Alternative Routes to Teaching: The Impacts of Teach for America on Student Achievement and Other Outcomes

Steven Glazerman
Daniel Mayer
Paul Decker

Abstract

This paper reports on a randomized experiment to study the impact of an alternative teacher preparation program, Teach for America (TFA), on student achievement and other outcomes. We found that TFA teachers had a positive impact on math achievement and no impact on reading achievement. The size of the impact on math scores was about 15 percent of a standard deviation, equivalent to about one month of instruction. The general conclusions did not differ substantially for subgroups of teachers, including novice teachers, or for subgroups of students. We found no impacts on other student outcomes such as attendance, promotion, or disciplinary incidents, but TFA teachers were more likely to report problems with student behavior than were their peers. The findings contradict claims that such programs allowing teachers to bypass the traditional route to the classroom harm students. © *2006 by the Association for Public Policy Analysis and Management*

INTRODUCTION

Public school teaching is a field with long-standing barriers to entry, particularly the requirement that teachers become certified by the state. Teacher certification can involve taking a minimum number of courses, receiving a credential from a teacher education program, conducting a minimum number of hours of student-teaching, passing an exam, or some combination of these requirements. Over the years, however, as teacher shortages have developed in particular areas and subjects, there has been continued interest in finding alternative routes into the classroom. The goal of alternative teacher preparation programs is to lower the barriers to entry and tap previously untapped pools of potential teachers, particularly for hard-to-fill teaching positions. This paper examines the impacts made by participants in one of the nation's most prominent alternative preparation programs, Teach for America (TFA).

TFA was founded in 1989 to address the educational inequities facing children in low-income communities across the United States by expanding the pool of teacher candidates available to schools in those communities. TFA recruits seniors and recent graduates of the nation's top colleges to find those who are willing to commit at least two years to teaching years in impoverished communities. The program focuses on graduates with strong academic records and leadership capabilities, regardless of whether they have prior exposure to teaching practice. TFA is particularly interested in candidates who can be effective teachers but, in the absence of

Manuscript received December 2004; review completed March 2005; revised manuscript submitted April 2005; revision review complete April 2005; revised manuscript submitted April 2005; accepted April 2005.

Journal of Policy Analysis and Management, Vol. 25, No. 1, 75–96 (2006)
© 2006 by the Association for Public Policy Analysis and Management
Published by Wiley Periodicals, Inc. Published online in Wiley InterScience (www.interscience.wiley.com)
DOI: 10.1002/pam.20157

TFA, would not otherwise consider a teaching career. Consequently, most TFA recruits have not had education-related majors in college and therefore have not followed the traditional route to teaching. However, TFA recruits stand out as high academic achievers. For example, the new TFA participants ("corps members") in 2003 had an average SAT score of 1,310 and a mean grade point average of 3.5.

TFA provides five weeks of intensive training, including coursework and student teaching, over the summer prior to the corps member's initial placement in the classroom. TFA reports that corps members spend about 70 hours per week on training-related activities during this period. As intensive as the summer training may be, it stands in contrast to the one to four years of coursework (with student teaching) that is typical of traditional teacher education programs. In addition to the TFA summer training, corps members take part in a one- to two-week, TFA-led induction in their assigned region. They also receive ongoing support from TFA staff and faculty located in each community, as well as from TFA national staff.

TFA has been highly successful in attracting people who meet its standards and the corps has expanded rapidly in recent years. Between 2000 and 2003, the TFA applicant pool grew almost fourfold (from 4,068 to 15,706) and the number of new corps members nearly doubled (from 868 to 1,656). Since the program began, more than 10,000 TFA corps members have taught more than 1.5 million students. In 2004, the program placed corps members in 22 urban and rural regions, an increase from 15 regions served in 2000.

Despite TFA's rapid expansion, there is little evidence whether teachers with strong academic backgrounds, but limited exposure to teaching practice, can be effective. Some critics argue that programs such as TFA are "loopholes" that permit unlicensed and under-trained teachers into the classroom simply as a way to address teacher shortages. Darling-Hammond (1994, 1996) has argued that TFA teachers "often have difficulty with curriculum development, pedagogical content knowledge, students' different learning styles, classroom management, and student motivation." Other researchers are more optimistic about the potential benefits of hiring teachers through programs such as TFA. Ballou and Podgursky (1998) argue that there is no evidence that formal teacher certification produces more qualified teachers, and that certification policies may discourage talented individuals from entering the profession. Two recent studies (Raymond, Fletcher, & Luque, 2001; and Laczko-Kerr & Berliner, 2002) attempted to assess the impact of TFA using non-experimental methods on samples drawn from single regions (Houston, TX, and an unnamed district in Arizona, respectively), and reached different conclusions. In both cases the impact of TFA was identified, in part, by making comparisons across teachers in different schools, with different students and other working conditions. The current study goes beyond the previous studies by using an experimental design; for example, by randomly assigning subjects to treatment conditions within grades within schools, and by working with a sample spread over several regions of the country.

This study examines the impact of TFA teachers on the students in their classrooms, compared with what would have happened in the absence of the TFA teachers. In the next section, we describe in detail our approach to estimating this counterfactual outcome. We estimate the "full" impact of TFA teachers on their students, which encompasses both a recruitment effect (for example, impacts due to changes to the type of person who chooses to teach in high poverty areas) and a productivity effect (for instance, impacts due to the effectiveness of training provided to these teachers relative to the training received by non-TFA teachers). Because both of these components are integral to the TFA program, the study was not designed to disentangle their separate influences on student outcomes.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

STUDY DESIGN AND DATA

This study addresses the question: Do TFA teachers improve (or at least, not harm) student outcomes relative to what would have happened in their absence? To measure the impact of TFA teachers on students, it would be ideal to compare the experience of students assigned to TFA teachers with the same students' experiences in the absence of TFA. Since this counterfactual cannot be directly observed, we approximated it by using a control group of non-TFA teachers teaching similar students in the same environment. Specifically, our strategy was to compare outcomes of students taught by TFA teachers with outcomes of students taught by non-TFA, or control, teachers *in the same schools and at the same grades*. We use this group of non-TFA teachers to represent the teachers that would have filled the open teachers slots in the absence of the TFA teachers.

In order to ensure that the TFA and control teachers have essentially identical classes of students, we secured permission from principals to randomly assign students to classrooms.[1] Without random assignment, school principals might have given the most challenging students to specific teachers, making classroom comparisons more reflective of student differences than of teacher performance.

For the analysis, "control teachers" included all teachers in the study who were not TFA corps members at the time of the study or at any time in the past. Therefore, the group included traditionally certified, alternatively certified, and uncertified teachers—any active teacher who came from any source other than TFA. "TFA teachers" included all teachers who entered the profession through TFA—both current TFA corps members in their first two years of teaching, and alumni (former corps members) who were still teaching.

We conducted two types of comparisons of TFA and control teachers. First, we compared classes taught by TFA teachers with classes taught by all control teachers, which could include both novice and veteran teachers. In this case, the average years of teaching experience was far higher for the control teachers than for the TFA teachers. To control for differences in teaching experience, we conducted a second type of comparison based on classes taught by novice TFA teachers with novice control teachers. "Novice teachers" included teachers in one of their first three years of teaching during the study year.

