations (row 6 of Table 5).

### *C. Effects within Teacher Subgroups*

The effects presented in Table 4 are averaged over a heterogeneous mix of both TFA and comparison teachers. In a given hiring decision, a school administrator may be faced with more

[10] This two-step method is an extension of the original approach specified by Lee (2009). Lee's original approach considers the case of no covariates; the approach can also be applied repeatedly within strata defined by covariates, with the final upper (or lower) bounds in the full sample being a weighted average of the upper (or lower) bounds obtained from the different strata. We do not form strata based on covariates due to the large number of covariates in our analysis and the fact that some of the key covariates are continuous. The two-step method described here accommodates any types or number of covariates as long as the functional form of the first-step equation is correct.

specific choices between TFA and non-TFA teachers with particular characteristics. To shed light on these choices, we estimate the effects of TFA teachers within sets of randomization blocks in which the TFA and comparison teachers exhibit particular configurations of characteristics. For all subgroups of randomization blocks considered, we find that TFA teachers raise student math achievement relative to their counterparts from other routes.

First, we consider the route through which the comparison teachers entered teaching. One criticism of alternative certification programs is that they provide insufficient preparation relative to traditional teacher preparation programs. To explore the validity of this criticism as it applies to TFA teachers, we estimate the effects of TFA teachers within randomization blocks in which TFA teachers are compared with teachers from traditional routes. We find no basis for this criticism; in fact, students of TFA teachers outperform those of traditionally certified teachers by 0.06 standard deviations (row 1 of Table 6). In a parallel analysis, we also find that students of TFA teachers outperform students of alternatively certified comparison teachers by 0.09 standard deviations (row 2 of Table 6).

Another common criticism of TFA is that it seeks teachers willing to make a two-year commitment to teaching, but does not expect them to remain over the longer term. Critics claim that too many TFA teachers leave teaching before they accumulate the experience needed to be as effective as their counterparts from other routes (Heilig and Jez 2010). We therefore consider a comparison that, based on the logic of this claim, ought to be most unfavorable to finding a positive effect of TFA teachers: inexperienced TFA teachers compared with experienced comparison teachers. From the perspective of a school administrator deciding how to fill a teaching position over the long run, this comparison mimics a worst case for hiring TFA teachers—one in which TFA teachers always leave at the end of two years and must be replaced by another inexperienced TFA teacher—versus the best case for hiring non-TFA teachers, in which non-TFA teachers will stay and become experienced.

We specify this analysis by identifying randomization blocks in which TFA teachers in their first two years of teaching are compared with non-TFA teachers with more than five years of teaching experience. Estimates from these randomization blocks indicate that inexperienced TFA teachers raise student math achievement relative to experienced comparison teachers, with an estimated effect similar to that in the full sample (row 3 of Table 6). In other words, high-poverty secondary schools should expect higher math achievement from hiring TFA teachers rather than non-TFA teachers for a given position, even if frequent turnover among TFA teachers would necessitate repeatedly filling the position with an inexperienced TFA teacher.

We also estimate effects separately within middle school grades (grades 6 to 8) and high school grades (grades 9 to 12), given the distinct contexts in the two grade spans. In particular, high school courses covered more advanced math, for which effective teaching might require different sets of instructional skills than the teaching of less advanced math. Moreover, the assessments taken by middle school students in the study were high-stakes, in that they served as inputs into school accountability measures, whereas the study-administered assessments taken by high school students were low-stakes. Despite these differences, our basic conclusion holds in both grade spans: TFA teachers have positive impacts relative to comparison teachers in both middle schools and high schools. Students of TFA teachers outscore those of comparison teachers by 0.06 standard deviations in middle schools (row 4 of Table 6) and 0.13 standard deviations in high schools (row 5 of Table 6).

## VI. Accounting for the Effect of TFA Teachers

In this section, we reconsider the debate about whether the quality of the teacher workforce can be improved by toughening requirements for the credentials that teachers must have in order to lead a classroom. One approach to informing this debate is to examine whether known differences in effectiveness between distinct groups of teachers can be statistically explained—in other words, predicted—by differences in their credentials or other easily observed aspects of their education and training. Insofar as credential gaps could account for effectiveness gaps between groups of teachers, there would be more support for the hypothesis that raising teacher credentials could improve the performance of the teacher workforce.

