MATHEMATICA Policy Research

REPORT

REVISED FINAL REPORT

Impacts of the Teach For America Investing in Innovation Scale-Up

February 8, 2017

Melissa A. Clark
Eric Isenberg
Albert Y. Liu
Libby Makowsky
Marykate Zukiewicz

Submitted to:
Teach For America
1413 K St. NW, 7th Floor
Washington, DC 20005

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

Project Director: Melissa A. Clark
Reference Number: 06889.740

This page has been left blank for double-sided copying.

ERRATUM

This report has been revised from the original version, released in March 2015, in response to an error discovered in the administration of one of the math assessments used for the evaluation—the Woodcock-Johnson Applied Problems assessment. This error caused scores on the assessment to be inappropriately constrained, which may have prevented us from reliably measuring Teach For America (TFA) teachers' impact on students' performance on this assessment. As a result, we have revised the report to exclude results from the Applied Problems assessment and rely only on other available assessment data for the evaluation.

The Woodcock-Johnson Applied Problems assessment, as correctly administered, contains 63 questions which increase in difficulty. Under established test administration procedures, the assessor begins the assessment with a pre-specified question that varies based on the student's grade level, with students at higher grade levels beginning with more difficult questions. The assessor then progresses through the remaining questions until the child answers six questions in a row incorrectly, at which point the assessment ends and a score can be assigned.

For the TFA-i3 evaluation, assessors administered the Applied Problems and other Woodcock-Johnson assessments to students individually. The assessment was programmed into laptop computers from which assessors read the questions and entered the students' responses. However, due to an error in programming specifications, the Applied Problems assessment stopped at the 29th item instead of allowing administration of the full 63. As a result, many students' scores were inappropriately constrained—they reached the end of the assessment before answering six items in a row incorrectly, and thus the score was not a valid estimate of their ability. We administered the assessment to students in prekindergarten through grade 2. The scores of 66 percent of these students (24 percent of prekindergarten students, 57 percent of kindergarteners, 88 percent of first graders, and 96 percent of second graders) were inappropriately constrained by the administration error, meaning that the students answered the 29th question and finished the assessment without having answered 6 questions in a row incorrectly.

Mathematica first discovered the error in test administration procedures for the Applied Problems assessment in fall 2016, when, upon re-examining the programming specifications for the assessment, we realized that the assessment program did not allow administration of the full 63 items. At this time we examined the raw response data and determined that, given the high proportion of students whose scores were artificially constrained, we would not be able to use the scores to reliably estimate impacts on students' math achievement. We contacted TFA in October 2016 to tell them about the error and its implications and inform them of our intention to revise the report, excluding the Applied Problems scores from our analysis.

This page has been left blank for double-sided copying.

ACKNOWLEDGEMENTS

This study would not have been possible without the contributions of many individuals and organizations. First and foremost, we are grateful for the cooperation of the school districts, schools, teachers, and students who participated in the study. We also thank the Teach For America staff who provided essential information about their program over the course of the study.

We are grateful for input on the study design provided by members of the study's technical working group, which included Margaret Burchinal, Laura Hamilton, Jane Hannaway, Brian Jacob, Helen Ladd, Michael Petrilli, Andrew Porter, and James Wyckoff. The study benefited greatly from their expertise. Mike Puma, the study's technical assistance liaison for the National Evaluation of Investing in Innovation, also provided valuable input on the study design.

The study also benefited from the contributions of many people at Mathematica. A large team of dedicated staff recruited districts and schools into the study. Survey Director Kathy Sonnenfeld led the study's data collection effort, with assistance from Barbara Kennen and Erin Panzarella. Alexander Johann, Nikhil Gahlawat, Chelsea Swete, and Kathryn Gonzalez provided excellent research and programming assistance. Phil Gleason provided valuable input on the study design, and Hanley Chiang and Barb Devaney provided thoughtful, critical reviews of the final report. Jennifer Baskwell provided expert production support.

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

This page has been left blank for double-sided copying.

EXECUTIVE SUMMARY

Teach For America (TFA) is a nonprofit organization that seeks to improve educational opportunities for disadvantaged students by recruiting and training teachers to work in low-income schools. The program uses a rigorous screening process to select college graduates and professionals with strong academic backgrounds and leadership experience and asks them to commit to teach for two years in high-needs schools. These teachers, called corps members, typically have no formal training in education but participate in an intensive five-week training program before beginning their first teaching job. TFA then provides them with ongoing training and support throughout their two-year commitment. TFA encourages teachers who complete their two-year commitment, known as TFA alumni, to continue to work to reduce educational inequity, whether by remaining in the classroom or by assuming roles of educational leadership and advocacy.

In 2010, TFA launched a major expansion effort, funded in part by a five-year Investing in Innovation (i3) scale-up grant of $50 million from the U.S. Department of Education. Under the i3 scale-up, TFA planned to increase the size of its teacher corps by more than 80 percent by September 2014, with the goal of placing 13,500 first- and second-year corps members in classrooms by the 2014–2015 school year and expanding to 52 regions across the country. While TFA ultimately fell short of the growth goals set in its scale-up application (Mead et al. 2015), by the 2012–2013 school year, the second year of the scale-up, it had expanded its placements by 25 percent, from 8,217 to 10,251 first- and second-year corps members.

Using a rigorous random assignment design to examine the effectiveness of TFA elementary school teachers in the second year of the i3 scale-up, Mathematica Policy Research found that first- and second-year corps members recruited and trained during the scale-up were as effective as other teachers in the same high-poverty schools in both reading and math. We found that TFA teachers in lower elementary grades (prekindergarten through grade 2) had a positive, statistically significant effect on students' reading achievement of 0.12 standard deviations, or about 1.3 additional months of learning for the average student in these grades nationwide. We also found that TFA teachers in grades 1 and 2 had a positive effect on student math achievement of 0.16 standard deviations, or about 1.5 additional months of learning. This difference was almost statistically significant at conventional levels (p-value = 0.054). We did not find statistically significant impacts for other subgroups of TFA teachers that we examined. Although the i3 scale-up expanded TFA placements at all grade levels, this analysis focuses only on teachers in prekindergarten through grade 5—36 percent of all TFA teachers recruited during the first two years of the scale-up—and the results pertain to this group of corps members.

A. Background

The most rigorous available prior evidence suggests that TFA teachers have been more effective than their non-TFA counterparts in math and about the same in reading. There have been two previous large-scale random assignment studies of TFA teachers. These studies randomly assigned students to classes taught by TFA teachers or classes taught by non-TFA teachers in the same grade and school. Random assignment ensured that the students taught by TFA and non-TFA teachers were similar at the start of the school year, so any differences in

students' test scores at the end of the school year could be attributed to the effectiveness of the teachers rather than to underlying differences in the students.

- The first experimental study (Decker et al. 2004) focused on TFA teachers in grades 1 through 5 during the 2001–2002 and 2002–2003 school years. The study found that students with TFA teachers performed as well as students with non-TFA teachers in reading and significantly better in math (by approximately 0.15 standard deviations).
- The second experimental study (Clark et al. 2013) examined the effectiveness of middle and high school math teachers from TFA during the 2009–2010 and 2010–2011 school years. It found that secondary math teachers from TFA were more effective than other math teachers in the same schools, increasing students' math achievement by 0.07 standard deviations.

Several well-designed nonexperimental studies have also examined the effects of TFA teachers on student achievement in New York City (Kane et al. 2008; Boyd et al. 2006), North Carolina (Xu et al. 2008; Henry et al. 2014), and Miami (Hansen et al. 2014). The studies used test score data and other student background characteristics to attempt to account for any underlying differences in the types of students assigned to TFA and non-TFA teachers in the same schools, and have compared TFA teachers with non-TFA teachers with similar years of experience. In math, the nonexperimental studies have found that TFA teachers perform better than other novice teachers (Xu et al. 2008; Henry et al. 2014; Hansen et al. 2014) or about the same (Kane et al. 2008; Boyd et al. 2006). In reading, some studies have found that TFA teachers perform about the same as other novice teachers in the same schools (Kane et al. 2008; Hansen et al. 2014), whereas other studies have found they perform either slightly better (Henry et al. 2014) or slightly worse (Boyd et al. 2006).

B. TFA's program model and implementation of the i3 scale-up

TFA seeks to improve student achievement by providing high quality teachers to high-needs schools. Key components of its approach include (1) recruiting applicants to the program; (2) selecting applicants it predicts have the potential to become effective teachers and asking them to make a two-year commitment to teaching in a high-needs school; (3) providing those who are selected and join the program, known as corps members, with five weeks of preservice training before they begin their first teaching job; (4) helping corps members find jobs in high-needs schools; and (5) providing ongoing training and support to corps members throughout their two-year commitment.

Recruitment. TFA recruits undergraduate and graduate students at college campuses across the country, as well as professionals. The program places a high priority on recruiting a racially and economically diverse set of corps members and on recruiting corps members to teach hard-to-staff subjects such as science, math, and special education. More than 48,000 applicants applied to join the 2012 TFA corps, including more than 5 percent of the graduating senior class at 135 colleges and universities.

Selection. TFA relies on an intensive, data-driven admissions process to select the candidates it predicts are most likely to succeed in the classroom. The process includes a web-based writing activity; a telephone interview; and a day-long, in-person interview that includes a one-on-one interview, a sample teaching lesson, and group discussions. At each stage of the

admissions process, TFA prioritizes the selection of candidates with the following attributes: (1) a commitment to reducing educational inequality; (2) demonstrated leadership ability and interpersonal skills to motivate others; (3) achievement in academic, professional, extracurricular, and/or volunteer settings; (4) perseverance in the face of challenges; (5) critical thinking skills; (6) organizational ability; and (7) respect for and ability to work with people from diverse background and experiences Approximately 17 percent of applicants for the 2012 corps were selected into the program; of these, 71 percent accepted the offer of admission.

Preservice training. After their acceptance into TFA, corps members are required to participate in a series of preservice training activities, the main component of which is a five-week, full-time residential summer program known as summer institute. During summer institute, corps members receive group instruction on curriculum, literacy, and diversity; teach summer school students under the supervision of experienced teachers; observe other teachers; receive written and oral feedback on teaching from advisors; attend small-group sessions to reflect on teaching practice; and participate in clinics designed to improve lesson-planning skills According to TFA staff, required summer institute activities in 2012 totaled at least 240 hours, with some variation by institute and the subject and grade level the corps member would be teaching.

Placement. TFA assigns corps members to the region where they will teach at the time of their acceptance into the program. Consistent with its goals for the i3 scale-up, TFA expanded from 40 regions in 2010–2011 to 43 regions in 2011–2012 (the first year of the scale-up) and to 47 regions in 2012–2013. In each region, corps members apply for positions in the public school districts, public charter schools, and community-based organizations in that region that have partnered with TFA. In the 2012–2013 school year, 84 percent of incoming corps members took jobs in high-poverty schools, defined as those in which 60 percent or more of students were eligible for free or reduced-price lunch. Nearly two-thirds of first-year corps members (65 percent) taught in traditional public schools, and approximately one-third (33 percent) taught in charter schools. In 2011, the first year of the scale-up, TFA placed 5,031 new teachers (a 12 percent increase from the prior year). In 2012, the second year of the scale-up, it placed 5,807 new teachers (a 15 percent increase from the first year).

Ongoing training and support. After partner schools and districts hire corps members, regional TFA staff provide them with ongoing training and support during their two-year commitment. This includes one-on-one coaching support, group meetings customized by grade and subject, and access to additional classroom resources and assessments via an online portal. Corps members in most regions must also complete alternative certification programs, state-defined routes through which people can begin teaching before completing all the requirements for state certification.

In our study of TFA's implementation of the i3 scale-up, we saw little evidence of substantive changes to TFA's approach during the first two years of the scale-up. However, we did see some declines in corps members' satisfaction with the program. For instance, the percentage of corps members who felt that the summer institute was critical for being an effective teacher fell from 85 to 75 percent from 2009–2010 (two years before the i3 scale-up) to the scale-up's second year, and the percentage reporting either positive or very positive overall satisfaction with the program declined from 64 to 57 percent over this period.

C. Study design, data collection, and analysis

We used a rigorous random assignment design to assess the effectiveness of TFA teachers recruited in the first two years of the i3 scale-up. Next, we describe the study design, study sample, data collection, and analysis.

Random assignment design. At the start of the 2012–2013 school year, we randomly assigned students in each participating school and grade level to a class taught by a TFA teacher or a class taught by a teacher from another certification route. The non-TFA teachers, whom we refer to as comparison teachers, were meant to represent the types of teachers who would have taught the students had TFA teachers not been teaching in a particular school. Random assignment ensured that there were no systematic differences between students assigned to TFA teachers and those assigned to the comparison teachers at the start of the school year. Therefore, any systematic differences in end-of-year achievement between the two groups could be attributed to the causal effect of being assigned to a TFA teacher rather than to a teacher from another certification route in the same school.