Which of these comparisons is more relevant is a matter of some debate. In the absence of TFA, the students in our sample would have been taught by a mix of novices and veterans found in their schools. However, one might assume that if a TFA teacher were not hired, then some other, presumably novice, non-TFA teacher would be hired in his or her place. Rather than try to identify which comparison is more relevant from an empirical perspective, we simply examined both the "all teachers" and "novice only" comparisons. Alternatively, one could argue that a third relevant counterfactual includes teachers who are new to the school, which includes the novice only group plus the veteran transfers. We conducted this comparison of newly hired controls along with other teacher subgroup comparisons in a sensitivity analysis.

[1] Securing participation in the study, including random assignment of students to classrooms, required negotiation with principals and resulted in a small number of schools either refusing to participate or refusing to complete the eligibility screening process. The number of eligible refusals is therefore unknown, but we estimate it to be less than five schools. We have no reason to believe that the relative effectiveness of TFA teachers in the schools in our sample is unrepresentative. Participation in the study and cooperation with random assignment were not requirements for schools to have TFA candidates placed there.

Before the start of the academic year, we made random class assignments for all students entering the targeted grades (grades 1 through 5, chosen because the classrooms are self-contained). Throughout the year we conducted roster checks to monitor and enforce the original assignments and we followed students who left the school, as long as they were still in the district. Because student and teacher mobility might be related to treatment status, we followed an "intent-to-treat" approach and based all test score comparisons on students' initial random assignment, even if they changed classrooms, left their school, or their teacher left. By monitoring random assignment, we verified that treatment-control crossover rates were low (less than 3 percent) and about the same in both directions (TFA to non-TFA and vice versa) and overall student attrition from the study was low (11 percent).[2] Two TFA teachers left during the observation period and were replaced by non-TFA teachers, but we estimated the impacts both with and without the affected schools and found it made little difference. We also conducted sensitivity tests to verify that neither student nor teacher mobility strongly influenced our findings.

The study included 6 regions—Baltimore, Chicago, Compton (Los Angeles), Houston, New Orleans, and the Mississippi Delta—of the 15 regions where TFA placed teachers at the time the study was designed.[3] The regions were selected after stratifying on urbanicity and student race. We excluded one region that had a policy of rearranging classrooms of students for reading instruction and we excluded one region whose district officials refused to consider participating because a central office reorganization was underway. Within each region, we randomly selected schools from those that had the staffing needed to support our design. For each eligible school, we selected every grade at which there was at least one TFA and one control teacher. Such grades-within-schools, which are the units at which random assignment was conducted, are referred to in this paper as "blocks." The final research sample, which is summarized in Table 1, consisted of 17 schools with 37 blocks, 100 classrooms, and nearly 1,800 students.

Table 1. Study sample.

Region	Number of Schools	Number of Comparison Blocks	Number of Classes Taught by: TFA Teacher	Novice Control Teacher	Veteran Control Teacher	Number of Students Taught by:[a] TFA Teacher	Novice Control Teacher	Veteran Control Teacher
Baltimore	3	6	7	1	8	137	18	147
Chicago	3	7	7	2	5	139	42	105
Houston	3	7	7	3	7	126	56	114
Compton	2	6	6	6	4	97	111	72
Mississippi Delta	3	6	12	2	10	201	31	146
New Orleans	3	5	5	1	7	85	21	117
Total	17	37	44	15	41	785	279	701

Source: TFA Evaluation Project tracking system.
[a] Includes students in the research sample who completed the spring achievement test.

[2] The rate at which students left their schools was higher than 11 percent, but we followed movers who remained in the district and assessed them in the spring.
[3] We distinguish between regions and school districts. In some cases, mostly in the large urban regions, TFA works with a single district in a region. In other cases, particularly in the rural regions, TFA works with multiple districts in a region.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The main source of data for this study was a set of achievement tests we administered in the fall (pre-test) and the spring (post-test) of the study year. We used an abbreviated form of mathematics and reading subtests of the Iowa Test of Basic Skills (ITBS). The response rate to the spring testing was 89 percent. We also collected data from school records and administered a survey of teachers at the end of the school year. The teacher survey measured personal characteristics, preparation for teaching, teaching experience, career expectations, professional development, teaching practices and beliefs, receipt of help in the classroom, and student behavior. Teachers in 98 of the 100 classrooms in the study completed a survey. Of those, 41 were TFA teachers and 57 were control teachers (18 novice teachers and 39 veterans).[4]

SAMPLE CHARACTERISTICS

Teacher Characteristics

TFA Teachers

The TFA teachers in the study sample resembled TFA teachers nationally. Using data from the TFA program, we compared the race/ethnicity, gender, and competitiveness of the college attended by the TFA corps members who participated in 2000 to 2002 (see Decker, Mayer, & Glazerman, 2004), and had the characteristics one would expect of a participant in the program. Just before entering TFA, they tended to have graduated from highly competitive colleges (70 percent) as judged by a prominent rater of college selectivity (Barron's Profiles of American Colleges, 2003). They had little or no formal training in pedagogy or student-teaching before they entered the program. Demographically they were diverse, but still more likely to be white (67 percent) than the students they taught or the other teachers in the schools where they taught and had a higher percentage of male teachers (31) than their non-TFA counterparts (13).

While the TFA teachers had little training before entering the program, they all went through TFA's summer institute, which included four weeks of student teaching. Many TFA teachers went on to get a regular teaching certificate (51 percent) and/or a master's degree in education (25 percent) at the same time they were teaching, typically to meet district or state requirements.

Control Teachers

When interpreting the impact estimates presented in this study, it is critical to understand what kinds of teachers represent the counterfactual to which TFA teachers are being compared. Control teachers varied widely in terms of training, experience, and credentials. They included both certified and uncertified teachers, teachers with and without degrees in education, and teachers with varying amounts of pre-service student teaching. Just over half of the control teachers had a degree in education (compared to 25 percent of TFA teachers): almost all were bachelor's degrees. In contrast, all of the TFA teachers who had studied education held master's degrees.

[4] Eight classrooms experienced turnover of teachers during the school year, so the numbers of novice controls, veteran controls, and TFA teachers who completed our spring questionnaire differed slightly from the numbers of those who began the school year.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

One might expect no TFAs to be certified and all non-TFAs to be certified, but that was not the case. Novice controls in particular had a low rate of certification, 38 percent, which was even lower than the percentage of TFA teachers who were certified by the time we conducted the survey (51 percent). Just over two-thirds of all control teachers (67 percent) were certified. When we controlled for school district (which at the time was a major factor in teachers' ability to remain uncertified) the differences disappeared, but these data underscore that our overall comparisons do not describe certified versus uncertified teachers.

Control teachers were less variable, and stood in starker contrast to TFA teachers, in terms of race/ethnicity and competitiveness of their undergraduate institutions. Three-quarters of control teachers were African American. Less than 3 percent had attended very competitive colleges. Almost 87 percent were female. Also, control teachers tended to be older, even controlling for teaching experience. This reflects the fact that TFA teachers typically entered teaching straight from college with a median age of 22, whereas the control teachers tended to finish college later (median age 24) and take longer after college to enter teaching (median age 27).

Student Characteristics

TFA works with schools that serve a disadvantaged, largely minority population. Nearly all students in our sample are from low-income families, and, on average, they are also low academic performers. Over 95 percent of the students in our sample were certified for free or reduced-price school lunch, compared with only about 41 percent of students nationwide. Two-thirds were African American and more than a quarter of the students were Hispanic. In addition, many of the students—more than a fifth—were overage for their grade.[5]

In academic performance, the typical student in our study starts off the year achieving far below the level of children in the same grade nationally. The average score for our sample members was 27 Normal Curve Equivalent (NCE) points in mathematics and 26 in reading. The NCE scale has a mean of 50 and standard deviation of 21 in a nationally representative norm group of children in the same grade. Ranked against students in this national norm group, these NCE scores would place our average sample member in the 14th percentile in math and the 13th percentile in reading.