Our experimental analysis has shown that TFA teachers are, on average, more effective than comparison teachers. Here, we use nonexperimental methods to examine the following question: To what extent could this difference in effectiveness have been predicted solely based on differences in credentials between these groups of teachers? To address this question, we assess the degree to which TFA teachers differ from non-TFA teachers in the prevalence of characteristics that are correlated with effectiveness. Because this analysis relies on correlations between credentials and effectiveness that do not have a causal interpretation, we regard this analysis as having the potential to produce suggestive, but not conclusive, evidence for the reasons TFA teachers are more effective than comparison teachers.

We consider a set of characteristics that could be readily observable on a teacher's resume at the time that a school administrator is making a hiring decision. These characteristics belong to four broad categories: (1) a measure of teachers' general academic ability, based on the selectivity ranking of their undergraduate institution; (2) measures of teachers' exposure to and knowledge of mathematics, based on the quantity of completed math coursework, prior use of math in a nonteaching job, and scores on the Praxis II tests of math knowledge; (3) measures of teachers' instructional training, including the extent of prior math pedagogy coursework, student teaching, and ongoing coursework during the school year; and (4) measures of teaching experience.

Table A.1 lists the specific variables in the analysis along with their sample means and standard deviations within the student-level analysis sample. All of these variables are based on self-reports from the teacher survey or on Praxis II scores collected by the study team. Given abundant evidence that the gains to experience decline with total experience (Hanushek et al. 2005; Rivkin et al. 2005), we capture teaching experience with a three-piece linear spline that allows for different marginal gains to experience in three different ranges of total experience—one to two years, three to five years, and more than five years.

In order for any of those characteristics to account for a positive portion of the difference in effectiveness between TFA and comparison teachers, two conditions are necessary. First, the characteristic must be associated with teacher effectiveness. Second, relative to non-TFA teachers, TFA teachers must show a greater extent of the characteristic if it is positively related to effectiveness, or a lesser extent if it is negatively related to effectiveness. We assess each of these two conditions in turn.

Accordingly, our analysis begins with estimating the associations between teacher characteristics and teacher effectiveness. We add a full vector of observed teacher characteristics to the main experimental estimation model:

$$(4) \qquad y_{ijk} = \alpha_k + \gamma_1 T_{ijk} + X_{ijk}\gamma_2 + C_{jk}\gamma_3 + \varepsilon_{ijk} \,,$$

where $C_{jk}$ is the full vector of teacher characteristics and all other variables are defined as in equation (1).[11] Although *student* math scores are the dependent variable, differences in student

[11] Measures of teacher characteristics have some missing data resulting from teachers' nonresponse to survey items and nonparticipation in Praxis II assessments. We account for this missing data by replacing missing values with imputed values from multiple imputation (Rubin 1987). Imputed values of a specified variable are stochastic draws from the conditional distribution of the variable given all other variables in the analysis. This procedure has the advantage of preserving the observed distributions of all variables and the observed relationships among the variables (Schafer and Graham 2002). Clark et al. (2013) provides details on the imputation models used in the analysis.

achievement across classrooms within randomization blocks are unbiased estimates of differences in teacher effectiveness due to random assignment. Therefore, because equation (4) includes block fixed effects, $\gamma_3$ can be properly interpreted as the association between teacher characteristics and teacher effectiveness.

Estimates of $\gamma_3$, shown in Table 7, indicate that few teacher characteristics readily observable on a resume are predictive of teacher effectiveness. Of the 11 variables measuring teacher characteristics, only two have a statistically significant association with effectiveness. First, second-year teachers are more effective than first-year teachers. Students assigned to second-year teachers are predicted to score 0.14 standard deviations higher than those assigned to first-year teachers, consistent with previous evidence that the largest gain in effectiveness from experience occurs between the first and second years of teaching (Hanushek et al. 2005; Boyd et al. 2006; Kane et al. 2008). Second, teachers' effectiveness is *negatively* associated with the amount of job-related coursework (for certification or advanced degrees) that they take during the school year. A teacher who takes 180 hours of coursework during the year—the average for teachers in the study who take any coursework at all—is predicted to lower student math achievement by 0.05 standard deviations relative to a teacher who takes no concurrent coursework.