Sample. We recruited sample members during the 2011–2012 school year to participate in the study the following school year. The final sample included 10 states, 13 school districts and other TFA placement partners, 36 schools, and 156 teachers (66 TFA and 90 comparison teachers). The sample of TFA teachers was limited to those recruited in the first two years of the scale-up, who were in their first or second year of teaching at the time of the study, whereas the comparison teachers included both novice and experienced teachers teaching in the same schools and grades as the TFA teachers. We randomly assigned 3,724 students to classes and obtained valid outcome test score data for 2,152 students. Of these students, 2,123 had valid outcome data for reading and 1,182 had valid outcome data for math.[1]

Data on characteristics of TFA and comparison teachers. In the spring of the study year, we administered a survey to teachers in the study to collect information on their professional background and experiences. The survey asked about teachers' educational background, teaching experience, preparation for teaching, support received during the school year, views toward teaching, and demographic characteristics.

Data on student outcomes. To measure student achievement outcomes, we collected end-of-year reading and math test scores from the 2012–2013 school year for all randomly assigned students with parental consent. In the lower elementary grades (prekindergarten through grade 2), we assessed students using reading and math assessments from the Woodcock-Johnson III achievement test.[2] In the upper elementary grades (3 to 5), in which annual reading and math

[1] We did not collect test score data for students who were randomly assigned but never enrolled in a study school, those who left a district before the end of the school year, or those whose parents did not consent for them to participate in the study. Rates of missing outcome data were very similar for students assigned to TFA and comparison teachers.

[2] We assessed students in reading and math in grades prekindergarten through 2. However, due to an error in test administration procedures for one of the math assessments (the Woodcock-Johnson Applied Problems assessment), we were unable to use those scores in the analysis. This assessment was the only math assessment for students in prekindergarten and kindergarten, and one of two for students in grades 1 through 2. Thus, our analysis of math

assessments were required by the federal No Child Left Behind Act, we collected state assessment data from district records. We also collected prior years' test scores from state assessments when available, along with other student background characteristics from district records.

Analysis. To estimate the effectiveness of TFA teachers relative to the comparison teachers, we compared end-of-year test scores of students assigned to the TFA teachers and those assigned to the comparison teachers. Because students in the study were randomly assigned to teachers, we can attribute systematic differences in achievement at the end of the study school year to the relative effectiveness of TFA and comparison teachers, rather than to the types of students taught by these two different groups of teachers. In addition to the impact analysis described in this report, the evaluation included an implementation analysis (Zukiewicz et al. 2015) that describes key features of TFA's program model and its implementation of the i3 scale-up.

D. Teach For America and comparison teachers in the sample

Understanding the characteristics of the TFA teachers in the sample and the teachers with whom they were compared can provide important context for interpreting the impact estimates. As expected, given that TFA follows a distinctive model for selecting and recruiting corps members and our approach to selecting the sample, we found many differences between the TFA and comparison teachers in the sample.

- **TFA teachers had substantially less teaching experience than comparison teachers.** As expected, given that our sample was limited to first- and second-year corps members, the TFA teachers in the study had significantly less teaching experience, on average, than comparison teachers. In all but one special case, TFA teachers were in their first or second year of teaching, compared with only 13 percent of comparison teachers. The TFA teachers had an average of 1.7 years of experience compared with 13.6 years among the comparison teachers.
- **The sample of TFA teachers was younger and included fewer racial or ethnic minorities than the sample of comparison teachers.** The average TFA teacher in the sample was 24 years old, compared with an average age of 43 among comparison teachers. About 90 percent of TFA teachers were female, compared with 99 percent of comparison teachers. About 70 percent of TFA teachers were white and non-Hispanic, compared with only 55 percent of comparison teachers.
- **TFA teachers were more likely than comparison teachers to have graduated from a selective college or university, but a substantial proportion of comparison teachers graduated from a selective school.** About 76 percent of TFA teachers in our sample had graduated from a selective college, compared with 40 percent of comparison teachers. TFA teachers were less likely than comparison teachers to have majored in early childhood education or elementary education, and more likely to have majored in a field unrelated to education.

achievement does not include any students in prekindergarten or kindergarten, because no valid math scores are available. We describe the error in administration procedures for the Applied Problems assessment and its implications in greater detail in Appendix A.

- **TFA teachers were less satisfied with many aspects of teaching.** For example, relative to comparison teachers, TFA teachers reported lower levels of satisfaction with their influence over school policies, support from administration, opportunities for professional development, and opportunities for professional advancement. However, similar percentages of TFA and comparison teachers were satisfied with the opportunities to help students and personal fulfillment offered by the teaching profession.
- **The comparison teachers in our sample were certified primarily through traditional routes.** About 85 percent of comparison teachers in the sample were from traditional routes and 15 percent were from other alternative routes to certification.

E. TFA impacts on math and reading achievement

On average, the TFA teachers in our sample were as effective as comparison teachers in both reading and math. In both subjects, differences in test scores between students assigned to TFA teachers and those assigned to comparison teachers were not statistically significant.

We found that TFA teachers in lower elementary grades (prekindergarten through grade 2) had a positive, statistically significant effect on student reading achievement of 0.12 standard deviations, or about 1.3 additional months of learning for the average student in these grades nationwide. We also found that TFA teachers in grades 1 and 2 had a positive effect on student math achievement of 0.16 standard deviations, or about 1.5 additional months of learning. This difference was almost statistically significant at conventional levels (p-value = 0.054). In neither subject did we find statistically significant differences for other grade levels or when we compared TFA teachers with novice comparison teachers. When we compared TFA teachers with traditionally certified teachers, we found that their students performed similarly in reading and that the students of TFA teachers scored about 0.10 standard deviations higher in math. However, this difference was not statistically significant at conventional levels (p-value = 0.098).

F. Conclusions

In this evaluation we documented TFA's experiences as it undertook an ambitious five-year scale-up effort, and we provided rigorous estimates of the program's effectiveness in the second year of the scale-up. We found that TFA elementary school teachers recruited in the first and second years of the i3 scale-up were as effective as other teachers in the same high-poverty schools in teaching both reading and math. We found that TFA teachers in lower elementary grades had a positive, statistically significant effect on student reading achievement and TFA teachers in grades 1 and 2 had a positive effect on student math achievement that was almost statistically significant at conventional levels.

Our main findings differ from some earlier studies showing that the effectiveness of TFA teachers differed by subject, with TFA teachers more effective at teaching math and just as effective as other teachers in teaching reading. We find differences across grade levels, with TFA teachers more effective in early elementary grades and as effective in upper elementary grades. Our study provides a snapshot of TFA's effectiveness at the elementary school level in the second year of the i3 scale-up, but it is possible that the effectiveness of TFA's teachers could either increase or decrease as the program continues to strive to meet the needs of schools with many high-poverty students.

I. INTRODUCTION

Teach For America (TFA) is a nonprofit organization that seeks to improve educational opportunities for disadvantaged students by recruiting and training teachers to work in low-income schools. The program uses a rigorous screening process to select college graduates and professionals with strong academic backgrounds and leadership experience and asks them to commit to teach for two years in high-needs schools. These teachers, called corps members, typically have no formal training in education but participate in an intensive five-week training program before beginning their first teaching job. TFA then provides them with ongoing training and support throughout their two-year commitment. TFA encourages teachers who complete their two-year commitment, known as TFA alumni, to continue to work to reduce educational inequity, whether by remaining in the classroom or by assuming roles of educational leadership and advocacy.

TFA was founded in 1989 and placed its first cohort of 384 corps members in classrooms in the 1990–1991 school year. Since that time, the program has launched several major expansion efforts, and in the 2010–2011 school year, TFA had more than 8,200 first- and second-year corps members teaching in 40 urban and rural regions across the country.

In 2010, TFA launched another major expansion effort, funded in part by a five-year Investing in Innovation (i3) scale-up grant of $50 million from the U.S. Department of Education. This was one of four i3 scale-up grants awarded in 2010—the scale-up grants were intended to fund the expansion of programs with rigorous evidence of prior effectiveness in improving student achievement. Through the i3 scale-up project, TFA planned to increase the size of its teacher corps by more than 80 percent by September 2014, with the goal of placing 13,500 first- and second-year corps members in classrooms by the 2014–2015 school year and expanding to 52 regions across the country.

TFA contracted with Mathematica Policy Research to conduct a rigorous independent evaluation of the i3 scale-up project's effectiveness, a requirement for all i3 scale-up grantees. The evaluation includes an analysis of TFA's implementation of the i3 scale-up and an impact analysis examining the effectiveness of TFA elementary school teachers (prekindergarten through grade 5) recruited under the scale-up. Because the evaluation, including analysis and reporting, was to be completed within the i3 grant period, the study includes only the first two cohorts of TFA teachers recruited as part of the scale-up effort. This report presents findings from the impact analysis.

A. Previous research on TFA

Because of its unconventional approach to recruiting and training teachers, TFA has generated some controversy. Critics have argued that TFA teachers are underprepared for the challenges of teaching in high-needs schools and that they tend to leave the profession before gaining the experience needed to teach effectively (Darling-Hammond 2011; Ravitch 2013). Proponents argue that TFA's rigorous screening process and intensive training provide an important source of effective teachers to high-needs schools and that many of its teachers continue to work to improve educational opportunity even after they complete their two-year teaching commitment (Rotherham 2009).

The most rigorous available prior evidence suggests that TFA teachers have been more effective than their non-TFA counterparts in math and about the same in reading. There have been two previous large-scale studies of TFA teachers that randomly assigned students to classes—the most rigorous possible research design. In both studies, students were randomly assigned to classes taught by TFA teachers or classes taught by non-TFA teachers in the same grade and school. Random assignment ensured that the students taught by TFA and non-TFA teachers were similar at the start of the school year, so any differences in student test scores at the end of the school year could be attributed to the effectiveness of the teacher rather than underlying differences in the students.

- The first experimental study (Decker et al. 2004) focused on TFA teachers in grades 1 through 5 during the 2001–2002 and 2002–2003 school years. The study found that students with TFA teachers performed as well as students with non-TFA teachers in reading and significantly better in math (by approximately 0.15 standard deviations). The impact on math was larger (0.26 standard deviations) when novice TFA teachers (those in their first or second year of teaching) were compared with novice non-TFA teachers.
- The second experimental study (Clark et al. 2013) examined the effectiveness of middle and high school math teachers from TFA during the 2009–2010 and 2010–2011 school years. It found that secondary math teachers from TFA were more effective than other math teachers in the same schools, increasing student math achievement by 0.07 standard deviations. TFA teachers in their first two years of teaching outperformed even the most experienced non-TFA teachers (those with more than five years of experience), again increasing student math achievement by 0.07 standard deviations (Chiang et al. 2014).

Several well-designed nonexperimental studies have also examined the effects of TFA teachers on student achievement in New York City (Kane et al. 2008; Boyd et al. 2006); North Carolina (Xu et al. 2008; Henry et al. 2014); and Miami (Hansen et al. 2014). The studies collectively span grade levels 3 through 12. They use test score data and other student background characteristics to attempt to account for any underlying differences in the types of students assigned to TFA and non-TFA teachers in the same schools. They also use teacher characteristics—especially teacher experience—to account for differences between teachers aside from their entry route into teaching. Because they account for teacher experience and school characteristics, these studies implicitly seek to compare the achievement of students of TFA teachers to the achievement of students of other novice teachers in the same schools.

In math, the nonexperimental studies have found that TFA teachers perform better than other novice teachers (Xu et al. 2008; Henry et al. 2014; Hansen et al. 2014) or about the same (Kane et al. 2008; Boyd et al. 2006). One study—Xu et al. (2008)—found that TFA high school teachers performed better than experienced teachers from other routes; the other studies did not investigate this question. In reading, some studies have found that TFA teachers perform about the same as other novice teachers in the same schools (Kane et al. 2008; Hansen et al. 2014), whereas other studies have found they perform either slightly better (Henry et al. 2014) or slightly worse (Boyd et al. 2006). Three of the studies reported results separately for upper elementary school teachers (the group most comparable to our own sample), and the findings for these teachers matched the overall findings for each study. Within the elementary school subsamples, Henry et al. (2014) found that TFA teachers outperformed other novice teachers in

both reading and math, Kane et al. (2008) found no difference between TFA and other novices, and Boyd et al. (2006) found that TFA teachers performed about the same as other novices in math but achieved smaller gains in reading.[3]

B. Goals for the evaluation

All i3 scale-up grantees were required to commission a rigorous independent evaluation of their scale-up efforts. Although TFA was awarded the grant based in part on past evidence of its effectiveness in improving student achievement, the program's effectiveness under the scale-up may differ from its effectiveness at its previous scale. Under the scale-up, TFA planned an ambitious 80 percent expansion of its teaching corps over the four years of the scale-up grant. The effectiveness of TFA's teachers recruited under the scale-up will depend on TFA's ability to attract enough high-quality applicants to meet its expansion goals without compromising its selection standards and its ability to expand its staff and infrastructure to keep pace with the growth of its teaching corps.