In terms of observable characteristics, the baseline data suggest that random assignment produced equivalent groups of students. We compared the students in TFA and control classrooms in terms of demographic characteristics—such as gender, race/ethnicity, age-for-grade, income—and tests scores and class characteristics such as class size. All of the treatment-control differences are small and none was statistically significant.

Not only were the baseline characteristics of the two groups equivalent, the mobility patterns were similar as well (Table 2). About 87 percent of the sample members stayed in the same classroom all year.[6] The percentages of students who switched classrooms from TFA to control or vice versa (crossovers), who moved

[5] Students were considered overage for grade if they were older than the most common age for that grade (7 years old for first graders, 8 years old for second graders, and so on) before September 1. For example, if a fourth-grade student turned 9 years old in August just before the school year began, she was overage. If she turned 9 years old in September, she was not overage.

[6] This figure includes a small number of students who transferred between classrooms of the same treatment status, such as control to control or TFA to TFA. Such transfers have a negligible effect on inferences about the impact of TFA and are treated as stayers in our analysis (since they "stay" with their original classification as a TFA or control student).

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Mobility rates of control and TFA students (percentages).

Mobility Type	Control Students	TFA Students	Difference[a]	Total
Stayer	87.8	86.0	–1.8	87.3
Crossover[b]	3.7	4.3	0.7	4.0
Mover within district	5.2	5.6	0.4	5.4
Mover out of district	2.3	2.9	0.6	2.5
Mover other/unknown	1.3	1.3	0.0	1.3
Sample size	1,094	875		1,969

Source: Student tracking system.
[a] Chi-squared test fails to reject the null hypothesis of equal distributions (p = 0.898); that is, the differences between TFA and control students are not statistically significant.
[b] "Crossover" refers to students who switched from a TFA classroom to a control classroom, or vice versa.

within the district, who moved out of the district, or who transferred out but could not be located, were about the same for students in both TFA and control classrooms. The differences between the two groups were not statistically significant.

ESTIMATION OF IMPACTS

Our procedure for estimating the impacts of TFA on student outcomes takes into account the nested structure of the data. In particular, it recognizes that students were randomly assigned to classrooms *within grades* within schools, that is, within blocks. Each block can be thought of as a mini-experiment. Therefore, the national study consists of 37 mini-experiments (composed of 100 classrooms). We describe the estimation in terms of a student-level equation nested within a block-level equation.

Student Level Model

The student-level model is a "post-test-on-pretest" regression, with dummy variables for each block (Equation 1):

$$Y_{ijk}^{spring} = \theta Y_{ijk}^{fall} + \beta' X_{ijk} + \alpha_k I_{ijk} + \delta_k I_{ijk} * TFA_{jk} + \varepsilon_{ijk} \quad (1)$$

where:
i indexes students; j indexes classrooms; k indexes blocks
Y = test score or other outcome
X = vector of student characteristics
I = block dummy indicator
TFA = treatment status dummy indicator
θ, α, β, δ = parameters to be estimated
ε = independent and identically distributed (iid) random error term

The vector X includes student-level control variables such as indicators for free lunch eligibility, Hispanic origin, and gender. These control variables play only a minor role, since schools tend to be homogeneous within blocks. The block dummies are interacted with TFA status, and the coefficients on those interaction terms represent the achievement growth for the TFA teacher(s) in each block, relative to the non-TFA teacher(s) in that block. The average of these 37 TFA impacts is the average impact of policy interest.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

While each block can include more than one TFA classroom or more than one control classroom—a typical configuration had one TFA and two control classrooms—the results presented here are based on a model that did not include a separate classroom level. With only one or two classrooms per treatment condition per block, allowing the block dummies to represent the average classroom fixed-effect accounted for most of the clustering of students and produced nearly the same result.

We estimated Equation 1 using linear regression with sample weights that correct for two features of the design. One is sample attrition. The analysis sample excluded 11 percent of the cases for which there was no spring test (usually because the student moved out of the district); therefore, we created non-response adjustment weights that are higher for students with background characteristics and pretest scores that are more similar to the leavers, allowing them to effectively stand in for their missing counterparts.

The other design feature is that the probability of assignment to TFA or control classrooms differs across blocks, mainly because the number of treatment or control teachers per grade was not uniform. This induces an arbitrary treatment-block correlation, but one that can easily be undone by giving each sample member a normalization weight proportional to the inverse of the sample size within treatment condition within block. Our sample weight combines the non-response adjustment with the normalization weight.

From Equation 1 we computed 37 block-specific impact estimates and a corresponding variance-covariance matrix that characterizes the estimation error associated with those estimates. Both are used in the block-level model to estimate the average impact overall.

Block Level Model

The simplest approach to estimating the overall impact of TFA is to average the 37 block-impact estimates (the unconditional mean). However, we estimated the average impact conditional on some control variables as described below. First, we note that the 37 block-level impacts are measured with estimation error, as shown in Equation 2:

$$\hat{\delta}_k = \delta_k + \omega_k \tag{2}$$

Using the coefficients on the treatment-block dummy indicators as regressors in the block-level model, we derive the following expression:

$$\hat{\delta}_k = \mu + \lambda' W_k + \{\eta_k + \omega_k\} \tag{3}$$

where W is the vector of block-specific variables and the composite error term captures both the estimation error from the student-level model and the sampling error from the block-level model. The block-specific variables contained in W can include grade level, school, or district. As a test of the robustness of the model, we included school dummy variables to address the problem of blocks from the same school having a common unmeasured component, such as the influence of a principal that could bias the estimates of impacts and standard errors. This effect was small, because most schools have only two blocks and we found the differences between blocks in the same school were just as large as differences between blocks from different schools.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

To estimate Equation 3 in the presence of a composite error term, we used the method of Hanushek (1974) to compute the error variance matrix for the WLS estimates.

The model represented by Equations 1 and 3 can be estimated for the full sample and for subgroups. We conducted analysis of subgroups defined by both student characteristics and teacher characteristics. For student subgroups the sample size in Equation 1 was smaller, but in most cases the number of blocks available for analysis in Equation 3 was unaffected. For some subgroups that were unevenly distributed, all the blocks might not have been represented in Equation 3. For teacher subgroups, we first dropped classrooms that were not members of the subgroup, and then we dropped blocks in which there did not remain at least one TFA and one control teacher.

IMPACTS ON STUDENT ACHIEVEMENT

The most important question this study addressed is whether students taught by TFA teachers performed at least as well on achievement tests as students taught by other teachers—in other words, whether the impacts on student achievement were non-negative—and we found that they were.

Positive Impact on Math, Zero Impact on Reading

Unadjusted Mean Differences

We first examined the differences in mean test scores, unadjusted for covariates, and found that students in TFA classrooms outperformed control students in mathematics (Figure 1). The figure shows the math percentile ranking of the average student in TFA and control classrooms in the fall and again in the spring, at the end of the school year.[7] The average control class students scored in the 15th percentile in the fall and remained in the 15th percentile at the end of the year. That is, control class students experienced typical achievement growth, shown in Figure 1 by the light-gray line.[8] In contrast, the average TFA class students increased their ranking from the 14th percentile to the 17th percentile over the same period. The difference in growth rates is statistically significant.