Next, we assess the direction and extent to which TFA teachers differ from comparison teachers on each of the two characteristics found to be associated with effectiveness. We estimate variants of equation (1) in which each of the two teacher characteristics, rather than student test scores, serve as the dependent variable, producing estimates of the within-block difference in the given characteristic between TFA and comparison teachers.[12] The first column of Table 8 shows estimates of the between-group differences. Consistent with the descriptive statistics discussed earlier in Section III, TFA teachers exhibit a significantly lower likelihood of having acquired a second year of teaching experience and are taking, on average, more education-related coursework during the school year (albeit not by a statistically significant margin).

The final column of Table 8 shows the *predicted* difference in effectiveness between TFA and comparison teachers—still expressed in student-level standard deviations—based solely on each

[12] We use the student-level analysis sample for estimating differences in characteristics between TFA and comparison teachers in order to maintain consistency with the level of analysis used in equation (4). However, conclusions are unchanged if we estimate these differences on the teacher-level data instead.

of the two characteristics found to be related to effectiveness. Each predicted difference represents the impact that would be expected based on the given characteristics. The predicted difference is equal to the product of the between-group difference in the characteristic (column 1) and the characteristic's association with effectiveness (redisplayed in column 2).

The negative values in the final column of Table 8 indicate that the observed credentials do not explain why TFA teachers are more effective than comparison teachers. Although teachers who acquire a second year of teaching experience are generally more effective than those who have not yet done so, TFA teachers are less likely than comparison teachers to have acquired a second year of teaching experience. Similarly, although the amount of concurrent coursework that teachers take is negatively related to effectiveness, TFA teachers take more concurrent coursework than comparison teachers (although the difference is not statistically significant). Based on these two characteristics alone, TFA teachers would have been predicted to be less effective than their counterparts from other routes into teaching—when, in fact, they are more effective.

## VII. Conclusions

Our study provides strong evidence that TFA math teachers in middle and high schools with highly disadvantaged students are more effective than the other math teachers teaching the same courses in the same schools. While the difference in effectiveness is not large enough to bring end-of-year scores on math assessments of disadvantaged students to the mean for the wider population, the difference in effectiveness is meaningful. Our estimate is that the difference is equivalent to about 2.6 months of math instruction. It is particularly striking that novice TFA teachers are more effective than even the more experienced non-TFA teachers who are teaching in the same schools. Thus, other attributes that boost the effectiveness of TFA teachers relative to non-TFA teachers outweigh the negative effect of the TFA teachers' relative inexperience. These findings are relevant to policymakers considering funding TFA, school districts considering using TFA teachers, and school administrators who select teachers.

Understanding *why* TFA teachers are more effective is of even greater policy interest as it may suggest more effective approaches to recruiting, screening, training, and supporting teachers in general. Our experimental study is unable to separate out the effect of TFA's recruiting and screening approach from the effect of its training and support. While we find TFA teachers have many different characteristics from non-TFA teachers in the same school, the reasons why TFA teachers are more effective than non-TFA teachers do not appear to lie in any of their credentials that would have been easily observed on a resume. Few credentials are associated with effectiveness at all, and for those that are—experience and not taking coursework while teaching—TFA teachers are at a distinct disadvantage relative to non-TFA teachers. Further research is needed to explore the extent to which TFA teachers' effectiveness is due to TFA's recruitment strategies, its approaches to identifying and selecting candidates based on criteria that go beyond easily observed credentials, and its methods for training and supporting its teachers.

## REFERENCES

Aaronson, Daniel, Lisa Barrow, and William Sander. “Teachers and Student Achievement in the Chicago Public High Schools.” *Journal of Labor Economics*, vol. 25, no. 1, 2007, pp. 95-135.

Angrist, Joshua D. and Guido. W. Imbens. “Two–Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity.” *Journal of the American Statistical Association*, vol. 90, no. 430, 1995, pp. 431-442.