For these reasons, it is important to document how TFA implemented the scale-up and to rigorously examine the impact of teachers recruited and trained during the scale-up period. The evaluation thus includes two main components:

1. The implementation analysis (Zukiewicz et al. 2015) describes key features of the scale-up implementation. It documents whether the scale-up was successful in increasing the number of TFA teachers and meeting TFA's other specified goals as well as examining whether TFA maintained fidelity to its core program model during the first two years of the scale-up.
2. The impact analysis, presented in this report, relies on random assignment of students to teachers to measure the relative effectiveness of TFA teachers compared with non-TFA teachers in the same grades and schools. We study TFA teachers in prekindergarten through grade 5 who were hired as part of the scale-up.

Although the i3 scale-up expanded TFA placements at all grade levels, the impact analysis focuses only on teachers in prekindergarten through grade 5, who made up 36 percent of all TFA teachers recruited during the first two years of the scale-up. Focusing our sample on a more limited set of grades allowed us to obtain a larger sample—and more precise impact estimates—for these particular grades. The study focused on prekindergarten through grade 5 because (1) the most rigorous experimental evidence at the elementary school level (Decker et al. 2004) is more than 10 years old, and there is more recent experimental evidence at the secondary level (Clark et al. 2013); (2) nonexperimental evidence has generally focused only on grades 4 and above; and (3) there is no previous rigorous evidence on the effectiveness of TFA teachers in prekindergarten and kindergarten, so including these grade levels allowed us to fill this gap in the literature for reading.[4] Although all TFA teachers in the study were hired during the scale-up, we do not attempt to distinguish between teachers who were hired as a result of the scale-up

[3] Henry et al. (2014) examined teachers in grades 3 through 5, whereas Kane et al. (2008) and Boyd et al. (2006) examined teachers in grades 4 and 5.

[4] We are unable to estimate impacts on math achievement at these grade levels due to an error in administration of the math assessment for students in prekindergarten and kindergarten.

compared with those who would have been hired even in its absence—the results reflect the combined impacts of these two groups.

To ensure the independence of the impact analysis, TFA staff reviewed the report for accuracy of information about the program but did not make any modifications to the findings. TFA staff also assisted in our efforts to recruit districts for the study by providing lists of corps member placements by district and school, and they provided information and data that we used to describe the program and implementation of the scale-up. However, they played no role in selecting schools and districts for the sample; randomly assigning students; testing students; collecting data on schools, teachers, or students in the impact analysis sample; or analyzing the data.

II. STUDY DESIGN, DATA, AND METHODS

We used a rigorous random assignment design to assess the effectiveness of TFA teachers recruited in the first two years of the i3 scale-up. In this chapter, we describe the study's design, data collection, and the methods we used for the analysis.

A. Experimental design

The study used an experimental design to assess the effectiveness of TFA teachers relative to teachers from other certification routes. Students in the same school and grade level were randomly assigned to a class taught by a TFA teacher or a class taught by a teacher from another route. The non-TFA teachers, whom we refer to as comparison teachers, were meant to represent the types of teachers who would have taught the students had TFA teachers not been teaching in a particular school. Random assignment ensured that there were no systematic differences between students assigned to TFA teachers (the treatment group) and those assigned to the comparison teachers (the control group). Therefore, any systematic differences in end-of-year achievement between the two groups could be attributed to the causal effect of being assigned to a TFA teacher rather than to a teacher from another certification route in the same school.

The experimental design allows us to estimate TFA elementary school teachers' effectiveness relative to other teachers in the same school, but it cannot tell us *why* TFA teachers' effectiveness may differ from that of other teachers. In particular, we cannot distinguish between differences in effectiveness due to the training that TFA teachers receive compared with the training of other teachers—many of whom were traditionally certified—and differences that may arise because of the background characteristics of TFA and comparison teachers, such as years of experience in teaching, college selectivity, college major, and academic ability. We describe the training that corps members receive in Chapter III and document the differences in teacher characteristics between the TFA and comparison groups in Chapter IV.

1. Eligible teachers

The study was designed to examine the effectiveness of TFA corps members recruited during the first two years of the i3 scale-up. Any first- or second-year TFA corps member teaching in the 2012–2013 school year (the second year of the i3 scale-up) was potentially eligible for the study sample. This included TFA corps members recruited in the first year of the scale-up (in their second year of teaching in the 2012–2013 school year) and those recruited in the second year of the scale-up (in their first year of teaching in the 2012–2013 school year). Teachers who had entered the profession via TFA prior to the scale-up and remained in the classroom after completing their two-year commitment—known as TFA alumni—were excluded from the sample, to maintain the study's focus on the effectiveness of the i3 scale-up.

Any non-TFA teacher teaching a class in the same school at the same grade level and covering the same subjects as a participating TFA teacher was potentially eligible to be a comparison teacher. This included both novice and experienced teachers; it also included traditionally certified teachers (those who completed a traditional university-based teacher certification program before they began teaching) and alternatively certified teachers (those who, like TFA teachers, began teaching before completing all requirements for certification).

Although the TFA teacher sample included only TFA teachers in their first or second years of teaching (current corps members at the time of the study), the comparison teacher sample included both novice and experienced teachers. Because the evaluation aimed to assess the short-term impact (as of the 2012–2013 school year) of TFA teachers recruited during the first two years of the scale-up, this provides a relevant comparison. If schools had not hired a TFA corps member in that year, students could have been taught by either a novice or experienced teacher from some other route to certification. Nonetheless, research has shown that experience is an important determinant of teacher effectiveness (Boyd et al. 2006; Kane et al. 2008; Papay and Kraft 2013). Although TFA asks its teachers to make only a two-year commitment to teaching, some corps members do continue beyond their two-year commitment. To the extent that TFA teachers' effectiveness increases with experience, our impact estimates may understate the longer-term impacts of TFA teachers recruited under the i3 scale-up because some may remain in teaching beyond their two-year commitment and become more effective with experience.

2. Eligible classes

Students in a given grade and school were randomly assigned between the classes of participating TFA and non-TFA teachers—we refer to the group of classes between which students were randomly assigned as a classroom match. A classroom match could contain one or more TFA teachers and one or more non-TFA teachers. All classes in a match must have been taught under similar circumstances—for instance, the classes taught by both the TFA and comparison teachers must have been in the same language (or combination of languages) to be included in a match. Of the 57 matches, 51 were taught in English and 6 were bilingual (Spanish/English) or for English language learners (ELL).

Most classes were self-contained, with a single lead teacher teaching both math and reading to the same class. However, in four classroom matches, instruction was departmentalized by subject, with different teachers for reading or math. In these cases, reading and math classes would form separate matches. Either or both subjects in a given grade and school could be included in a separate match as long as at least one class in that subject was taught by a TFA teacher and at least one was taught by a comparison teacher. Of the 57 matches, the TFA and comparison teachers taught only math in one classroom match and they taught only reading in three classroom matches—the rest of the matches included instruction in both reading and math.

Classes in prekindergarten through grade 5 were eligible for the study—any school with an eligible classroom match at this grade level was eligible for the study, regardless of the school's overall grade configuration. Eligible schools included traditional elementary schools (kindergarten through grade 5), charter schools, and community-based prekindergarten programs. As discussed in Chapter I, while the i3 scale-up expanded TFA placements at all grade levels, the impact analysis focuses only on teachers in prekindergarten through grade 5 because (1) the most rigorous experimental evidence at the elementary school level (Decker et al. 2004) is more than 10 years old, and there is more recent experimental evidence at the secondary level (Clark et al. 2013); (2) nonexperimental evidence has generally focused only on grades 4 and above; and (3) there is no previous rigorous evidence on the effectiveness of TFA teachers in prekindergarten and kindergarten, so including these grade levels allowed us to fill this gap in the

literature for reading.[5] Because we included only elementary school teachers, we do not draw conclusions about the effectiveness of secondary school TFA teachers, who made up 64 percent of all TFA teachers recruited during the first two years of the scale-up and were not eligible for inclusion in the study.

B. Recruitment of placement partners, schools, and teachers

We recruited sample members during the 2011–2012 school year to participate in the study the following school year. The final sample included 10 states, 13 school districts and other TFA placement partners, 36 schools, 57 classroom matches, and 156 teachers (Table II.1).[6] The study sample included all TFA and comparison teachers who taught matched classes, the students who were randomly assigned to those classes, and the schools and placement partners in which those classes were located. Appendix A provides details on the numbers of placement partners, schools, and potential classroom matches that were involved in each stage of recruitment.

Table II.1. Number of states, placement partners, schools, classroom matches, and teachers in the study

	Number of study units
States	10
TFA placement partners	13
Traditional public school districts	11
Charter schools and charter management organizations	1
Community-based organizations	1
Schools	36
Classroom matches	57
Classroom matches in reading analysis sample	56
Classroom matches in math analysis sample	32
Teachers (total)	156
TFA teachers	66
Comparison teachers	90
Teachers in math analysis sample	83
TFA teachers	34
Comparison teachers	49
Teachers in reading analysis sample	154
TFA teachers	65
Comparison teachers	89

Source: Mathematica evaluation tracking system.

Notes: A community-based organization is an early childhood education program that is not part of a district or charter school.

TFA = Teach For America.

[5] We are unable to estimate impacts on math achievement at these grade levels due to an error in administration of the math assessment for students in prekindergarten and kindergarten.

[6] TFA's placement partners include traditional public school districts, charter schools or charter management organizations, and community-based organizations that run prekindergarten programs.

1. Recruitment of districts and other placement partners

We focused our recruitment efforts on districts and other TFA placement partners with large concentrations of elementary teachers from TFA. Using fall 2011 teacher placement data from TFA, we identified placement partners with the largest numbers of TFA elementary school teachers, and we contacted 70 of them prior to the study year. In 28 of those 70 placement partners, we contacted schools directly to explore eligibility. We conducted random assignment in schools within 15 placement partners, and 13 ultimately remained in the study.[7] As expected, given our focus on placement partners with large numbers of elementary school TFA placements, the 13 placement partners in the study tended to have more elementary school placements than the typical TFA placement partner (with an average of 50 elementary school placements in study placement partners, compared with an average of 8 across all placement partners).

2. Recruitment of schools

Any school with an eligible classroom match in the 2012–2013 school year was eligible for the study. Within placement partners that allowed us to contact their schools directly, we contacted schools in the spring prior to the study year to identify those that were likely to have an eligible match in the upcoming year. We prioritized contacting schools with first-year TFA corps members in the 2011–2012 school year because these corps members were likely to be eligible for our study the following school year. Although placements of incoming corps members for the 2012–2013 school year were not all known at the time we conducted recruitment, we found that many schools with corps members in the 2011–2012 school year were also planning to hire new corps members for the study school year.

We also placed priority on contacting schools with potential matches in prekindergarten and kindergarten in an effort to oversample matches at these grade levels. Prior to this study, there was no experimental evidence on the effectiveness of TFA teachers in prekindergarten and kindergarten, and oversampling allowed us to obtain more precise impact estimates for teachers in these grade levels.

In each school, we gathered information about the school structure and teaching assignments to determine whether the school was likely to have any eligible classroom matches in the following school year. For example, we obtained data on the number of teachers per grade and whether students were grouped in any way that would prevent random assignment. Of the 313 schools we initially contacted, the final sample of 36 schools consisted of those that had eligible classroom matches, agreed to allow random assignment of students, and provided verification that students had been placed into classes in accordance with the results of the random assignment.

Even though study schools were not randomly selected from the full set of elementary schools employing TFA teachers nationwide, the study schools were similar to elementary schools employing TFA teachers nationwide along many dimensions (Table II.2). Both sets of schools served predominantly students from racial and ethnic minority groups. Less than

[7] We dropped two placement partners from the study sample because the schools in those placement partners failed to implement random assignment.

8 percent of students at both the average study school and the average elementary school with TFA teachers nationwide were white, non-Hispanic; about one-half of students at both types of schools were black, non-Hispanic; and more than one-third were Hispanic. About 80 percent of students at both types of schools were eligible for free or reduced-price lunch. Consistent with TFA's mission to place its corps members in schools in low-income communities, schools in the study sample and schools employing TFA teachers nationwide were on average considerably more disadvantaged than the average elementary school nationwide.

There were also some differences between study schools and all TFA schools nationwide, some of which may have been due to our recruitment approach and study eligibility requirements. Because charter schools were typically smaller than average and therefore less likely to have eligible classroom matches, they were less likely to be included in the study. Although only about 3 percent of the study sample was made up of charter schools, almost 26 percent of TFA elementary schools nationwide were charter schools. There were also differences between groups in how the schools were distributed across regions of the United States. The majority of study schools (82 percent) were located in the South, whereas only about half of all TFA elementary schools nationwide were located in that region. TFA elementary schools in the Northeast and West were underrepresented in the study sample.