For reading achievement, we found that the average student in TFA and control classrooms experienced the same growth rate. The average sample member increased by the equivalent of about 1 percentile point during the study year. The nearly parallel lines in Figure 2 demonstrate the similarity in these growth rates. The initial treatment-control difference of a single percentile point is not statistically significant.

Regression-Adjusted Estimates

The findings shown in Figures 1 and 2 are not regression-adjusted, but they are confirmed when subjected to formal modeling and hypothesis testing. The regression-adjusted estimates, which give a better picture of the size and statistical signifi-

[7] All calculations in this report used normal curve equivalent (NCE) scores, which are translated into percentile rankings for ease of interpretation.

[8] A flat line is a sign of normal growth because all rankings are expressed relative to a nationally representative norm group, which also experienced fall-to-spring achievement growth.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

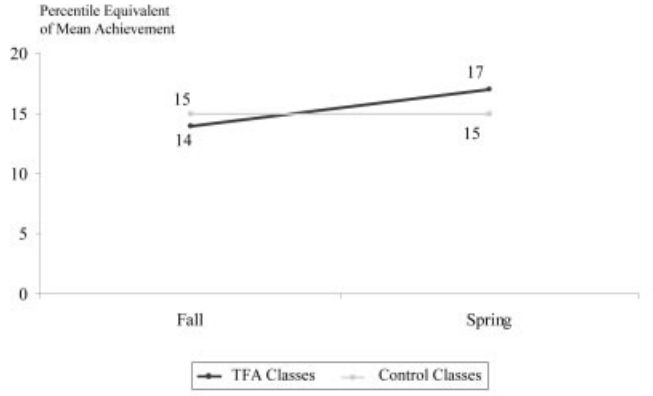

Figure 1. Fall-to-spring change in achievement: Math.

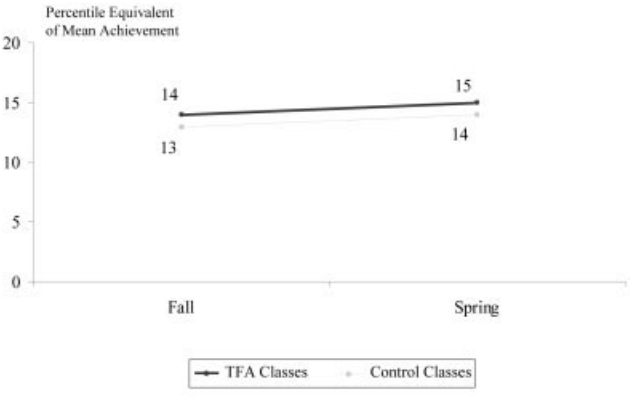

Figure 2. Fall-to-spring change in achievement: Reading.

cance of these findings, are shown in Table 3, along with the other regression coefficient estimates of the base case model specification. We report all impact estimates in terms of NCEs, a scale normed to a nationally representative population with a mean of 50 and standard deviation of 21.06. Using this metric, the impact on math achievement is 2.4 NCEs, which is significantly different from zero. This corresponds to a standardized effect size of 0.15; that is, 15 percent of a standard deviation.[9]

The positive impact of TFA on math scores is statistically significant, but is it large enough to imply that TFA teachers produce meaningfully greater math achievement? When expressed in grade equivalents, the math achievement advantage TFA teachers offered appears to be meaningful. The impact translates into about 10 percent of a grade equivalent, suggesting that the advantage to TFA stu-

[9] The standard deviations used in effect size calculations are 15.9 for math and 17.1 for reading. These are the standard deviations estimated from the control group's fall (baseline) test.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3. Main test score regression estimates.

	Mathematics		Reading	
	Coefficient Estimate	Standard Error	Coefficient Estimate	Standard Error
Student Level Model				
Explanatory Variables				
Baseline Math Score	0.52 ***	(0.02)	0.23 ***	(0.02)
Baseline Reading Score	0.16 ***	(0.02)	0.45 ***	(0.02)
Female	-1.59 ***	(0.61)	0.11	(0.58)
Over Age for Grade	-1.79 **	(0.85)	-3.53 ***	(0.81)
Free Lunch	2.67	(2.47)	-1.38	(2.36)
African American	4.04 **	(1.99)	4.50 **	(1.90)
Hispanic	4.97 **	(2.21)	3.61 *	(2.11)
Block Dummies	Included		Included	
Block*TFA Dummies	Included		Included	
Sample Size (Students)	1,692		1,710	
Adjusted R-squared	0.869		0.869	
Block Level Model				
Explanatory Variables				
Constant (TFA Effect)	2.43 ***	(0.73)	0.56	(0.62)
DiffPercent Nonresearch	-0.12	(0.09)	-0.19 **	(0.07)
Sample Size (Blocks)	37		37	
Adjusted R-squared	0.028		0.108	

Source: Scores from the Iowa Test of Basic Skills.

Note: All test scores are expressed in NCEs, whose average score nationally is 50 and standard deviation is 21.06.

*Significantly different from zero at the .10 level, two-tailed test.
**Significantly different from zero at the .05 level, two-tailed test.
***Significantly different from zero at the .01 level, two-tailed test.

dents corresponds roughly to an additional month of instruction. Comparisons with other evaluation findings also suggest that the TFA impacts on math achievement are meaningful. An often-cited benchmark for assessing impacts on education performance is the effect of reducing elementary school class size, from an average of 23, to 15 students, which has been reported to have a single-year effect size of about 0.23, based on a large-scale experimental study in Tennessee (Finn & Achilles, 1999). Therefore, when compared with the effect of reduction in class size, the magnitude of the TFA impact on math scores—an effect size of 0.15—is about 65 percent of the effect of a reduction in class size of eight students.

The estimated impact on reading scores, also shown in Table 3, was very close to zero and was not statistically significant. The point estimate of 0.56 NCEs corresponds to an effect size of 0.03.

In both the mathematics and reading score regressions, the coefficients on the student-level explanatory variables are presented as well. The coefficients on the pre-test measures are positive, as one would expect. The coefficients on demographic variables such as income (proxied by a dummy for eligibility for free or reduced price lunch) and race reflect the relationship of those characteristics to test scores *conditional on pre-test*. When the pre-test measures are excluded, many of the demographic variable coefficients change signs. Regardless, the estimated TFA effect is hardly changed by inclusion of such variables, because schools (and therefore blocks) tend to be homogeneous in terms of race, ethnicity, and income. The block-level model includes a variable that measures the difference in the block

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

between the TFA classrooms and control classrooms in the percentage of students who were "non-research." Non-research students were primarily those who transferred into the study schools during the school year, after random assignment, and a small number of students (about 3 percent of the original total) whose parents opted them out of the study or were assigned to a specific teacher before random assignment. These students were not included in our sample, but their presence could have affected the classroom environment that the research students were exposed to.

Impacts Similar for Different Types of Teachers

To address variations on the study's main research question, we estimated impacts for various subgroups of teachers. We focused on the subgroup defined according to teacher experience. Some might argue that a fair test of TFA would be to compare its teachers against a group of non-TFA teachers who began teaching around the same time, not against a mixed group that includes some 20- and 30-year veterans, as the full sample does. To examine this "novice-only" comparison, we estimated the impacts for TFA and control teachers with three or fewer years of experience. Based on this comparison, we found the impact of TFA on math scores was 4.1 NCEs (Table 4). This corresponds to an effect size of 0.26. The impact estimate for novice teachers, however, was sensitive to model specification, because the novice comparison is based on only 11 comparison blocks (25 classrooms), about one-quarter of the full sample. By including or excluding different control variables, the impact estimate (not shown) ranged from 3.0 to 6.2.