Angrist, Joshua D., and Jorn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press, 2009.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. “How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement.” *Education Finance and Policy*, vol. 1, no. 2, 2006, pp. 176–216.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.” NBER Working Paper No. 17699. Cambridge, MA: National Bureau of Economic Research, December 2011.

Clark, Melissa A., Hanley S. Chiang, Tim Silva, Sheena McConnell, Kathy Sonnenfeld, Anastasia Erbe, and Michael Puma. “The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs.” Princeton, NJ: Mathematica Policy Research, September 2013.

Clark, Melissa, Sheena McConnell, Kristin Hallgren, Daniel Player, and Alison Wellington. “Evaluating Highly Selective Programs That Provide Alternative Routes to Teacher Certification: Feasibility and Design Issues.” Princeton, NJ: Mathematica Policy Research, March 28, 2008.

Constantine, Jill, Daniel Player, Tim Silva, Kristin Hallgren, Mary Grider, and John Deke. “An Evaluation of Teachers Trained Through Different Routes to Certification. Final Report.” Princeton, NJ: Mathematica Policy Research, 2009.

Darling-Hammond, Linda. “Teaching and Knowledge: Policy Issues Posed by Alternate Certification for Teachers.” *Peabody Journal of Education*, vol. 67, no. 3, Spring 1990, pp. 123-154.

Darling-Hammond, Linda. “How Teacher Education Matters.” *Journal of Teacher Education*, vol. 51, no. 3, May/June 2000, pp. 166-73.

Darling-Hammond, Linda, Deborah J. Holtzman, Su Jin Gatlin, and Julian Vasquez Heilig. "Does Teacher Preparation Matter? Evidence About Teacher Certification, Teach For America, and Teacher Effectiveness." *Education Policy Analysis Archives*, vol. 13, no. 42, October 2005, pp. 1068–2341.

Decker, Paul T., Daniel P. Mayer, and Steven Glazerman. "The Effect of Teach For America on Students: Findings from a National Evaluation." Princeton, NJ: Mathematica Policy Research, 2004.

Donald, Stephen G. and Kevin Lang. "Inference with Difference-in-Differences and Other Panel Data." *The Review of Economics and Statistics*, vol. 89, no. 2, 2007, pp. 221–233.

Duncan, Greg and Richard Murnane, editors. *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances.* New York: Russell Sage Foundation Press, 2011.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven Rivkin. "The Market for Teacher Quality." NBER Working Paper 11154. Cambridge, MA: National Bureau of Economic Research, February 2005.

Heilig, Julian Vasquez, and Su Jin Jez. "Teach For America: A Review of the Evidence." East Lansing, MI: The Great Lakes Center for Education Research and Practice, June 2010.

Hess, Frederick M. "Tear Down This Wall: The Case for a Radical Overhaul of Teacher Certification." Washington, DC: Progressive Policy Institute, November 2001.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*, vol. 2, no. 3, 2008, pp. 172–177.

Horvitz, D. G. and D. J. Thompson. "A Generalization of Sampling Without Replacement From a Finite Universe." *Journal of the American Statistical Association*, vol. 47, no. 260, 1952, pp. 663–685.

Jacob, Brian A. "The Challenges of Staffing Urban Schools with Effective Teachers." *The Future of Children*, vol. 17, no. 1, 2007, pp. 129–153.

Imbens, Guido W. and Joshua D. Angrist. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, vol. 62, no. 2, 1994, pp. 467–475.

Imbens, Guido W., and Donald B. Rubin. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies*, vol. 64, 1997, pp. 555-574.

Ingersoll, Richard M., and Henry May. "The Magnitude, Destinations, and Determinants of Mathematics and Science Teacher Turnover." *Educational Evaluation and Policy Analysis*, forthcoming, 2012.

Ingersoll, Richard M., and David Perda. "The Mathematics and Science Teacher Shortage: Fact and Myth." CPRE Research Report #RR-62. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education, 2009.

Jacob, Brian A. "The Challenges of Staffing Urban Schools with Effective Teachers." *The Future of Children*, vol. 17, no. 1, 2007, pp. 129–153.

Kane, Thomas, Jonah E. Rockoff, and Douglas Staiger. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review*, vol. 27, 2008, pp. 615–631.