Table II.2. Characteristics of study schools with TFA teachers compared with all elementary schools with TFA teachers and all elementary schools nationwide

	Study schools with TFA teachers[a]	All elementary schools with TFA teachers[b]		All elementary schools nationwide[c]	
Characteristic	Mean	Mean	*p*-Value of difference from study schools	Mean	*p*-Value of difference from study schools
Racial/Ethnic distribution of students					
Percentage Asian, non-Hispanic	1.4	3.4	0.000**	4.1	0.000**
Percentage Black, non-Hispanic	48.1	51.4	0.544	15.4	0.000**
Percentage Hispanic	40.3	34.2	0.218	21.4	0.000**
Percentage White, non-Hispanic	7.9	7.9	0.975	54.5	0.000**
Percentage other race/ethnicity	2.4	3.1	0.291	4.6	0.000**
Student socioeconomic status					
Percentage eligible for free/reduced-price lunch	78.7	81.1	0.536	52.3	0.000**
Percentage Title I-eligible schools	96.7	97.5	0.787	80.1	0.000**
Enrollment and staffing					
Average total enrollment	560.0	569.7	0.842	451.5	0.000**
Average enrollment per grade	77.6	77.7	0.988	77.6	0.992
School type					
Percentage traditional public school[d]	97.1	74.0		94.1	
Percentage public charter school	2.9	26.0		5.9	

Characteristic	Study schools with TFA teachers[a] Mean	All elementary schools with TFA teachers[b] Mean	All elementary schools with TFA teachers[b] *p*-Value of difference from study schools	All elementary schools nationwide[c] Mean	All elementary schools nationwide[c] *p*-Value of difference from study schools
Chi-squared test of difference in distributions			0.000**		0.309
School location					
Percentage urban	88.2	75.6		27.5	
Percentage suburban	8.8	17.5		41.6	
Percentage rural	2.9	6.9		30.9	
Chi-squared test of difference in distributions			0.098		0.000**
Census Bureau region					
Percentage in Northeast	0.0	12.7		16.4	
Percentage in Midwest	14.7	17.3		25.8	
Percentage in South	82.4	50.8		33.9	
Percentage in West	2.9	19.2		23.9	
Chi-squared test of difference in distributions			0.000**		0.000**
Sample Size	**34**	**1,263**		**59,790**	

Source: TFA placement data; Common Core of Data, Public Elementary/Secondary School Universe Survey, 2011–2012.

[a]Estimates for study schools include only 34 schools. Comparable data are not available for the two early childhood programs in the sample.

[b]Estimates are based on public elementary or charter schools in which new TFA teachers were placed in the 2011–2012 and 2012–2013 school years. Comparable data are not available for early childhood programs run by community-based organizations.

[c]Estimates include all schools with at least one grade from prekindergarten to grade 5.

[d]Traditional public schools are non-charter schools.

**Difference between this group and study schools with TFA teachers (first column) is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

3. Classroom matches and teachers in the final study sample

The final set of 57 classroom matches in the study spanned all elementary grade levels from prekindergarten through grade 5. In 54 percent of the matches, there were two teachers—one TFA teacher and one comparison teacher. In the rest, there were additional teachers of one or both types (Appendix Table A.1). In total, there were 66 TFA teachers in the study sample—the math analysis sample included 34 TFA teachers and the reading analysis sample included 65. This sample was large enough to reliably detect effects on student reading achievement as small as 0.14 standard deviations. The sample size for the math analysis sample was smaller (due to the fact that we were unable to include matches in grades prekindergarten and kindergarten) but was large enough to detect impacts of 0.15 standard deviations, or about the size of TFA's effects on math achievement found by Decker et al. (2004), as discussed further in Section D of Appendix A.

Our sample differed from the full set of TFA teachers on several characteristics (Table II.3); many of the differences can be attributed to our recruitment strategy. First, because we targeted schools with TFA teachers in the school year prior to the study year, a lower percentage of study teachers were first-year corps members compared with TFA corps members nationally. Second, we deliberately recruited a large number of schools with potential matches in prekindergarten or kindergarten to allow for more precise estimation for this subgroup; this led to an overrepresentation of prekindergarten or kindergarten study teachers compared with all TFA elementary school teachers. Third, because charter schools were less likely to have eligible classroom matches, study teachers were far more concentrated in regular public schools than the group as a whole. Table II.3 documents other ways in which our sample of teachers was similar to or different from TFA teachers nationally.

To adjust for the underrepresentation of first-year corps members and overrepresentation of early childhood teachers in the sample, we created weights to rescale each classroom match such that each grade level and cohort represented the same percentage of the study sample as their percentage in the full population of TFA elementary corps members nationwide in the 2012–2013 school year. We did not adjust for the underrepresentation of charter school teachers; to have done so would have assigned undue weight to the single charter school match in the sample. The weights, discussed further in Appendix A, scale down the contribution to the impact estimates of grade-level and corps year groups that are overrepresented in the sample (early childhood teachers and second-year corps members) and scale up the contribution of groups that are underrepresented.

4. Representativeness of the study sample

Ideally, to estimate the effectiveness of all TFA teachers recruited under the i3 scale-up, we would have randomly sampled TFA teachers from the full set of all TFA teachers recruited over the full period of the i3 scale-up, included all their students in the study sample, and collected data on a wide array of outcomes these teachers could have affected. For a variety of reasons, related to the timeframe and resources available for the evaluation, requirements of the random assignment design, practical considerations for sample recruitment, and district requirements for study participation, this approach was not possible. The following features of the evaluation design and sample selection limit our ability to generalize findings to the full population of TFA teachers recruited under the scale-up or the full set of students taught by these teachers:

1. The analysis of reading test scores focuses only on teachers in prekindergarten through grade 5, who made up 36 percent of all TFA teachers recruited during the first two years of the scale-up. The analysis of math test scores is further limited to teachers in grades 1 through 5, who made up 29 percent of all TFA teachers recruited during the first two years of the scale-up.
2. Because the evaluation, including analysis and reporting, was to be completed within the i3 grant period, the study includes only the first two cohorts of TFA teachers recruited as part of the scale-up and does not include the third or fourth cohorts of teachers.
3. Because the evaluation only includes TFA teachers in their first or second year of teaching (also because of the timeframe available for the evaluation), impact estimates do not reflect the longer-term effectiveness of some TFA teachers recruited under the scale-up who may have chosen to remain in teaching beyond their two-year commitment.

Table II.3. Comparison of sample to TFA teachers nationally in 2012–2013 school year

Characteristic	TFA study teachers	All elementary TFA teachers	
	Percentage	Percentage	*p*-Value of difference from study teachers
Corps year			
2011 (second year in TFA)	63.6	43.1	
2012 (first year in TFA)	36.4	56.9	
Chi-squared test of difference in distributions			0.001**
Age (average years)[a]	23.5	23.9	0.027*
Female	90.8	82.4	0.019*
Race/ethnicity			
Asian, non-Hispanic	7.6	5.2	
Black, non-Hispanic	10.6	13.6	
Hispanic	6.1	10.5	
White, non-Hispanic	68.2	62.3	
Other, non-Hispanic	7.6	8.4	
Chi-squared test of difference in distributions			0.491
Received Pell Grant	36.4	34.3	0.723
College selectivity[b]			
Most selective	24.2	33.4	
More selective	50.0	40.6	
Selective	10.6	14.1	
Not selective or unranked	15.2	11.9	
Chi-squared test of differences in distributions			0.217
Grade level			
Prekindergarten–kindergarten	42.4	19.8	
Grades 1–2	36.4	36.0	
Grades 3–5	21.2	44.3	
Chi-squared test of difference in distributions			0.000**
School type			
Traditional public[c]	93.9	58.0	
Public charter	4.6	38.3	
Bureau of Indian Affairs	0.0	0.8	
Catholic	0.0	0.1	
Early childhood center	1.5	2.5	
Private	0.0	0.4	
Chi-squared test of difference in distributions			0.000**
Number of teachers	**66**	**7,325**	

Sources: Study data from the Mathematica evaluation tracking system; national data from TFA admissions and placement data.

[a]Age is calculated as of September 1, 2012.

[b]TFA defines selective colleges as those ranked by *U.S. News & World Report* as "selective," "more selective," or "most selective." Information on selectivity is only collected for schools from which TFA has received five or more applications in any year between 2010 and 2013. In addition, TFA no longer uses these selectivity data internally, so there are many colleges that are classified as unranked.

[c]Traditional public schools are noncharter schools.

*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

4. Only reading and math achievement are included in the analysis because of a lack of available test score data in other subjects. Thus, the impact estimates do not reflect student performance in other domains in which the TFA teacher may differentially affect student achievement.
5. The experimental design necessarily limited the sample to TFA teachers for whom this design was feasible—those teaching in a classroom match opposite a non-TFA teacher—and may have led to an underrepresentation of particular types of schools where the study was less likely to be feasible. For instance, as discussed above, charter schools were less likely to have eligible classroom matches and are underrepresented in the sample. TFA teachers' impacts may have differed in schools that did not have eligible matches.
6. As discussed above, particular features of our recruiting approach led to an overrepresentation of teachers in prekindergarten and kindergarten and of second-year corps members in our sample. Even with our use of sample weights to scale down the contribution of these groups to our impact estimates, findings do not generalize to the full population of TFA teachers, but reflect the effectiveness of the particular teachers in our sample when the sample is weighted to more closely resemble the national population of elementary TFA teachers in terms of grade level and corps year.
7. Only 10 of TFA's 49 regions were included in the sample. TFA teachers' impacts may have differed in other regions.
8. Participation was voluntary. The sample of schools included only those that agreed to participate. In addition, as described in Section C, we only had test score data for students whose parents consented for them to participate in the study and who were available on the day of testing.

For all these reasons, the evaluation provides evidence on the effectiveness of a particular set of TFA teachers recruited under the i3 scale-up, for a particular set of students in particular subjects, rather than for the full set of TFA teachers recruited under the i3 scale-up or the full set of students taught by these teachers.

C. Selection and assignment of students

We randomly assigned students to classes to ensure that similar sets of students were assigned to TFA and comparison teachers within each classroom match at the start of the school year. Before the start of the study school year, schools sent us lists of students to be enrolled in the identified classroom matches. We randomly assigned the students to the classes, specifying the teacher for each class. The schools then placed students in classes in accordance with the random assignment results. We also randomly assigned students who needed to enter one of the classes after this initial assignment but before the end of the first two weeks of the school year; schools called a study hotline to request assignments for these late enrolling students. On a limited basis, schools could explicitly request a specific assignment for a given student, in which case the student was excluded from the study. We did not randomly assign students who enrolled after the first two weeks of school, and we excluded these students from the study. If a school refused to implement the random assignments for a given match or if the composition of the classes changed after school staff implemented the random assignments (for instance, the classes were departmentalized, with separate teachers for math and reading) and the school did not allow

us to redo random assignment, then that match was dropped from the study. We provide additional details on the random assignment process in Appendix A.

By the end of the first two weeks of the school year, we had randomly assigned 3,590 students in matches that included math and 3,679 students in matches that included reading (Table II.4). We attempted to obtain test score data for all randomly assigned students, and we include all randomly assigned students with valid end-of-year reading or math test scores in the impact analysis. The math analysis includes 1,182 students and the reading analysis includes 2,123.

Overall attrition rates were high, particularly for the math sample, but the attrition rates of students from the TFA and comparison groups were similar, reducing concern about selective attrition that might have compromised the randomized design. In math, the overall attrition rate was around 68 percent, and in reading it was around 42 percent. (Table II.4). The main reason we lacked test score data for students in both the reading and math samples was that their families did not consent to their participation in the study.[8] In addition, as described in greater detail in Appendix A, due to the error in administration of the math assessment for students in prekindergarten and kindergarten, we lacked valid math scores for these students. Within each subject, attrition rates were similar across TFA and comparison groups. We provide additional details on sample attrition in Appendix A.

Table II.4. Attrition from the student sample

	Number of students			Attrition rate		
	Assigned to TFA teachers	Assigned to comparison teachers	Total	Assigned to TFA teachers	Assigned to comparison teachers	Total
Math						
Randomly assigned	1,476	2,114	3,590			
Randomly assigned and had valid test score data	477	705	1,182	67.7%	66.7%	67.1%
Reading						
Randomly assigned	1,521	2,158	3,679			
Randomly assigned and had valid test score data	877	1,246	2,123	42.3%	42.3%	42.3%

Source: Mathematica evaluation tracking system.

TFA = Teach For America.

Among students included in the analysis for either reading or math, characteristics are similar between those assigned to TFA teachers and those assigned to comparison teachers (Table II.5). This suggests that random assignment was properly implemented and that student attrition due to lack of end-of-year tests did not lead to differences in baseline characteristics between the two groups. Those assigned to TFA teachers and those assigned to comparison teachers were statistically similar in terms of baseline characteristics. Only one difference was statistically significant at the 5 percent level or below: students assigned to comparison teachers

[8] Although parental consent for study participation was not required by federal law, many school districts required us to obtain written consent from parents for students to participate.

were more likely to be Asian than were students assigned to the TFA teachers. Because we examined multiple characteristics, it is possible that this single case of a statistically significant difference was the product of chance differences in the two samples.