The impact of TFA on reading for the novice-only comparison followed a similar pattern, although the estimated impact was closer to zero. Using the main regression model, the reading impact of TFA for novice teachers was 1.1 NCEs, which was not statistically significant. Under alternative regression models, the estimate was about 0.7 NCEs, on average.[10]

We also estimated the impact for first-year TFA teachers, only to see if there was a difference from an extra year on the job for these teachers. We found the impact for first-year TFA teachers in math was lower than the full-sample impact—1.8 compared to 2.4. The precision of this new estimate is also very low, because of the smaller number of first-year teachers from which to generalize (standard error is 1.7 NCEs), so this impact is not significantly different from zero. Not surprisingly, the impact on math scores at 2.5 NCEs was slightly higher for second-year TFAs than it was for the full sample. (In addition to having one more year of experience in the classroom, many second-year TFAs had earned a master's degree in education.)

To examine the role of certification, we repeated the exercise, this time comparing all TFA teachers with certified teachers only. We found the math impact was 1.9 NCEs. It was 3.1 when we compared TFA teachers to their uncertified counterparts. These findings imply that certified teachers outperformed uncertified ones, on average. However, the difference in impacts for the two groups is small, and the standard errors are large, so these different estimates between the subgroup of certified teachers and the full sample are as likely due to chance as they are to a real certification effect. Our data cannot distinguish between the two.

[10] Samples of students and comparison blocks used for different teacher subgroup analyses are not mutually exclusive. For example, students in the TFA classrooms that had both novice and veteran control teachers in the same grade were included in both the novice comparison and the veteran comparison.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4. Impacts on test scores, teacher subgroups (NCEs).

	Mathematics		Reading		Sample Size		
Subgroup Comparison	Impact Estimate	Standard Error	Impact Estimate	Standard Error	Blocks	Classes	Students
Full sample	2.43***	(0.73)	0.56	(0.62)	37	100	1,715
Experience							
Novice TFAs versus novice controls	4.13***	(1.24)	1.06	(1.19)	11	25	432
All TFAs versus veteran controls	2.71***	(0.97)	0.45	(0.70)	31	79	1,370
First-year TFAs versus all controls	1.81	(1.70)	-0.90	(0.99)	12	32	526
Second-year and veteran TFAs versus all controls[b]	2.55***	(0.74)	1.09	(0.71)	29	77	1,320
All TFAs versus recently hired controls	2.46***	(0.79)	1.12	(0.75)	24	57	994
Certification							
All TFAs versus certified controls	1.92*	(0.94)	0.01	(0.75)	27	70	1,216
All TFAs versus uncertified controls	3.12**	(1.11)	1.01	(0.95)	14	36	620
Uncertified TFAs versus all controls	3.21***	(0.98)	-0.34	(0.92)	19	58	973

Source: Scores from the Iowa Test of Basic Skills, administered by Mathematica Policy Research, Inc.
Note: All test scores are expressed in NCEs, whose average score nationally is 50 and standard deviation is 21.06.
[a] Control group means and impacts are regression-adjusted. The regression model controls for baseline test scores, gender, race/ethnicity, eligibility for free or reduced-price lunch, age (whether overage for grade), and percentage of students in the classroom who were not in the research sample.
[b] Some TFA teachers continue to teach in the same schools beyond their two-year commitment. In our sample, there were five TFA teachers in their third year, one in their fifth year, and one in their sixth year. Of the rest, 15 were in their first year of teaching, and 22 were in their second year.
* Significantly different from zero at the .10 level, two-tailed test.
** Significantly different from zero at the .05 level, two-tailed test.
*** Significantly different from zero at the .01 level, two-tailed test.

For all the teacher subgroup comparisons above, the impact on reading changes in a similar pattern, but the impacts were consistently small—no more than 1.2 NCE points in absolute value—and were not statistically significant.

Impacts Similar for Different Types of Students

We examined the hypothesis that the TFA teachers might have more success with some types of students than with others and found little evidence to support such a claim. Instead, the impact of TFA appeared to be generally consistent across a broad spectrum of subgroups. A key constraint in testing the hypothesis is that the study was not designed specifically for subgroup analysis. As one begins to look at subgroup impacts, it becomes difficult to distinguish true differences from chance differences, because the size of subgroups is often small. For most subgroups of students, however, the pattern of impacts was similar to that of the full sample.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The impacts of TFA were similar across boys and girls and across racial/ethnic groups. Table 5 shows that the impacts on math scores were positive and significant for both boys and girls, and the impacts on reading were not significantly different from zero for both groups. The impacts on math scores for the race/ethnic groups that were large enough to estimate separate impacts—African American and Hispanic students—were 2.1 and 1.9 NCE points, respectively.[11] The estimate of the impact for African American students was highly sensitive to inclusion of a few comparison blocks that consisted of just two or three students. These were classrooms in largely Hispanic schools. Removing the outlier blocks resulted in an impact on African American students' math scores of 2.4 to 2.5 NCEs, depending on the threshold used. The corresponding impacts on reading scores differed. The impact on African American students' reading scores was less than one point. For Hispanic students, it was more than two points (although not statistically significant). It is interesting that the impact on African American students' reading scores in regions with a majority of Hispanic students was also high, more than 3.4 NCE points (not presented in the table). These findings—that the impacts on the reading scores in regions with a majority of Hispanic students were higher than the full sample—are based on only 10 comparison blocks, but could provide a lead for further research.

Estimates of separate impacts by age-within-grade were hampered somewhat by sample size considerations. Those who are not overage were in the majority, and the impacts for this group are similar to the full-sample estimates. Those who were older than their classmates or whose age was unknown formed smaller groups, for whom it is difficult to estimate the impacts precisely. The same was true for impacts by student mobility—most students remained in the school where they took the baseline achievement test.

Estimated TFA impacts were similar across students with different baseline achievement scores and students in different grades. We divided the student population into three groups based on their level of initial achievement and found the impacts on math scores for the three subgroups fell in a narrow range—between 2.1 and 2.3 NCE points. The estimated impacts on reading varied more widely—from –0.5 for the middle achievers to 1.1 for the high achievers, with an impact of 0.5 for the low achievers—although all of these estimates are fairly close to zero. The fact that these estimates did not ascend or descend uniformly according to achievement level is consistent with the idea that the range of estimates reflects random fluctuation in estimates that one would expect when examining subgroups that are one-third the size of the overall sample. Similarly, the grade-level impacts shown in Table 5 do not show any patterns, but fluctuate randomly around the impact estimates for the overall sample. Except for grade 1, the estimates of the math impact are more than two NCE points for each grade level. The estimates of the reading impact are less than two NCE points in absolute value for every grade level.

Test Score Impact Findings Robust

To examine the sensitivity of the impact findings, we tested alternative specifications and sub-samples. Based on the sensitivity analyses, we found that the impact of TFA on mathematics achievement, estimated under various assumptions, ranged

[11] One limitation in estimating impacts by race/ethnicity is that the groups are not similarly distributed across regions. Hence, we cannot effectively isolate variation in impacts by race/ethnicity from variation in impacts by region, and these estimates should be interpreted cautiously.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5. Impacts on test scores, student subgroups (NCEs).