Lee, David. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, 2009, vol. 76, no. 3, pp. 1071–1102.

Mayer, Daniel P., Paul T. Decker, Steven Glazerman, and Timothy W. Silva. "Identifying Alternative Certification Programs for an Impact Evaluation of Teacher Preparation." Cambridge, MA: Mathematica Policy Research, April 2003.

Monk, David H. "Recruiting and Retaining High-Quality Teachers in Rural Areas." *The Future of Children*, vol. 17, no. 1, 2007, pp. 155–174.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. "Teachers, Schools, and Academic Achievement." *Econometrica*, vol. 73, no. 2, March 2005, pp. 417–458.

Rockoff, Jonah. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review: AEA Papers and Proceedings*, vol. 94, no. 2, 2004, pp. 247-252.

Rubin, Donald B. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley, 1987.

Schafer, Joseph L., and John W. Graham. "Missing Data: Our View of the State of the Art." *Psychological Methods*, vol. 7, no. 2, 2002, pp. 147–177.

Walsh, Kate, and Sandi Jacobs. "Alternative Certification Isn't Alternative." Washington, DC: Thomas B. Fordham Institute, September 2007.

Xu, Zeyu, Jane Hannaway, and Colin Taylor. "Making a Difference? The Effects of Teach For America in High School." *Journal of Policy Analysis and Management*, vol. 30, 2011, pp. 447-469.

**Table 1. Rates of Mobility and Nonmissing Outcome Data within the Randomization Sample**

| | Percentages of Students | |
|---|---|---|
| | Assigned to TFA Teacher | Assigned to Comparison Teacher |
| All Students | 100.0 | 100.0 |
| Has valid end-of-year score | 79.5 | 78.5 |
| Stayed in study classrooms and with originally assigned type of teacher | 77.5 | 77.4 |
| And has valid end-of-year score | 69.0 | 68.4 |
| Stayed in study classrooms but switched to opposite type of teacher | 2.2 | 2.4 |
| And has valid end-of-year score | 1.8 | 1.8 |
| Transferred to non-study classroom in the same school | 7.9 | 7.7 |
| And has valid end-of-year score | 4.4 | 3.6 |
| Left study school | 12.5 | 12.5 |
| And has valid end-of-year score | 4.2 | 4.6 |
| Number of Students in the Randomization Sample | 2,884 | 2,906 |

**Table 2. Summary Statistics of Baseline Covariates in the Analysis Sample**

| | Treatment Mean | Control Mean | Difference |
|---|---|---|---|
| **Baseline Achievement, Expressed as *z*-Score within the Statewide Distribution** | | | |
| Baseline math score | -0.512 | -0.504 | -0.008 |
| | [0.870] | [0.853] | (0.013) |
| Baseline reading score | -0.514 | -0.510 | -0.005 |
| | [0.908] | [0.893] | (0.014) |
| **Demographic Group Dummy Variables** | | | |
| Old for grade | 0.073 | 0.064 | 0.009* |
| | [0.261] | [0.245] | (0.005) |
| Grade is below modal grade in randomization block | 0.011 | 0.014 | -0.003 |
| | [0.113] | [0.117] | (0.001) |
| Grade is above modal grade in randomization block | 0.021 | 0.016 | 0.005 |
| | [0.143] | [0.127] | (0.003) |
| Retained in same grade between previous and current year | 0.022 | 0.024 | -0.002 |
| | [0.146] | [0.154] | (0.003) |
| Female | 0.486 | 0.500 | -0.015 |
| | [0.500] | [0.500] | (0.009) |
| Black, non-Hispanic | 0.621 | 0.625 | -0.004 |
| | [0.487] | [0.484] | (0.008) |
| Hispanic | 0.283 | 0.277 | 0.005 |
| | [0.452] | [0.448] | (0.008) |
| Non-black, non-Hispanic | 0.096 | 0.098 | -0.002 |
| | [0.300] | [0.297] | (0.006) |
| Receives free or reduced-price lunch | 0.899 | 0.905 | -0.007 |
| | [0.305] | [0.293] | (0.009) |
| English language learner | 0.080 | 0.084 | -0.004 |
| | [0.276] | [0.277] | (0.006) |
| Has an Individualized Education Program | 0.064 | 0.060 | 0.004 |
| | [0.245] | [0.237] | (0.005) |
| Number of Students | 2,292 | 2,281 | |

Note: In the columns for treatment and control means, standard deviations are in brackets; in the column for the treatment-control difference, standard errors are in parentheses. Means are regression-adjusted for randomization block fixed effects. Treatment-control differences and standard errors are based on a regression of the specified variable on a treatment dummy and randomization block dummies, accounting for sample weights and clustering at the teacher level. Statistical significance at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively.