Consistent with TFA's goal of serving disadvantaged students, the students in the study tended to have low baseline achievement, be from low-income families, and be members of racial and ethnic minority groups. Among students for whom we have baseline test score data, on average they scored below the mean on their state tests in math (average z-score of -0.05) and reading (average z-score of -0.21) in the year prior to the evaluation. These scores indicate that the average sample member with baseline scores would rank at about the 48th percentile in math relative to other students in the same state and grade, and at about the 42nd percentile in reading. The majority of students (84 percent) were eligible for free or reduced-price lunch. About 47 percent of students were black, and 42 percent were Hispanic. About one-third of students had limited English proficiently and 7 percent had an individualized education plan (IEP) for a special education program or services. Compared with national averages, fewer students in the sample had IEPs, but more students were black, Hispanic, and eligible for free or reduced-price lunch and had limited English proficiency.[9] Compared with students in the 2004 study of elementary teachers (Decker et al. 2004), the students in this study were higher achieving and less likely to be from low-income families. The students in the 2004 study ranked at about the 14th percentile in math and the 13th percentile in reading, and 95 percent were eligible for free or reduced-price lunch.

Table II.5. Average baseline characteristics of students in the math or reading analysis who were assigned to TFA teachers or comparison teachers (percentages unless otherwise indicated)

Characteristic	Analysis sample	Assigned to TFA teachers	Assigned to comparison teachers	Difference between TFA and comparison	*p*-Value
Baseline math score (average z-score)[a]	-0.05	-0.14	0.04	-0.18	0.357
Baseline reading score (average z-score)[a]	-0.21	-0.21	-0.21	0.00	0.985
Female	47.2	47.2	47.2	0.1	0.981
Race and ethnicity					
Asian, non-Hispanic	1.7	0.9	2.5	-1.5	0.006**
Black, non-Hispanic	46.6	47.0	46.1	0.8	0.605
Hispanic	41.7	42.5	40.9	1.6	0.388
White, non-Hispanic	7.3	7.4	7.1	0.2	0.846
Other, non-Hispanic	2.8	2.2	3.3	-1.1	0.142
Eligible for free/reduced-price lunch	83.7	84.5	82.9	1.6	0.270
Limited English proficiency	33.7	33.2	34.1	-0.8	0.634
Individualized education plan	6.9	7.8	6.0	1.8	0.146
Number of students	**2,152**	**895**	**1,257**		

[9] We compared statistics for students in the sample to statistics taken from the 2011–2012 Schools and Staffing Survey (Goldring et al. 2013) at the elementary level for a sample of schools and the 2011–2012 Common Core of Data at the district level for all schools.

Characteristic	Analysis sample	Assigned to TFA teachers	Assigned to comparison teachers	Difference between TFA and comparison	*p*-Value
Number of teachers	**156**	**66**	**90**		
Number of classroom matches	**57**	**57**	**57**		
Number of schools	**36**	**36**	**36**		

Source: District administrative records.

Note: Means and percentages are weighted with sample weights and adjusted for classroom match fixed effects; *p*-values are based on a regression of the specified characteristic on a TFA indicator and classroom match indicators, accounting for sample weights.

[a]Baseline test scores were only available for students in grades 4 and 5. In the math analysis, 143 students had baseline test scores, as did 199 of the students in the reading analysis.

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

TFA = Teach For America.

We also examined the proportion of students in each study class at the beginning and end of the school year who were not randomly assigned (either because schools requested an exemption for particular students or because students enrolled after the random assignment period). Even though randomly assigned students were similar at baseline, the composition of their non-randomly assigned peers could potentially affect the achievement of students in particular classes. Of the students who enrolled in a study class before or during the first two weeks of school, 97 percent were randomly assigned (Table II.6). Rates were similar in the classes of TFA teachers and the classes of comparison teachers, differing by just one percentage point. The remaining students were exempted from random assignment at the school's request.

Table II.6. Changes in composition of study classes during the school year

	Average number of students per teacher (unless otherwise indicated)		
	All study classes	Classes of TFA teachers	Classes of comparison teachers
Enrolled in study classes before the end of the first two weeks of school			
Number of students	20.6	20.5	20.7
Number of students who were randomly assigned	20.0	20.0	19.9
Percentage of students who were randomly assigned	96.7	97.3	96.2
Listed on end-of-year class rosters			
Number of students	21.8	21.5	22.0
Number of students who were randomly assigned and stayed in originally assigned class	16.0	16.0	16.1
Percentage of students who were randomly assigned and stayed in originally assigned class	73.6	74.5	72.9
Number of teachers	**156**	**66**	**90**

Source: Mathematica evaluation tracking system.

Note: Table excludes students who were randomly assigned before the start of the school year but never attended a study school.

There was some student movement into and out of the study classes after the random assignment period. Some students transferred out of their originally assigned classes and some late-enrolling students were placed by schools into study classes after the first two weeks of the school year. Despite this mobility, study classes remained primarily composed of research sample members throughout the year. On end-of-year class rosters, 74 percent of students in study classes had been randomly assigned to those classes originally, with similar rates in the classes of TFA teachers and the classes of comparison teachers (Table II.6).

D. Attrition of teachers from the sample

Of 156 teachers in the initial sample, 9 left after the school year began (Table II.7). Three TFA teachers left; in two cases, they were replaced by TFA teachers and in the other case by a non-TFA teacher. Six comparison teachers left, one of whom was replaced by a TFA teacher and the rest of whom were replaced by non-TFA teachers. Most of the departing teachers left in the spring semester, with just one TFA and one non-TFA teacher departing in the fall semester.

Table II.7. Teacher turnover

	Number of TFA teachers	Number of comparison teachers
Start of school year	66	90
Stayed through end of school year	63	84
Replaced by teacher of same type	2	5
Replaced by teacher of opposite type	1	1

Source: Mathematica evaluation tracking system.

TFA = Teach For America.

For the study, we retained all of these classroom matches, including all students in the group (TFA or comparison) to which they were initially assigned, even in the one case in which a TFA teacher was replaced by a non-TFA teacher, and the one case in which a non-TFA teacher replaced a TFA teacher. We considered the turnover of these nine teachers to be part of the "TFA effect." In other words, the risks associated with having to replace a TFA or non-TFA teacher with a backup teacher were incorporated into our measure of the relative effectiveness of TFA teachers compared with teachers from other routes. However, we examine the sensitivity of our results to this decision in Appendix B.

E. Data used in the study

We collected data from a variety of sources, listed in Table II.8.

1. Data on students

We attempted to collect data on reading and math achievement and demographic characteristics for all randomly assigned students for whom we received parental consent to collect these data.

Student achievement outcome data. To measure student achievement outcomes, we collected end-of-year reading and math test scores from the 2012–2013 school year for all randomly assigned students with parental consent. However, due to an error in administration of

the math assessment for students in prekindergarten and kindergarten, we were unable to include these scores in the math analysis. In the lower elementary grades (prekindergarten through 2 for reading and grades 1 and 2 for math), we assessed students using reading and math assessments from the Woodcock-Johnson III. This test can be administered in either English or Spanish and has a reliability for student ages 6 to 9 of over 0.90 for the reading tests and greater than 0.85 for the math test that we used (McGrew et al. 2007). In the upper elementary grades (3 to 5), in which annual reading and math assessments were required by the federal No Child Left Behind Act, we collected state assessment data from district records. We also collected prior years' test scores from state assessments where available.

Table II.8. Data sources for the evaluation

Domain	Data source	Schedule of data collection
Math achievement		
Grades 1–2	Study-administered Woodcock-Johnson III Normative Update Tests of Achievement, Calculation subtest	Spring 2013
Grades 3–5	District administrative records	Summer/fall 2013
Reading achievement		
Prekindergarten–grade 2	Study-administered Woodcock-Johnson III Normative Update Tests of Achievement, Letter-Word Identification subtest (prekindergarten–grade 2) and Passage Comprehension subtest (kindergarten–grade 2)	Spring 2013
Grades 3–5	District administrative records	Summer/fall 2013
Baseline student characteristics	District administrative records	Summer/fall 2013
Baseline student achievement	District administrative records (grades 4–5 only)	Summer/fall 2013
Student mobility	Class rosters	Summer 2012, fall 2012, winter 2013, spring 2013
Teachers' route to certification	Teacher background form	Summer/fall 2012
Teachers' characteristics, attitudes, and practices	Teacher survey	Spring 2013
School characteristics	Common Core of Data	Spring 2014
TFA program characteristics and scale-up implementation	Program administrator interviews	Summer 2011–winter 2012, summer 2012–winter 2013
	TFA program data and internal survey data	Spring 2012–spring 2014

TFA = Teach For America.

- **Outcome test scores for students in lower elementary grades.** To assess the achievement of students in lower elementary grades, we administered a series of tests from the Woodcock-Johnson III Normative Update Tests of Achievement in the spring of the 2012–2013 school year. Students took Woodcock-Johnson tests that were appropriate for their grade level. In math, students in grades 1 and 2 took the Calculation subtest (as well as the Applied Problems subtest, for which scores were dropped due to an error in test administration). In reading, students in prekindergarten through grade 2 took the Letter-Word Identification subtest, and all but those in prekindergarten took the Passage Comprehension subtest. We provide details on how we assessed students using the Woodcock-Johnson test in Section G of Appendix A.

- **Outcome test scores for students in upper elementary grades.** To measure the achievement of upper elementary students, we used scores from state reading and math assessments. We obtained these data from district records. Because these annual assessments are used to track student progress, we expected them to be closely aligned with course content and to measure accurately the math and reading skills teachers had covered during the school year. Students typically took assessments in English, although in a few bilingual and ELL classes, students took the test in Spanish. As long as most students in the classroom match took the test in the same language, we used the test scores of the students who took the test in the language of the majority of students in that classroom match and excluded the test scores of students taking the test in the other language. For example, if 40 students in a classroom match took the test in English, and 2 students took the test in Spanish, we would use the test scores from the 40 English-language tests and drop the 2 test scores from the Spanish-language tests. This ensured that the tests taken by students in both TFA and comparison classes were comparable.

Student baseline characteristics. To improve the precision of the impact estimates, we collected data on student baseline characteristics from district or school records. Where available, we collected students' scores from state reading and math assessments in the school year prior to the impact evaluation (2011–2012). These data were only available for students in grades 4 and 5 (who were in grades 3 and 4 in the previous school year). In addition to baseline test scores, we collected information on student demographic characteristics, including date of birth, grade, gender, race/ethnicity, free or reduced–price lunch eligibility, special education status or whether the student had an IEP, and whether the student had limited English proficiency.

2. Data on teachers

TFA status. Before the study year, we verified the certification route (TFA, some other alternative route, or traditional route) of all teachers whose classes could potentially be included in classroom matches by asking the teachers (or school administrators if the teachers were unavailable) to complete a brief form with this information.

Professional background and experiences. In the spring of the study year, we administered a survey to teachers in the study to collect information on their professional background and experiences. The survey asked about teachers' educational background, teaching experience, preparation for teaching, support received during the school year, views toward teaching, and demographic characteristics.

3. Data on schools

Data on schools provided important contextual information for the evaluation, allowing us to compare the characteristics of schools in the sample to all elementary schools in which corps members were placed in the study school year and all elementary schools nationwide. Using the Common Core of Data, a comprehensive database of the universe of public schools in the United States, we assembled data on school characteristics, including grade span, enrollment, percentage of students eligible for free or reduced-price lunch, and the racial/ethnic distribution of the student body.

4. Data on TFA

To describe TFA's program and its implementation of the i3 scale-up, we used both qualitative and quantitative data. We conducted semi-structured interviews with 17 members of TFA's senior staff following the first and second years of the scale-up. TFA provided data on corps member admissions, placement, training, and support provided to its corps members. It also provided data from internal surveys it administers to all its corps members. To track the implementation of scale-up activities, we collected information on broad organizational plans and data from key program areas (recruitment, selection, training and support, and placement).

F. Overview of analytic approach

We estimated the causal effect of TFA teachers on elementary student reading and math achievement based on the experimental design. Because students in the study were randomly assigned to teachers, we attribute any differences in achievement at the end of the study school year to the relative effectiveness of TFA teachers and comparison teachers rather than to the types of students taught by these two different groups of teachers.

Outcome measures. The outcome measures for this study were student achievement in math and reading. Because tests at the upper elementary school level differed across state, grade level, and subject area and differed from the study-administered tests at the lower elementary level, we converted the original scale scores to z-scores (original scores minus the mean score divided by the standard deviation of the scores) in order to scale the outcome variable comparably across all classroom matches. For both the Woodcock-Johnson and state assessments, we computed z-scores using means and standard deviations from the broadest possible reference population. For upper elementary school students, we used published means and standard deviations for each test for all students in each state and grade. For lower elementary school students, in which all students took the same assessment, we separately converted broad reading *W* scores and broad math *W* scores to z-scores using the means and standard deviations for each subject and age group provided by the test publisher.