	Mathematics		Reading		Sample Size		
Subgroup Comparison	Impact Estimate[a]	Standard Error	Impact Estimate[a]	Standard Error	Blocks	Classes	Students
Full Sample	2.43***	(0.73)	0.56	(0.62)	37	100	1,715
Gender							
Females	2.83***	(0.96)	0.14	(0.82)	37	100	843
Males	1.95*	(1.02)	0.71	(0.89)	37	100	872
Race/Ethnicity							
African American	2.06	(1.81)	0.30	(0.91)	25	69	976
Hispanic	1.89	(1.34)	2.10	(1.58)	13	33	442
Overage for Grade							
Overage	1.23	(1.17)	0.10	(0.66)	37	100	305
Not Overage	1.93**	(0.84)	0.17	(0.80)	37	100	1,191
Missing Age	4.67*	(1.37)	1.97	(2.31)	4	12	205
Mobility Status							
Stayers	2.56***	(0.73)	0.35	(0.72)	37	100	1,622
Movers	0.37	(4.20)	2.08	(2.36)	28	70	89
Initial Achievement							
Low	2.32**	(1.11)	0.51	(0.89)	37	100	464
Middle	2.08	(1.37)	-0.54	(0.88)	37	100	580
High	2.27*	(1.33)	1.14	(0.88)	37	100	671
Grade Level							
Grade 1	1.26	(1.23)	1.09	(1.18)	9	23	320
Grade 2	3.21	(3.44)	1.97	(2.42)	4	10	171
Grade 3	2.58	(1.56)	-1.36	(1.31)	11	34	574
Grade 4	3.14**	(1.05)	0.75	(1.52)	9	25	480
Grade 5	2.58	(2.18)	1.00***	(0.14)	4	8	170

Source: Scores from the Iowa Test of Basic Skills, administered by Mathematica Policy Research, Inc.
Note: All test scores are expressed in NCEs, whose average score nationally is 50 and standard deviation is 21.06.
[a] Impact estimates are regression-adjusted. The benchmark regression model controls for baseline test scores, gender, race/ethnicity, eligibility for free or reduced-price lunch, age (whether overage for grade), and percentage of students in the classroom who were not in the research sample.
* Significantly different from zero at the .10 level, two-tailed test.
** Significantly different from zero at the .05 level, two-tailed test.
*** Significantly different from zero at the .01 level, two-tailed test.

from 2.0 to 3.0 NCEs (which corresponds to a range of 0.13 to 0.19 standard deviation units) and was always statistically significant. The estimated impact on reading achievement ranged from –0.4 to 0.8 NCEs (which corresponds to a range of effect sizes from –0.03 to 0.05), with none of the estimates being statistically significant. All the values within each range led to the same general conclusion—students in TFA classrooms outperformed students in control classrooms in math, and they performed about the same, on average, in reading.

Table 6 shows some illustrative results of the sensitivity analyses. In Alternative Specification 1 in Table 6, we used test score *gains* between the fall and spring as the outcomes. The gain model is more restrictive than the more general model,

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6. Impacts on test scores, sensitivity analyses (NCEs).

	Mathematics		Reading		Sample Size		
Subgroup Comparison	Impact Estimate[a]	Standard d Error	Impact Estimate[a]	Standard d Error	Blocks	Classes	Students
Base model	2.43***	(0.73)	0.56	(0.62)	37	100	1,715
Alternative specifications:							
(1) Gain scores as dependent variable	2.87***	(0.76)	–0.35	(1.02)	37	100	1,732
(2) Errors-in-variables model, reliability = 0.90	2.50***	(0.73)	0.49	(0.66)	37	100	17,15
(3) Errors-in-variables model, reliability = 0.75	2.76***	(0.73)	0.26	(0.79)	37	100	1,715
(4) Included school fixed effects	2.00**	(0.89)	–0.08	(0.66)	37	100	1,715
(5) Dropped blocks if Spanish-language test	2.43***	(0.82)	0.22	(0.61)	33	89	1,551
(6) Adjusted for floor effects (censored regression)	2.01***	(0.71)	0.76	(0.62)	37	100	1,715
(7) Included stayers only	2.56***	(0.72)	0.35	(0.72)	37	100	1,622

Source: Scores from the Iowa Test of Basic Skills, administered by Mathematica Policy Research.
Note: All test scores are expressed in NCEs, whose average score nationally is 50 and standard deviation is 21.06.
[a] Impact estimates are regression-adjusted. The benchmark regression model controls for baseline test scores, gender, race/ethnicity, eligibility for free or reduced-price lunch, age (whether overage for grade), and percentage of students in the classroom who were not in the research sample.
* Significantly different from zero at the .10 level, two-tailed test.
** Significantly different from zero at the .05 level, two-tailed test.
*** Significantly different from zero at the .01 level, two-tailed test.

which allows baseline test scores to have a varying effect on achievement in the spring. Using the score gain model, the estimated impact on math scores was 2.87 and on reading scores was –0.35.

One rationale for using the gain score model rather than the unrestricted post- or pre-test regression is that the pre-test is measured with error, which, if included in the regression model, could bias the coefficient estimates, including the treatment effect. In Alternative Specifications 2 and 3, we included results of an errors-in-variables correction that use alternative estimates of the reliability of the test to adjust the cross-product matrix. While the errors-in-variables correction affects the pre-test coefficients, the TFA impact estimates presented in Table 6 only increased slightly.

In Alternative Specification 4, we included dummy variables for each school to account for school fixed effects—school-specific effects on test scores that are fixed over time. Under this specification, the impact estimate for math was 2.00 with a standard error of 0.89, which makes it significant at the 0.05 level. The impact estimate for reading was –0.08.

We also estimated the model with and without Spanish-language test takers. A number of classrooms in our sample (accounting for 8.7 percent of the students) provided instruction in Spanish, so we administered a Spanish-language version of the test to these students. To see whether these Spanish-language scores (which had to be adjusted to be comparable to the English-language scores) were influencing

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the results, we estimated the impacts separately with these classes excluded (Alternative Specification 5 in Table 6). The math impact excluding the Spanish-language test takers was 2.43, and the impact on reading was 0.22.

We used different methods to correct for floor effects, which occurred because some students received the minimum possible test score. About 7 percent of the sample members received the minimum test score in math, and a similar number did so in reading, suggesting that the test itself was unable to discriminate between low and very low achievers. We used a censored regression model to account for these cases (Alternate Specification 6 in Table 6), and the resulting impacts were 2.01 for math and 0.75 for reading.

Finally, in Alternative Specification 7, we estimated impacts for stayers only. This test removes the effects of students who left their assigned classrooms during the school year to transfer to another school or to cross over from a treatment to a control classroom, or vice versa. There is always a concern that including crossovers might bias the impacts toward zero by attributing performance gains from better teachers to worse teachers and vice versa. (For the benchmark analysis, we classified mobile students according to the classroom to which they were randomly assigned.) Noting that the crossover rates both out of and into TFA classrooms were about the same (4 percent), we estimated the impact on just those students who stayed in their designated classroom. The TFA impact on math for stayers was 2.56, and the impact on reading was 0.35.

We also examined impacts with different control variables in the regression model,[12] different sample weighting schemes, and different exclusion rules, to leave out the small number of classrooms or schools where teachers reported having received help from teacher aides or other teachers or where any possible anomalies might have occurred. In all cases, the impact estimates fell within the ranges described earlier.