**Table 3. Characteristics of TFA and Comparison Teachers (Percentages Unless Otherwise Indicated)**

| Characteristic | Teach For America Teachers | Comparison Teachers | Difference |
|---|---|---|---|
| **Demographic Characteristics** | | | |
| Age (average years) | 24.5 | 37.9 | -13.4***<br>(1.3) |
| Female | 60.9 | 79.4 | -18.4**<br>(8.0) |
| Black, non-Hispanic | 7.8 | 57.1 | -49.3***<br>(7.1) |
| Hispanic | 4.7 | 12.7 | -8.0<br>(5.0) |
| White, non-Hispanic | 89.1 | 30.2 | 58.9***<br>(7.0) |
| **Educational Background** | | | |
| Bachelors degree from selective college or university | 81.3 | 22.7 | 58.5***<br>(8.1) |
| Majored in math | 7.8 | 25.6 | -17.8**<br>(7.5) |
| Majored in secondary math education | 0.0 | 16.3 | -16.3***<br>(5.7) |
| Majored in other math-related subject | 26.6 | 11.6 | 14.9**<br>(7.4) |
| **Average Scores on Math Content Knowledge Test** | | | |
| Praxis II Mathematics Content Knowledge Test | 162.0 | 140.1 | 21.9**<br>(7.9) |
| Praxis II Middle School Mathematics Test | 179.8 | 158.3 | 21.6***<br>(3.7) |
| **Teaching Experience At End of Study Year** | | | |
| 1-2 years | 82.8 | 9.5 | 73.3***<br>(6.0) |
| 3-5 years | 17.2 | 20.6 | -3.4<br>(7.0) |
| More than 5 years | 0.0 | 69.8 | -69.8***<br>(5.8) |
| Average years | 1.9 | 10.1 | -8.3***<br>(0.9) |
| **Coursework During School Year** | | | |
| Took coursework during school year | 50.0 | 20.6 | 29.4***<br>(8.1) |
| Average hours of coursework during school year | 89.4 | 49.9 | 39.5*<br>(23.5) |
| Number of Teachers | 64 | 63 | |

Note: Standard errors in parentheses. Statistical significance at the 1, 5, and 10 percent levels is denoted by ***, **, and *, respectively. Selective colleges are those ranked by *Barron's Profiles of American Colleges 2003* as being very competitive, highly competitive, or most competitive. Other math-related subjects include statistics, engineering, computer science, finance, economics, physics, and astrophysics. We have scores on the Praxis II Mathematics Content Knowledge Test for 15 TFA teachers and 11 comparison teachers. We have scores on the Praxis II Middle School Mathematics Test for 45 TFA teachers and 40 comparison teachers.

**Table 4. Effects of TFA Math Teachers Relative to Comparison Teachers**

| | | Assuming Upper Bound for Compliance | | Assuming Lower Bound for Compliance | |
|---|---|---|---|---|---|
| | End-of-Year Math Score (Intent-to-Treat) | Fraction of Snapshots Enrolled with a TFA Teacher (First Stage) | End-of-Year Math Score (LATE) | Fraction of Snapshots Enrolled with a TFA Teacher (First Stage) | End-of-Year Math Score (LATE) |
| | (1) | (2) | (3) | (4) | (5) |
| Randomly Assigned to TFA Teacher (=1) | 0.07***<br>(0.02) | 0.96***<br>(0.01) | | 0.80***<br>(0.01) | |
| Fraction of Snapshots Enrolled with a TFA Teacher | | | 0.08***<br>(0.02) | | 0.09***<br>(0.02) |
| Control Group Mean | -0.60 | 0.02 | -0.59 | 0.10 | -0.57 |
| First-stage F-statistic | | 20868.8 | | 3346.8 | |
| Number of Blocks | 111 | 111 | 111 | 111 | 111 |
| Number of Teachers | 136 | 136 | 136 | 136 | 136 |
| Number of Students | 4,573 | 4,573 | 4,573 | 4,573 | 4,573 |