Estimation method. We estimated the effectiveness of TFA corps members relative to comparison teachers using a regression model. Because teachers in the same classroom matches were assigned similar students at the beginning of the year, we could have estimated the effectiveness of TFA corps members by subtracting the average test scores of the students of comparison teachers from the average test scores of students of TFA teachers. Instead, the regression approach built upon simple test score differences in two ways: (1) allowing comparisons to be made within the same classroom match and (2) enhancing the precision of the estimates by using information on student baseline characteristics to better predict their end-of-year achievement. We included indicators for each classroom match so that comparisons were made only within the same match. In the regression-based approach, the average effectiveness of TFA teachers was similar to a weighted average of the effectiveness of each TFA teacher relative to the comparison teacher(s) in each match. Matches with more students received more weight in the analysis. We accounted for student demographic information for all students. For students in districts and grades for which prior-year test score data were available (grades 4 and 5), we accounted for these prior-year test scores as well. We provide more details on the estimation method, including descriptions of the sensitivity analyses we conducted, in Appendix A.

Subgroup analyses. In addition to the main impact estimates, we estimated the impact of TFA teachers for several subgroups. For math, we estimated impacts for four subgroups: (1) lower elementary students (grades 1 and 2), (2) upper elementary students (grades 3 to 5), (3) TFA teachers compared with other teachers in their first two years of teaching, and (4) TFA teachers compared with traditionally certified comparison teachers. For reading, we estimated impacts for five subgroups: (1) early childhood student (prekindergarten and kindergarten), (2) lower elementary students (prekindergarten to grade 2), (3) upper elementary students (grades 3 to 5), (4) TFA teachers compared with other teachers in their first two years of teaching, and (5) TFA teachers compared with traditionally certified comparison teachers. We analyzed early childhood teachers as a subsample for reading because there are no previous studies of the effectiveness of TFA early childhood teachers. We also included these teachers as part of the lower elementary sample for reading to increase the statistical power for that subgroup.

This page has been left blank for double-sided copying.

III. TFA'S PROGRAM MODEL AND IMPLEMENTATION OF THE I3 SCALE-UP

In this chapter we describe TFA's program model and the extent to which TFA maintained core elements of its approach as it expanded during the first two years of the scale-up, to provide context for interpreting the study's impact estimates. We discuss five key components of TFA's approach: (1) recruiting applicants to the program; (2) selecting applicants; (3) providing those who are selected and join the program, known as corps members, with preservice training before they begin their first teaching job; (4) helping corps members find jobs in high-needs schools; and (5) providing ongoing training and support to corps members throughout their two-year commitment. More details about TFA's program model and its implementation of the i3 scale-up are provided in Zukiewicz et al. (2015).

A. Recruitment

TFA recruits undergraduate and graduate students at college campuses across the country as well as professionals. The program places a high priority on recruiting a racially and economically diverse set of corps members and on recruiting corps members to teach hard-to-staff areas such as science, math, and special education. More than 48,000 applicants applied to join the 2012 TFA corps, including more than 5 percent of the graduating senior class at 135 colleges and universities.

Undergraduate recruitment. During undergraduate recruitment, recruitment teams conduct outreach on college campuses, meeting with prospective applicants both in person and online. The teams seek to raise student awareness of the program through the use of media campaigns, on-campus presentations, and partnerships with student organizations. Typically, the teams work with undergraduate "campus campaign coordinators," students working as part-time TFA employees who help TFA conduct publicity campaigns and identify potential applicants on their campuses. The recruitment teams also learn about promising candidates from interested students themselves and via referrals from university alumni, professors, and administrators. They then target recruitment efforts to the individuals they believe are best qualified for the program, contacting promising candidates to discuss the program, answer their questions, and encourage them to apply.

As a part of its expansion effort under the i3 scale-up, TFA increased recruitment among less selective colleges, with the understanding that highly qualified individuals, particularly those from low-income backgrounds, often attend less selective schools that are closer to their homes because of economic constraints. Between the year prior to the scale-up and the second year of the scale-up, TFA expanded its outreach from 370 to 573 campuses, with the largest increases at schools in the second and third tiers of selectivity (those ranked "more selective" and "selective") as well as those that were not ranked by *U.S. News & World Report* (Table III.1).[10] TFA staff said that although the recruitment of students at these lower-ranked schools increased under this new recruitment strategy, the organization did not modify or reduce its applicant standards, such as grade point average or leadership experience. Instead, recruitment teams

[10] TFA recruitment staff said they no longer use the selectivity data internally, so there are many colleges that are classified as unranked.

expanding to new, less selective campuses sought to recruit the top students that they believed would meet the program's qualifications.

Table III.1. Number of colleges in which TFA recruited before and during the i3 scale-up

	Pre-scale-up cohort	First two scale-up cohorts	
	2009–2010 academic year	2010–2011 academic year	2011–2012 academic year
	Recruitment for entering TFA cohort 2010–2011	Recruitment for entering TFA cohort 2011–2012	Recruitment for entering TFA cohort 2012–2013
Selectivity of colleges[a]			
Most selective	66	66	67
More selective	182	186	214
Selective	73	75	109
Less selective	36	33	44
Least selective	2	2	2
Unranked	11	4	137
Type of college			
Historically black colleges and universities	25	25	38
Hispanic Association of Colleges and Universities	30	30	41
All universities	**370**	**370**	**573**

Source: TFA recruiting data.

[a]Based on *U.S. News & World Report* college rankings. Information on selectivity is only collected for schools from which TFA has received five or more applications in any year between 2010 and 2013. In addition, TFA no longer uses these selectivity data internally, so there are many colleges that are classified as unranked.

TFA = Teach For America.

Recruitment of professionals and graduate students. In recent years, TFA has increased its recruitment of graduate students and professionals with experience in the corporate or nonprofit sectors. A centralized team of recruitment staff conducts most professional recruitment across the country. Responsibility for recruiting graduate students is shared by this centralized team and the on-campus recruitment teams. Most communication with graduates and professionals is by telephone or online, and most meetings are conducted via webinar or video call. Among incoming corps members in fall 2012, 17 percent had post-college professional experience and 6 percent were graduate students immediately prior to entering the corps.

Corps member diversity. TFA places a high priority on recruiting racial and ethnic minorities and corps members from low-income backgrounds. In an effort to increase corps member diversity, TFA recruitment teams partner with both campus-based and national organizations that serve racial and ethnic minorities on college campuses. TFA also places special emphasis on recruiting students from historically black colleges and universities, the Hispanic Association of Colleges and Universities, and public university systems known for their racial and ethnic diversity. They expanded recruitment from 25 to 38 historically black colleges and universities and from 30 to 41 schools in the Hispanic Association of Colleges and Universities between the year prior to the scale-up and the second year of the scale-up (Table III.1). Recruiters also target applicants from low-income backgrounds by recruiting

candidates who attended programs that serve low-income communities such as Posse, Prep for Prep, INROADS, KIPP charter schools, and Summer Search.

B. Selection

TFA relies on an intensive, data-driven admissions process to select the candidates who it predicts are most likely to succeed in the classroom. The process includes four stages: an online application; a web-based writing activity; a phone interview (which the most promising applicants are allowed to bypass); and a day-long, in-person interview that includes a one-on-one interview, a sample teaching lesson, and group discussions. At each stage of the admissions process, TFA prioritizes the selection of candidates with the following attributes:

- Commitment to reducing educational inequality
- Demonstrated leadership ability and interpersonal skills to motivate others
- Achievement in academic, professional, extracurricular, and/or volunteer settings
- Perseverance in the face of challenges, ability to adapt to changing environments, and a strong desire to improve and develop
- Critical thinking skills, including the ability to accurately link cause and effect and to generate relevant solutions to problems
- Organizational ability, including planning well and managing responsibilities effectively
- Respect for and ability to work with individuals from diverse background and experiences

At each stage of the selection process, the TFA selection committee considers the opinion and judgment of TFA staff who have either reviewed the application or spoken with the applicant to determine whether a candidate will move onto the next round. In addition, at each stage of the process, TFA staff use a mathematical selection model that helps guide decisions about whether applicants will progress to the next stage. This model, which TFA updates annually, uses recruitment, selection, and student achievement data from previous cohorts of corps members to determine the factors associated with corps member effectiveness and then uses these factors to predict the effectiveness of each new applicant. For components of the selection process that are qualitative in nature, such as observations of sample lessons given by candidates during the final round of interviews, TFA staff use scoring rubrics to rate candidate performance, and those quantified values are also entered into the selection model. Approximately 17 percent of applicants for the 2012 corps were selected into the program, and of these, 71 percent accepted the offer of admission.

In the first two years of the scale-up, the period covered by this evaluation, TFA fell just short of the growth goals it laid out in its i3 application. In 2011, the first year of the scale-up, it placed 5,031 new teachers (a 12 percent increase from the prior year, and just below its target of 5,300). In 2012, the second year of the scale-up, TFA placed 5,807 new teachers (a 15 percent increase from the first year, and short of its target of 6,000). More recent data for the final years of the scale-up show that TFA's growth slowed and it failed to meet its targets for those years

(Mead et al. 2015).[11] Nonetheless, over the first two years of the scale-up, the focal period for this evaluation, TFA expanded the number of first and second year corps members by 25 percent.

To provide evidence on whether TFA maintained its selection standards as it increased the size of its corps, we compared data on the characteristics of admitted corps members from the first two years of the scale-up and the two years prior. There were few apparent changes in the corps member characteristics we examined over this period (Table III.2). In the first two years of the scale-up, as in the two prior years, 90 percent or more of selected corps members held a bachelor's degree from a "selective," "more selective," or "most selective" college as ranked by *U.S. News & World Report.* More than one-third of corps members held a bachelor's degree from "most selective" colleges across those four years. Consistent with TFA's planned expansion of recruitment efforts to lower ranked colleges, there was a slight increase in the proportion of admitted corps members from colleges ranked "selective," "not selective," or unranked and a slight decrease in the proportion from those ranked "most selective" and "more

Table III.2. Accepted applicants to TFA program during the first two years of the TFA i3 scale-up

	Pre-scale-up cohorts		First two scale-up cohorts	
	Entering TFA cohort 2009–2010	Entering TFA cohort 2010–2011	Entering TFA cohort 2011–2012	Entering TFA cohort 2012–2013
Percentage of applicants accepted	15.8	14.7	14.8	17.0
Percentage of accepted applicants who join TFA	75.4	74.2	73.9	71.2
Academic background				
College selectivity[a]				
Most selective	39.8	38.6	38.9	36.1
More selective	43.1	41.2	41.1	40.5
Selective	10.2	11.7	10.9	13.4
Not selective or unranked	6.8	8.5	9.0	10.0
Average undergraduate GPA	3.6	3.6	3.6	3.6
Average SAT score	1,325	1,314	1,327	1,319
Demographic characteristics				
Percentage from racial or ethnic minorities	30.0	33.5	34.5	36.5
Percentage from disadvantaged background[b]	24.2	26.9	30.3	33.9
Overall sample size	**5,349**	**6,022**	**6,802**	**8,185**

Source: TFA admissions data.

[a]Selective colleges include colleges ranked by *U.S. News & World Report* as "selective," "more selective," or "most selective." Information on selectivity is only collected for schools from which TFA has received five or more applications in any year between 2010 and 2013. In addition, TFA no longer uses these selectivity data internally, so there are many colleges that are classified as unranked.

[b]Percentage from disadvantaged backgrounds measured by Pell Grant receipt.

TFA = Teach For America.

[11] According to Mead et al. (2015), TFA placed 5,400 new corps members in 2014, well below its goal of 7,500. That study, which is based on analysis of data and documents from TFA and interviews with current and former TFA staff, concludes that both improving economic conditions that increased employment options for graduating college students and external criticisms of TFA may have contributed to TFA's inability to meet its growth targets for the final years of the scale-up.

selective" over this period. The average undergraduate grade point average of new corps members remained constant at 3.6 over all four years, and the average combined math and verbal SAT score remained relatively constant, ranging from 1,314 to 1,327 over this period. Consistent with its efforts to expand recruitment of racial and ethnic minorities and candidates from low-income backgrounds, TFA increased the diversity of its corps over this period—the percentage of corps members from racial or ethnic minorities increased from 30 to 37 percent, and the percentage from a disadvantaged background (measured by Pell Grant receipt) increased from 24 to 34 percent.

C. Preservice training

Once corps members are accepted into TFA, they are required to participate in a series of preservice training activities, the main component of which is a five-week, full-time residential summer program known as summer institute. Prior to summer institute, corps members are asked to complete a series of independent study activities and attend a regional induction session. Following summer institute, they are asked to attend a post-institute training located in the region in which they will teach.[12] TFA officials estimated that corps members were assigned between 299 and 311 hours of preservice work in 2012.