In addition to the specification checks above, we examined whether the findings might be sensitive to outliers. We used two methods for checking for outliers. One was to examine the impacts separately by region. The other was to examine the distribution of block-specific impacts, where each block is a group of teachers in the same school at the same grade.

The range of estimates across the six regions varies around the overall estimate for the study, with the impact on mathematics scores ranging from just below zero to 5.5 NCE points and the impacts on reading ranging from just below zero to 2 NCE points. Because each of the six regions represents only one-sixth of the sample, the region-specific impact estimates are imprecise. The variation across regions appears similar to what one might expect from ordinary sampling variation about an overall mean.

Another test for outliers was to examine the distribution of impacts at the block level. Block-specific impacts represent an even finer grain of analysis than the district-specific impacts, so the individual estimates are even less precise. Nevertheless, the pattern is informative.

The distributions of impacts by block for mathematics and reading suggest that the general findings are not driven by outliers. The math impacts by block form a neat bell curve centered on 4 NCEs, with most of the estimates being positive or just

[12] We found that the coefficients on covariates such as pretest and gender differed by grade. However, the less restrictive model with more grade interactions produced no difference in the treatment effects and therefore was not incorporated into the benchmark model. Grade level was one of the student subgroups we examined in Table 7.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

under zero. Only two blocks appeared to be outliers from the rest, and they were negative. Eliminating those outliers would *increase* the estimate of the impact of TFA on math. For reading, the impact estimates also follow an approximately bell-shaped distribution, centered on zero. This is consistent with a story of zero difference in reading, with sampling error surrounding the estimates.

IMPACTS ON OTHER OUTCOMES

In addition to administering achievement tests, we examined other outcomes using data from school records, district records, and teacher reports on classroom management. From these sources, we were able to measure retention in grade, assignment to summer school, disciplinary incidents, tardiness, chronic absence, and the extent to which student behavior disrupted the class. Estimated impacts for most of these outcomes were not statistically significant, as explained below, but they raise important issues for future research.

We found no strong evidence that students in TFA classrooms were either more or less likely to attend summer school or be held back in grade. As Table 7 shows, the differences between TFA and control students—less than 1 percentage point in grade retention (TFA students being held back with slightly greater frequency) and less than 1 percentage point in summer school attendance—were not statistically significant. When the comparison was restricted to novice teachers, we found slightly larger differences, but they were still not significantly different from zero and did not go in a consistent direction (positive or negative).

We also found that TFA teachers had no impact on absenteeism or disciplinary incidents (Table 7). Absenteeism was measured in two ways: (1) number of days absent, and (2) percentage of students who we defined as chronically absent—absent more than 10 percent of the time while enrolled. For both number of days and percent chronically absent, the differences between TFA and control students were small—0.5 days and 0.5 percent, respectively—and not statistically significant.

The estimated impacts on disciplinary incidents were also not statistically significant. We examined two measures of disciplinary incidents: (1) the percentage of students who were ever suspended or expelled, and (2) the number of days suspended. As one would expect in elementary schools, such disciplinary incidents were rare, occurring for about 11 percent of the sample, averaging less than a quarter of one day of suspension per student. TFA had no impact on the number of days suspended, which averaged close to zero for both control students and TFA students. The estimated TFA impact on probability of suspension or expulsion was more substantial, but again not statistically significant.

Teacher-reported experiences in the classroom seem to paint a different picture of absenteeism, discipline, and other classroom management outcomes. TFA teachers were significantly more likely to report that student disruptions and physical conflicts among students in their classrooms were a "serious" problem (Table 8). While 17 percent of the control teachers said that physical conflicts among students were a serious problem, more than a third of the TFA teachers said they were a serious problem. In addition, TFA teachers reported significantly more class interruptions to deal with student disruptions—24 interruptions in the past week reported by TFA teachers, on average, compared with 14 reported by control teachers. In addition, TFA teachers reported greater verbal abuse and a greater frequency of student absenteeism, but the differences were not statistically significant. Differences in the rest of the teacher-reported outcomes in Table 8 were smaller and not statistically significant.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 7. Impacts on other student outcomes.

	Control Mean[a]	TFA Mean	Impact	P-value	Sample Size Blocks	Classrooms	Students
Other academic outcomes							
Retained in grade	12.09	13.03	0.94	0.536	31	84	1,596
Attended summer school	30.52	30.92	0.40	0.884	37	100	1,912
Absenteeism							
Number of days absent	8.31	8.83	0.52	0.415	36	97	1,783
Chronically absent (percentage)	15.07	15.60	0.52	0.794	36	97	1,775
Disciplinary incidents							
Number of days suspended	0.23	0.28	0.04	0.578	31	84	1,574
Ever suspended or expelled (percentage)	10.55	13.31	2.77	0.177	31	84	1,574

Source: Data from school and district records.
[a] Control group means and impacts are regression-adjusted. The regression model controls for baseline test scores, gender, race/ethnicity, eligibility for free or reduced-price lunch, and age (whether overage for grade), as well as percentage of students in the classroom who were not in the research sample.
* Significantly different from zero at the .10 level, two-tailed test.
** Significantly different from zero at the .05 level, two-tailed test.
*** Significantly different from zero at the .01 level, two-tailed test.

Table 8. Impacts on teacher-reports of classroom problems.

Outcome	Control Mean[a]	TFA Mean	Impact	P-value	Number of Teachers
Teacher reports a serious problem with attendance/tardiness (percentage)					
Student tardiness	12.9	16.2	3.3	0.669	96
Student absenteeism/class-cutting	8.6	17.1	8.6	0.237	96
Teacher reports a serious problem with behavior (percentage)					
Physical conflicts among students	17.1	34.3	17.1*	0.073	96
Verbal abuse of teachers	4.3	14.3	10.0	0.107	96
General misbehavior (for example, students talking in class, refusal to follow classroom rules)	22.9	30.0	7.1	0.460	96
Problems in the most recent week (average number)					
Students tardy or absent without excuse	4.5	6.6	2.1	0.108	94
Teacher interrupted class to deal with student disruptions	13.7	24.0	10.2*	0.061	94
Teacher sent child out of the room	2.4	2.6	0.3	0.795	95

Source: Teacher survey.
* Statistically significant at the 0.10 level, two-sided test.
** Statistically significant at the 0.05 level, two-sided test.
*** Statistically significant at the 0.01 level, two-sided test.

There are at least two potential explanations for the findings of no impacts based on school-reported outcomes and potentially harmful impacts based on teacher self-reports. One is that TFA teachers had different expectations and perceptions than control teachers about student behavior, which could lead them to interrupt the class more often for disruptive students or be more prone to describing their

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

students' behavior as problematic. This explanation seems plausible, since TFA teachers and control teachers come from substantially different backgrounds before teaching. Another possibility is that TFA teachers actually had more difficulty managing their classrooms, which resulted in an objective increase in physical conflicts, verbal abuse, and disruption of class time. Because the results presented here on classroom management are inconclusive, further research is needed to fully understand the impacts of TFA on student behavior in the classroom.

SUMMARY AND IMPLICATIONS

The TFA teacher recruitment and training strategy produces teachers who differ in important ways from the other teachers in their schools. The TFA teachers in our sample had strong academic backgrounds, but they had less extensive teacher training than the control teachers in the same schools—before entering the classroom, TFA teachers were less likely to have education degrees, be fully certified, or have substantial student teaching experience. However, the differences were modest, primarily because the control teachers tended to be a diverse group with respect to their training. Many control teachers in these schools, like their TFA counterparts, did not have education degrees; and many were not fully certified and did not have extensive student teaching experience prior to entering the classroom. Hence, in evaluating the impact of TFA teachers in our study, the appropriate counterfactual was not a set of fully certified teachers with education majors and substantial student teaching experience, but rather a diverse group with mixed training.