Note: Standard errors clustered by teacher are in parentheses. All regressions control for randomization block dummies, the variables listed in Table 2, a set of dummy variables indicating the number of years that elapsed between the baseline and outcome test score, and a set of dummy variables (one for each main covariate) indicating that a missing value of a given covariate has been replaced by a placeholder constant. Control group means listed in the 2SLS columns are control complier means calculated from the approach specified in Imbens and Rubin (1997). Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

**Table 5. Effects of TFA Math Teachers Relative to Comparison Teachers: Sensitivity Analyses**

| Model | Estimated Effect of TFA Teachers | Sample Sizes | | |
|---|---|---|---|---|
| | | Blocks | Teachers | Students |
| **Main Model** | 0.07*** (0.02) | 111 | 136 | 4,573 |
| **Alternative Estimation Approaches** | | | | |
| (1) No covariates except randomization block dummies | 0.07*** (0.02) | 111 | 136 | 4,573 |
| (2) No analysis weights | 0.07*** (0.02) | 111 | 136 | 4,573 |
| **Dropping Particular Randomization Blocks** | | | | |
| (3) Drop blocks in which percentage assigned nonrandomly in first month > 10 percent or percentage on final enrollment snapshot who had entered nonrandomly > 25 percent | 0.06** (0.02) | 78 | 112 | 3,434 |
| (4) Drop blocks with supplemental math classes | 0.11*** (0.03) | 58 | 72 | 2,460 |
| **Accounting for Selection Bias from Attrition** | | | | |
| (5) Lower bound for effect, based on Lee (2009) | 0.05*** (0.02) | 111 | 136 | 4,545 |
| (6) Upper bound for effect, based on Lee (2009) | 0.11*** (0.02) | 111 | 136 | 4,545 |

Note: Standard errors clustered by teacher are in parentheses. Each row of the table represents a different regression. Estimated effects are intent-to-treat effects. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

**Table 6. Effects of TFA Math Teachers Relative to Comparison Teachers within Teacher Subgroups**

| Type of Comparison | Estimated Effect of TFA Teachers | Sample Sizes | | |
|---|---|---|---|---|
| | | Blocks | Teachers | Students |
| (1) TFA teachers versus comparison teachers from traditional routes | 0.06** (0.03) | 58 | 82 | 2,477 |
| (2) TFA teachers versus comparison teachers from less selective alternative routes | 0.09*** (0.02) | 53 | 58 | 2,096 |
| (3) Inexperienced TFA teachers versus experienced comparison teachers [a] | 0.07** (0.03) | 66 | 85 | 2,815 |
| (4) TFA teachers versus comparison teachers within middle school grades | 0.06*** (0.02) | 83 | 103 | 3,373 |
| (5) TFA teachers versus comparison teachers within high school grades | 0.13*** (0.03) | 28 | 33 | 1,200 |

Note: Standard errors clustered by teacher are in parentheses. Each row of the table represents a different regression. Estimated effects are intent-to-treat effects. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

[a] Inexperienced teachers are defined as those in their first two years of teaching. Experienced teachers are defined as those with more than 5 years of teaching experience.

**Table 7. Associations between Teacher Characteristics and Teacher Effectiveness**

| Independent Variable | Dependent Variable: Student's End-of-Year Math Score (*z*-score) |
|---|---|
| Graduated from selective college or university (=1) | 0.028<br>(0.038) |
| Number of college-level math courses taken is above sample median (=1) | -0.019<br>(0.033) |
| Used college-level math in nonteaching job (=1) | -0.054<br>(0.044) |
| Score on Praxis II Test in Math Content Knowledge (*z*-score) | 0.018<br>(0.037) |
| Score on Praxis II Test in Middle School Math (z-score) | 0.023<br>(0.018) |
| Number of hours of math pedagogy instruction during training is above sample median (=1) | -0.025<br>(0.034) |
| Number of days of student teaching in math during training is above sample median (=1) | -0.009<br>(0.033) |
| Hours of education-related coursework during the school year (divided by 10) | -0.003***<br>(0.001) |
| Has more than one year of teaching experience (=1) | 0.142***<br>(0.041) |
| Number of additional years of teaching experience beyond two total years (until teacher has five total years of experience) | -0.030<br>(0.021) |
| Number of additional years of teaching experience beyond five total years | -0.001<br>(0.004) |
| Number of Blocks | 111 |
| Number of Teachers | 136 |
| Number of Students | 4,573 |