Pre-institute work. Prior to beginning the summer institute program, all new corps members must complete a series of activities designed to serve as an introduction to TFA's overall approach and the Teaching As Leadership Rubric, a framework that guides all TFA training activities offered prior to and during a corps member's two-year commitment.[13] Corps members are asked to complete a set of eight required activities as part of their independent study, including reading curriculum texts, watching video clips of classroom instruction, and providing written responses to preservice materials. They must also conduct two in-person observations of a veteran teacher and respond to a series of questions regarding the teacher observations they conducted. According to TFA staff, required pre-institute activities in 2012 totaled 42.5 to 46.5 hours, depending on the grade level the corps member would be teaching.

Regional induction. Before summer institute, corps members attend an induction program in the region where they will teach. Induction serves to introduce corps members to the curricula and policies specific to the region where they will teach and to familiarize them with TFA's mission. Several regions also offer optional small-group orientation sessions. During the first two years of the i3 scale-up, TFA granted its regions greater autonomy to tailor the content and length of regional inductions to the schools and districts where corps members in that region would teach. Therefore, the content and length of the inductions varied across regions, but in 2012 they typically required 16 to 24 hours (two to three days) of training.

[12] A TFA region is a geographic cluster of school districts, charter schools, and community-based early childhood programs. It may contain a single large urban district; a small number of geographically clustered mid-sized districts; or a large number of small, geographically clustered rural districts.

[13] The Teaching as Leadership Rubric is a framework of six principles and 28 discrete teacher actions that TFA believes to be the roadmap to effective teaching. The six principles are (1) set big goals, (2) invest students and their families in working hard to reach the big goal, (3) plan purposefully, (4) execute effectively, (5) continuously increase effectiveness, and (6) work relentlessly.

Summer institute. As the main component of its preservice training, TFA provides corps members with a five-week training during the summer institute program. TFA typically holds summer institute programs on university campuses and runs summer school programs in partnership with local school districts. In 2012, corps members attended summer institutes in nine locations, including Atlanta, Chicago, Houston, Los Angeles, the Mississippi Delta, New York, Philadelphia, Phoenix, and Tulsa. Summer institute includes the following activities:

- Receiving group instruction on curriculum, literacy, and diversity
- Teaching summer school students under the supervision of experienced teachers
- Observing other teachers
- Receiving written and oral feedback on teaching from advisors
- Attending small-group sessions to reflect on teaching practice
- Participating in clinics designed to improve lesson-planning skills

According to TFA staff, required summer institute activities in 2012 totaled at least 240 hours, with some variation by institute and the subject and grade level the corps member would be teaching.

There were a few changes in the preservice training TFA provided to corps members in the first two years of the scale-up relative to the two previous years that we were able to discern in data provided by TFA (Table III.3). For instance, the number of hours of curriculum and literacy sessions assigned during summer institute decreased from 60 in 2009 (two years prior to the scale-up) to 52 in 2012 (the second year of the scale-up). The percentage of corps members conducting student teaching in the subject of their future placement increased from 56 to 64 percent between 2009 and 2012, whereas the percentage teaching in the grade of their future placement decreased from 52 to 44 percent between 2009 and 2011 but then increased back to 54 percent in 2012.

There were also some changes in corps members' perceptions of preservice training, as measured by an internal survey TFA conducts with its new corps members after each summer institute.[14] In all four years examined, almost 75 percent of corps members agreed or strongly agreed with the statement that "within TFA I feel part of a community where corps members help each other increase collective impact" immediately following summer institute. However, the percentage who felt that the summer institute was critical for being an effective teacher fell from 85 to 75 percent from 2009 to 2012, and the percentage reporting positive or very positive overall satisfaction with TFA at the end of their preservice training fell from 69 to 61 percent over this same period.

[14] TFA attempted to survey all corps members who attended summer institute and achieved a response rate of at least 97 percent across all years in the analysis, from 2009–2010 to 2012–2013 school years.

Table III.3. Corps member preservice training during the first two years of scale-up (percentages unless otherwise indicated)

	Pre-scale-up cohorts		First two scale-up cohorts	
	Entering cohort 2009–2010	Entering cohort 2010–2011	Entering cohort 2011–2012	Entering cohort 2012–2013
Summer Institute training sessions				
Hours of curriculum and literacy sessions assigned[a]	60	63	63	52
Hours of corps member advisor-led sessions assigned[a]	38	36	36	33
Student teaching placement				
Taught in subject of future placement	56	53	56	64
Taught in grade level of future placement	52	54	44	54
Perceptions of preservice training				
Agreed or strongly agreed that "within TFA I feel part of a community where corps members help each other increase collective impact"	77.1	78.7	75.5	74.6
Agreed or strongly agreed that summer institute was critical in efforts to become a successful teacher	84.7	83.8	82.0	74.8
Positive or very positive overall satisfaction with TFA	69.3	71.7	65.9	60.7
Sample size	**3,919**	**4,449**	**5,003**	**5,850**

Source: TFA preservice training data and end-of-institute surveys.

[a]Based on number of hours assigned on the national level. Hours may vary by institute.

TFA = Teach For America.

D. Placement

Consistent with its goal of placing corps members in high-needs schools, TFA partners with local education agencies (LEAs) comprising low-income, high-needs schools, as measured by the percentage of student who qualify for free or reduced-price lunch.[15] Partner LEAs include public school districts, public charter schools, and community-based organizations (for prekindergarten placements). In 2012–2013, nearly two-thirds of first-year corps members (65 percent) taught in traditional public schools and about one-third (33 percent) taught in charter schools (Table III.4). Consistent with its goals for the i3 scale-up, TFA expanded from 40 regions in 2010–2011 to 43 regions in 2011–2012 (the first year of the scale-up) and to 47 regions in 2012–2013.

TFA assigns corps members to the region where they will teach at the time that they are accepted into the program, taking into account corps members' preferences, the alignment of corps member qualifications with local teaching requirements (as determined by previous coursework and professional history), and the staffing needs of schools within each region. In each region, corps members apply for positions with TFA partner LEAs that have vacancies, including public school districts, public charter schools, and community-based organizations (CBOs).

[15] TFA considers low-income schools to be those schools in which at least 60 percent of students qualify for free or reduced-price lunch.

Table III.4. Placements of TFA's entering cohorts during the first two years of the TFA i3 scale-up (percentages unless otherwise indicated)

	Pre-scale-up cohorts		First two scale-up cohorts	
	2009–2010	2010–2011	2011–2012	2012–2013
Grade level				
Prekindergarten and kindergarten	8.6	6.7	7.4	6.9
Grades 1–5	28.0	27.4	28.9	29.3
Grades 6–8	32.3	32.7	32.7	30.6
Grades 9–12	31.2	33.1	31.0	33.2
Group				
General education	84.0	88.7	84.8	85.3
Special education	12.2	7.7	10.8	10.7
English language learners	3.8	3.5	4.4	4.0
School type				
Traditional public[a]	69.8	65.0	65.1	65.3
Public charter	27.0	32.9	32.7	32.9
Private	0.5	0.3	0.4	0.4
Early childhood	1.5	0.9	0.9	0.9
Catholic	0.3	0.0	0.1	0.1
Bureau of Indian Affairs	0.9	0.9	0.7	0.5
Poverty level[b]				
High percentage free or reduced-price lunch	83.3	82.2	85.6	84.1
Overall sample size[c]	**4,035**	**4,469**	**5,027**	**5,825**

Source: TFA placement data and Common Core of Data.

[a]Traditional public schools are noncharter schools.

[b]Schools are defined as high poverty if 60 percent or more of the student population qualifies for free or reduced-price lunch.

[c]Sample sizes for our analyses differ slightly from official TFA statistics on number of corps members cited earlier in the report, which classify corps members who take a leave of absence according to the year in which they were admitted rather than the year in which they actually began teaching.

TFA = Teach For America.

Corps members are hired through the same hiring process as other beginning teachers in their district or school. Most corps members interview across multiple LEAs in a region prior to finding a position. In some cases, where districts centrally assign all of their teachers, districts will hire corps members before identifying the schools where the corps members will be placed. In other LEAs, where principals make hiring decisions, corps members submit resumes to specific schools. Typically, corps members interview with LEAs between January and September, with the majority of interviews taking place during the summer before the corps members are to begin teaching. In 2012, approximately 40 percent of corps members were offered positions by schools or districts by late June, and 96 percent of corps members had been hired by the beginning of the school year. Though TFA does not guarantee teaching positions for all corps members, just 1 percent failed to secure a classroom placement in 2012. Most corps members who did not secure a placement failed to do so because they did not pass certification tests required by districts or states and therefore were ineligible to teach.

The types of classes and schools in which corps members were placed changed little between the two years prior to the scale-up and the first two years of the scale-up. Around 7 to 9 percent of incoming corps members taught prekindergarten or kindergarten over all four years examined, with about 30 percent of corps members in each of the grade ranges 1 through 5, 6 through 8, and 9 through 12. Around 85 percent of placements in all four years were in general

education classes, with 8 to 12 percent in special education and around 4 percent in English language learner (ELL) classes in all four years. Between 65 and 70 percent of placements were in traditional public schools and 27 to 33 percent were in charter schools. The poverty level of the schools in which corps members were placed remained relatively constant as well, with around 85 percent of corps members placed in low-income schools in all four years.[16]

E. Ongoing training and support

Once corps members are hired by partner schools and districts, regional TFA staff provides them with ongoing training and support during their two-year commitment. This includes one-on-one coaching support, group meetings specialized by grade and subject, and access to additional classroom resources and assessments via an online portal. Corps members in most regions must also complete alternative certification programs, state-defined routes through which individuals can begin teaching before completing all the requirements for state certification.

Round Zero. Following summer institute, corps members return to the regions where they will teach in the fall for a regional orientation, typically known as "Round Zero" or "First Eight Weeks." This training focuses on building relationships with students and their families; developing a vision and goals for their classroom; and working with state standards and district requirements to develop long-term instructional plans for the year, daily lesson plans, and assessments. Given the variation in district requirements and student populations across regions, the content of the regional orientations varies from region to region. As a supplement to in-person activities, several regions provide corps members with additional online modules to complete as preparation for their teaching placement.

Managers of Teacher Leadership Development. During their two-year commitment, corps members receive individualized support from their Manager of Teacher Leadership Development (MTLD), an instructional coach who provides one-on-one coaching and observational feedback. MTLDs work with corps members to prepare an individualized plan for the corps member's professional development that includes regular observation from the MTLD and often other skilled instructors. Following observations, MTLDs offer feedback to corps members on their teaching practice and provide suggestions for improvement. In addition to formal observations and debriefings, MTLDs also collect data on student progress toward goals for each corps member and provide corps members with resources tailored to the specific grade and subject area taught. TFA matches corps members to MTLDs either based upon grade and subject area or based upon the geographic location of a corps member's school, depending on the region. Additional TFA support staff specializing in specific subject areas and teaching strategies supplement the support provided by MTLDs. According to data from surveys TFA conducted of its corps members, at least 60 percent of corps members interacted with their MTLDs at least three times a month in the first two scale-up years, as in the year prior to the scale-up.

Ongoing group meetings. Over the course of the school year, corps members also regularly attend small-group and large-group meetings, designed as a venue through which to share best practices and resources. Regions utilize a variety of approaches to provide this group instruction.

[16] TFA considers low-income schools to be those schools in which at least 60 percent of students qualify for free or reduced-price lunch.

Some regions use "learning team" sessions, which are led by current corps members or alumni and are generally specialized by grade and subject area. In addition, some regions offer online modules targeted toward certain grades, content areas, or instructional practices.

Online resources. TFA provides its corps members with a number of online tools and resources through its TFAnet online community to help support and improve their teaching practices. These include sample student assessments, lesson plans, and other instructional planning tools; online trainings; video examples of model classrooms; and online forums in which corps members can discuss best practices.

Alternative certification programs. Prior to beginning their first teaching assignment, all corps members must receive state teaching certification (a license, certificate, credential, or permit) and be considered "highly qualified" under federal law and according to state-specific requirement. Because most corps members have not completed a traditional college-based education program before teaching, they are considered "nontraditional" or "alternative route" teachers in most states. As a part of their alternative certification program, corps members in most states receive added support and also must complete coursework as they progress toward the next level of certification or licensure. Depending on the region, corps members can complete coursework through a state-approved certification provider such as a school district, nonprofit organization, or local college or university. In 16 regions, TFA is itself a state-approved certification program in which regional corps members enroll. In many regions, corps members have the option of completing a master's degree by the end of their two-year teaching commitment.