Our estimates show that TFA teachers had a positive impact on the math achievement of their students—average math scores were higher among TFA students than among control students, and the difference was statistically significant. TFA teachers did not have an impact on reading achievement—average reading gains were comparable among the TFA and control students. The findings regarding math and reading impacts were fairly consistent across subgroups and they were robust to changes in modeling assumptions and specifications. Our estimates also suggest that TFA teachers had larger impacts on both math and reading achievement when compared with novice control teachers than when compared with all control teachers, but the limited sample size for the novice teacher estimate does not support a definitive conclusion on this point.

There is no easy explanation for the pattern of positive impacts on math and no impacts on reading. Differences among teachers in math-related college majors do not account for the larger effect on math scores. Only 25 percent of the teachers in the study had college majors or minors in subjects that could be considered quantitative, and the proportions were similar for TFA and control teachers. With respect to reading, TFA program staff indicated to the authors that they had been trying to improve the reading component of the summer training, and that those efforts had not been in place until after the study year. It is possible that since those changes were implemented, TFA now has a positive impact on reading scores as well as math scores.

Estimates for other student outcomes did not reveal any other impacts of TFA. TFA teachers were more likely than control teachers to report having had problems with student disruptions and physical conflicts, but this fact may simply reflect differences between TFA and control teachers' expectations and perceptions regarding student behavior, rather than actual differences between classrooms.

The positive impacts of TFA teachers on student test scores should not be interpreted as evidence that traditional teacher preparation routes provide training inferior to that provided by TFA. First, as noted above, the control teachers in our com-

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

parisons included many who entered the profession through nontraditional routes. Second, this study was designed to assess the combined recruitment and productivity effects of TFA. We did not try to disentangle these two effects, because the combined effect is most relevant for policymakers.

Nevertheless, our findings have important implications for a variety of stakeholders. Program funders, program operators, and policymakers at the state and federal levels have an enduring interest in finding ways to attract and retain high-quality teachers in low-income communities. District officials and school staff in such areas have an especially practical interest in the same question, particularly in the short term, with federal requirements under No Child Left Behind to place a highly qualified teacher in every classroom. Finally, parents and children in low-income communities are most directly affected by decisions about who will teach in their schools. The findings have implications for each of these groups.

From the perspective of a community or a school faced with the opportunity to hire TFA teachers, the findings of this study suggest that TFA offers an appealing pool of candidates. First, the positive impacts on math scores suggest that by hiring TFA teachers, a school can expect to increase the average math achievement of its students without lowering their reading achievement. Second, the consistent pattern of positive or zero impacts on test scores across grades, regions, and student subgroups suggests that there is little risk that hiring TFA teachers will reduce achievement, either for the average student or for most subgroups of students. Finally, TFA teachers are paid the same as other teachers, so schools pay no direct costs for the achievement increase and school districts typically contribute only $1,500 per corps member to offset screening and recruiting costs. This contrasts with other interventions, such as reduction in class size, that have been shown to increase achievement but that entail substantial direct costs.

One could expand this reasoning to conduct a larger assessment of whether, from society's perspective, TFA is a cost-effective way to attract teachers to low-income schools. However, a full cost-effectiveness assessment would require information on a number of factors our study does not directly address. For example, although TFA teachers are paid on the same salary scale as their counterparts, they may create hidden costs if they leave their jobs sooner—for example, at the end of their two-year commitment—and have to be replaced more frequently than their non-TFA peers. Measuring such costs would be difficult, because the retention rates of TFA and non-TFA teachers are not well documented. Our data showed no significant difference in within-year attrition rates, but since they cover only a single school year, they cannot be used to compare attrition rates over time between the TFA and control teachers. Hanushek, Kain, and Rivkin (2004) show that teacher attrition rates are particularly high in schools that serve large numbers of academically disadvantaged students—exactly the types of schools where TFA places teachers. Therefore, there is no reason to presume that TFA teachers have a higher attrition rate than other new teachers in the same schools.

From the perspective of TFA and its funders, our findings clearly show that the organization is making progress toward its primary mission of reducing inequities in education—it supplies low-income schools with academically talented teachers who contribute positively to the academic achievement of their students. The success of TFA teachers is not dependent on teachers having extensive exposure to teacher practice or training. Even though TFA teachers generally lack any formal teacher training beyond that provided by TFA, they produce higher student test scores than the other teachers in their schools—not just other novice teachers or uncertified teachers, but also veterans and certified teachers.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Finally, these findings provide important information to policymakers who are trying to improve the educational opportunities of children in poor communities. The findings that many of the control teachers were not certified or did not have formal pre-service training highlights the need for programs or policies that offer the potential of attracting good teachers to schools in the most disadvantaged communities. The findings presented here show that TFA is one such program.

ACKNOWLEDGMENTS

This research was supported by grants from the Smith Richardson Foundation, William and Flora Hewlett Foundation, and the Carnegie Corporation. We thank Richard Murnane, Jeff Smith, and Mark Dynarski, as well as the anonymous referees for their careful review. The study could not have been done without gracious cooperation from numerous school superintendents, principals, teachers, and other district officials and school staff whom we cannot thank by name but who were enormously patient and generous with their time.

STEVEN GLAZERMAN is Senior Researcher at Mathematica Policy Research in Washington, DC.

DANIEL MAYER is Assistant Superintendent for Curriculum, Instruction, and Assessment in the Maynard Public Schools in Maynard, MA.

PAUL DECKER is Vice President and Managing Director for Human Services Research at Mathematica Policy Research in Princeton, NJ.

REFERENCES

Ballou, D., & Podgursky, M. (1998). The case against teacher certification. Public Interest, 132, 17–29.

Barron's Profiles of American Colleges. (2003). Hauppauge, NY: Barron's Educational Series, Inc.

Darling-Hammond, L. (1994). Who will speak for the children? How Teach for America hurts urban school children. Phi Delta Kappan, 76, 21–34.

Darling-Hammond, L. (1996). What matters most: A competent teacher for every child. Phi Delta Kappan, 77, 193–201.

Decker, P., Mayer, D., & Glazerman, S. (2004). The effects of Teach for America on students: findings from a national evaluation. Princeton, NJ: Mathematica Policy Research, Inc.

Finn, J., & Achilles, C. (1999). Tennessee's class size study: Findings, implications, misconceptions. Educational Evaluation and Policy Analysis, 21(2), 97–110.

Hanushek, E. A. (1974). Efficient estimators for regressing regression coefficients. American Statistician, 28(2), 66–67.

Hanushek, E. A., Kain, J. F., & Rivkin, S. (2004). Why public schools lose teachers. Journal of Human Resources, 39(2), 326–354.

Laczko-Kerr, I., & Berliner, D. (2002). The effectiveness of Teach for America and other under-certified teachers on student academic achievement: A case of harmful public policy. Education Policy Analysis Archives, 10, 37.

Raymond, M., Fletcher, S., & Luque, J. (2001). Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas. Stanford, CA: CREDO, Hoover Institution, Stanford University.

15206688, 2006, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pam.20157 by Vanderbilt University Medical, Wiley Online Library on [10/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License