Note: Standard errors clustered by teacher are in parentheses. Estimates come from a single regression that also controls for a treatment dummy, randomization block dummies, the variables listed in Table 2, a set of dummy variables indicating the number of years that elapsed between the baseline and outcome test score, and a set of dummy variables (one for each student-level covariate) indicating that a missing value of a given student-level covariate has been replaced by a placeholder constant. Missing values of the teacher-level variables shown in the table are accounted for with multiple imputation. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

**Table 8. Extent to which Observed Teacher Characteristics Account for the Impact of TFA Teachers**

| Teacher Characteristic | (1) Difference in Characteristic Between TFA and Comparison Teachers | (2) Association Between Characteristic and Teacher Effectiveness (student *z*-score units) | (3) Predicted Difference in Effectiveness Between TFA Teachers and Comparison Teachers (student *z*-score units) |
|---|---|---|---|
| Hours of education-related coursework during the school year (divided by 10) | 2.64<br>(1.86) | -0.003***<br>(0.001) | -0.01 |
| Has more than one year of teaching experience (=1) | -0.33***<br>(0.05) | 0.142***<br>(0.041) | -0.05 |

Note: In columns 1 and 2, standard errors clustered by teacher are in parentheses. In column 1, differences between TFA and comparison teachers are estimated from a student-level regression in which the indicated teacher characteristic is regressed on a treatment dummy, randomization block dummies, the variables listed in Table 2, a set of dummy variables indicating the number of years that elapsed between the baseline and outcome test score, and a set of dummy variables (one for each covariate) indicating that a missing value of a given covariate has been replaced by a placeholder constant. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

**Table A.1. Summary Statistics of Teacher Characteristics Examined in Accounting for the Impact of TFA Teachers**

| Teacher Characteristic | Mean | Standard Deviation |
|---|---|---|
| **Measure of General Academic Ability** | | |
| Graduated from selective college or university (=1) [a] | 0.56 | 0.50 |
| **Measures of Exposure to and Knowledge of Math** | | |
| Number of college-level math courses taken is above sample median (=1) [b] | 0.42 | 0.49 |
| Used college-level math in nonteaching job (=1) | 0.23 | 0.42 |
| Score on Praxis II Test in Math Content Knowledge (*z*-score) | 0.00 | 1.00 |
| Score on Praxis II Test in Middle School Math (*z*-score) | 0.00 | 1.00 |
| **Measures of Instructional Training** | | |
| Number of hours of math pedagogy instruction during training is above sample median (=1) [c] | 0.41 | 0.49 |
| Number of days of student teaching in math during training is above sample median (=1) [d] | 0.32 | 0.47 |
| Hours of education-related coursework during the school year (divided by 10) | 6.35 | 12.67 |
| **Measures of Teaching Experience** | | |
| Has more than one year of teaching experience (=1) | 0.79 | 0.41 |
| Number of additional years of teaching experience beyond two total years (until teacher has five total years of experience) | 1.40 | 1.37 |
| Number of additional years of teaching experience beyond five total years | 3.63 | 6.66 |
| Number of Blocks | 111 | |
| Number of Teachers | 136 | |
| Number of Students | 4,573 | |

Note: Summary statistics are calculated from the student-level analysis sample.

[a]Selective colleges are those ranked by *Barron's Profiles of American Colleges* as very competitive, highly competitive, or most competitive.

[b]Teacher at the median took 5 college-level math courses.

[c]Teacher at the median had 21 to 40 hours of math pedagogy instruction.

[d]Teacher at the median had 16 to 20 days of student teaching.