Measuring teacher effectiveness. TFA encourages corps members to set both academic and personal goals for students and to use a variety of formal and informal assessments to monitor student development. TFA uses assessment data gathered by TFA corps members in combination with longitudinal teacher-linked data gathered from districts, states, and national test publishers to measure the effectiveness of its teachers relative to "high-performing" teachers nationwide, defined as teachers at the 75th percentile of student achievement growth. TFA deems corps members "effective" if their students' test score growth over the school year is the same as that achieved by a high-performing teacher and "highly effective" if their students' test score growth is one and a half times that achieved by a high-performing teacher. In 2012–2013, 32 percent of first-year teachers and 41 percent of second-year teachers were rated highly effective, and 32 percent of first-year teachers and 78 percent of second-year teachers were rated highly effective or effective according to this internal metric.

Although corps members' perceptions of TFA and the ongoing support they were provided were generally favorable over the full period examined according to TFA's internal end-of year corps member surveys, perceptions grew less favorable in each year, both pre-scale-up and into the first scale-up year (Table III.5).[17] For instance, more than half of corps members agreed or strongly agreed with the statement that "within TFA I feel part of a community where corps members help each other increase collective impact" according to an end-of year survey, but this

[17] As with the survey it conducts at the end of summer institute, TFA attempts to survey all corps members in its end-of-year survey. Response rates for first-year corps members were above 90 percent during all years in the analysis, from 2009–2010 to 2012–2013.

percentage declined over the period examined, from 64 percent in the 2009–2010 school year to 57 percent in the 2012–2013 school year. The percentage of corps members reporting either positive or very positive overall satisfaction with the program also declined over this period, from 64 percent in 2009–2010 to 48 percent in the 2011–2012 school year and 57 percent in the 2012–2013 school year. Corps members' views on the usefulness of individual components of the training and support remained relatively constant over this period, with the exception of views on online resources—the percentage of corps members who agreed or strongly agreed that the online resources aided their teaching declined from 61 to 35 percent between the 2009–2010 and 2012–2013 school years.

Table III.5. Corps member perceptions following first year of teaching (percentages unless otherwise indicated)

	Pre-scale-up cohorts		First two scale-up cohorts	
	Entering cohort 2009–2010	Entering cohort 2010–2011	Entering cohort 2011–2012	Entering cohort 2012–2013
Overall perceptions of TFA at end of school year				
Agreed or strongly agreed that "within TFA I feel part of a community where corps members help each other increase collective impact"	64.1	59.0	52.4	56.9
Positive or very positive overall satisfaction with TFA	64.0	58.5	47.9	57.1
Perceptions of ongoing support (agreed or strongly agreed that components aided teaching)				
Coaching from MTLDs	58.4	54.8	52.2	54.7
Online resources	60.9	50.9	41.7	34.7
Group learning activities	42.9	39.7	33.8	39.3
Alternative certification programs	31.5	23.7	27.6	33.0
Overall sample size	**3,582**	**3,906**	**4,247**	**4,925**

Source: TFA end-of-year surveys.

MTLDs = Managers of Teacher Leadership Development; TFA = Teach For America.

This page has been left blank for double-sided copying.

IV. TEACH FOR AMERICA AND COMPARISON TEACHERS IN THE STUDY

To provide context for the estimates of the effectiveness of TFA teachers presented in Chapter V, in this chapter we use information from the teacher survey to compare the characteristics of the TFA and comparison teachers in the study sample. We found many differences between the two types of teachers—they differed in their background characteristics, experience, preparation for teaching, support received throughout the school year, and attitudes toward teaching.

Compared with comparison teachers, TFA teachers in the sample:

- Were younger and less likely to be female and members of racial or ethnic minorities
- Were more likely to have graduated from a selective college or university
- Were less likely to have majored in early childhood or elementary education
- Had fewer years of teaching experience
- Reported completing similar amounts of pedagogy instruction but fewer days of student teaching in their preparation for teaching
- Were more likely to have taken coursework during the study school year, were more likely to have had a formal mentor during that year, and spent more time in professional development
- Spent more time in a typical week planning and preparing for classroom instruction, but less time helping other teachers plan instruction for their classes
- Were less satisfied with many aspects of teaching
- Were less likely to plan to spend the rest of their career as a classroom teacher

The comparison teachers included both teachers from traditional routes to certification and those from other alternative routes to certification—85 percent of comparison teaches were from traditional routes, and 15 percent were from other alternative routes. The proportion of comparison teachers from alternative routes was lower than in the prior experimental evaluations of TFA. In the 2004 study of elementary teachers (Decker et al. 2004), about a third of comparison teachers were from alternative routes, and in the 2013 study of secondary math teachers (Clark et al. 2013), 41 percent were from alternative routes.

A. Demographic characteristics

TFA teachers differed from comparison teachers in age, gender, and race/ethnicity (Table IV.1). As expected, given that the sample of TFA teachers was limited to teachers recruited under the i3 scale-up who were typically in their first or second year of teaching, TFA study teachers were on average significantly younger than comparison teachers. TFA teachers were significantly less likely to be female, and they were less likely to be members of racial or ethnic minorities. Almost 70 percent of TFA teachers were white, non-Hispanic compared with only 55 percent of comparison teachers (this difference was only statistically significant at the 10 percent level). TFA teachers were significantly more likely to be Asian and significantly less

likely to be black than were comparison teachers. Comparison teachers in the study were closer in age to the average elementary teacher nationwide than were TFA teachers, but TFA teachers looked more like the average elementary teacher in terms of gender and racial/ethnic distribution.

Table IV.1. Demographic characteristics of TFA and comparison teachers in the study and all elementary teachers nationwide (percentages unless otherwise indicated)

	Elementary teachers nationwide	TFA teachers	Comparison teachers	Difference between TFA and comparison teachers	*p*-Value
Age (average years)	42.4	24.4	42.8	-18.4**	0.000
Female	89.3	89.8	98.6	-8.8*	0.025
Race/ethnicity[a]					
Asian, non-Hispanic	1.7	11.9	2.7	9.1*	0.039
Black, non-Hispanic	7.1	11.9	34.2	-22.4**	0.003
Hispanic	8.7	6.8	11.0	-4.2	0.410
White, non-Hispanic	81.2	69.5	54.8	14.7+	0.086
Number of teachers	**1,626,800**	**59**	**76**		

Source: Data for elementary teachers nationwide from the Schools and Staffing Survey Teacher Questionnaire, 2011–2012; data for study teachers from the teacher survey.

Note: Information on study teachers is based on teachers in the study classrooms at the start of the school year.

[a]Racial and ethnic categories for study teachers are not mutually exclusive, so percentages may sum to more than 100.

+Difference is statistically significant at the 0.10 level, two-tailed test.
*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

B. Educational background

As expected, given that TFA focuses its recruitment efforts on the most competitive undergraduate institutions and on candidates without formal training in education, the educational background of TFA teachers in the study differed significantly from that of the comparison teachers (Table IV.2). As was the case in past studies of TFA, TFA teachers were more likely to have graduated from a selective college or university than were comparison teachers (76 versus 40 percent).[18] However, a higher percentage of comparison teachers in this study graduated from a selective college or university than in the past studies (Decker et al. 2004; Clark et al. 2013). In the 2004 study of elementary teachers, only 2 percent of comparison teachers had graduated from a selective school, and in the 2013 study of secondary math teachers, 23 percent of comparison teachers had graduated from a selective school. TFA teachers were less likely than comparison teachers to have majored in early childhood education or

[18] College selectivity data reported here for the teachers in our study and the prior random assignment studies are based on rankings from *Barron's Profiles of American Colleges 2013*. Selective colleges are those ranked as very competitive, highly competitive, or most competitive, and highly selective colleges are those ranked as highly competitive or most competitive. In contrast, data on college selectivity of all TFA corps members reported in Chapters II and III were collected by TFA and are based on *U.S. News & World Report* college rankings.

elementary education, and more likely to have majored in a field unrelated to education. They were also less likely to have any graduate degree and a graduate degree in education.

Table IV.2. Educational background of TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Bachelor's degree				
From a highly selective college or university[a]	23.6	5.2	18.5**	0.005
From a selective college or university[b]	76.4	39.7	36.7**	0.000
Major[c]				
Early childhood or prekindergarten general education	5.4	27.4	-22.1**	0.001
Elementary general education	14.3	53.2	-38.9**	0.000
Other education-related field	5.4	9.7	-4.3	0.382
Non-education-related field	83.9	25.8	58.1**	0.000
Major or minor[c]				
Early childhood or prekindergarten general education	5.4	30.6	-25.3**	0.000
Elementary general education	16.1	54.8	-38.8**	0.000
Other education-related field	10.7	12.9	-2.2	0.716
Non-education-related field	91.1	37.1	54.0**	0.000
Graduate degree				
Any graduate degree	8.5	38.2	-29.7**	0.000
Graduate degree in education	3.4	35.5	-32.1**	0.000
Early childhood or prekindergarten general education	0.0	7.9	-7.9*	0.027
Elementary general education	3.4	15.8	-12.4*	0.019
Other education-related field	0.0	17.1	-17.1**	0.001
Non-education-related field	5.1	2.6	2.5	0.458
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Highly selective colleges are those ranked by *Barron's Profiles of American Colleges 2013* as being highly competitive or most competitive.

[b]Selective colleges are those ranked as very competitive, highly competitive, or most competitive.

[c]Percentages might not sum to 100 if some sample members had a degree in more than one subject.

*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

C. Teaching experience

TFA teachers in the study had significantly less teaching experience, on average, than comparison teachers (Table IV.3), which is expected given that the TFA sample was limited to first and second year corps members, whereas there was no limit on the experience of comparison teachers. The TFA teachers had been teaching an average of 1.7 years compared with 13.6 years among the comparison teachers. The comparison teachers in this study were, on average, more experienced than the comparison teachers in past studies of TFA. In the 2004 TFA study, the median comparison teacher had been teaching for 6 years, and in the 2013 study, the average teacher had been teaching for 10.1 years.

Table IV.3. Teaching experience of TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Teaching experience (end of study year)				
Years of teaching experience (average)	1.7	13.7	-12.0**	0.000
1 or 2 years of teaching experience	98.3	11.8	86.5	
1 year of teaching experience	28.8	2.6	26.2	
2 years of teaching experience	69.5	9.2	60.3	
3 to 5 years of teaching experience[a]	1.7	11.8	-10.1	
More than 5 years of teaching experience	0.0	76.3	-76.3	
Chi-squared test of difference in distributions				0.000**
Sample size	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]A single TFA teacher reported being in the third year of teaching and had completed two of these years prior to joining TFA. This teacher was eligible for the TFA teacher sample because the teacher was trained under the i3 scale-up.

**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Almost all of the TFA teachers (99 percent) were in their first or second year of teaching, compared with only 13 percent of comparison teachers.[19] About a third of TFA teachers in the sample were in their first year of teaching and 68 percent were in their second year. As noted in Chapter II, first-year TFA teachers were somewhat underrepresented in the study sample—among the full population of TFA elementary school teachers recruited under the i3 scale-up, 56 percent were in their first year of teaching. Almost 80 percent of comparison teachers had more than five years' experience, and none of the TFA teachers in the sample had this much experience.

D. Teacher training

Although TFA and comparison teachers reported completing similar amounts of pedagogy instruction as part of their teacher training, TFA teachers reported completing significantly fewer days of student teaching, on average (Table IV.4). TFA teachers were less likely to report that they felt extremely or very prepared for their first teaching job (15 versus 55 percent) and that the instruction they received before their first teaching job was extremely or very helpful (39 versus 66 percent), compared with comparison teachers. However, these estimates should be interpreted with caution due to potential recall bias (Tourangeau et al. 2000). Because the comparison teachers had been teaching an average of 14 years at the time of the survey, they might have had a more difficult time accurately remembering the components of their teacher training and their preparedness for their first teaching job. In contrast, the TFA teachers all completed their initial training within the past one or two years and might have had more reliable recollections of their training experience.

[19] A single TFA teacher reported that she was in her third year of teaching and had completed two of these years prior to joining TFA. Because she was trained under the i3 scale-up, she was eligible for the sample.

Table IV.4. Training of TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Average hours of pedagogy or teaching strategies instruction as part of teacher training[a]	70.5	60.3	10.2	0.141
Days of student teaching as part of teacher training (average)[b]	27.9	46.2	-18.2**	0.000
No days	10.2	11.8	-1.7	
1 to 15	16.9	11.8	5.1	
16 to 60	57.6	34.2	23.4	
More than 60	15.3	42.1	-26.9	
Chi-squared test of difference in distributions				0.005**
Minutes per day spent teaching as part of teacher training (average)[c]	38.8	40.5	-1.7	0.556
Felt extremely or very prepared for first teaching job[d]	15.3	55.3	-40.0**	0.000
Felt instruction received to become a teacher before first teaching job was extremely or very helpful[e]	39.0	65.8	-26.8**	0.002
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Teachers were asked, "As part of your training to become a teacher, did you receive any instruction in pedagogy or teaching strategies?" If so, "In total, how many hours of instruction in pedagogy or teaching strategies did you receive?" Possible responses were none, 1 to 4, 5 to 20, 21 to 40, 41 to 60, 61 to 80, 81 to 100, and more than 100. To construct average hours of pedagogy training, we created a continuous variable equal to zero for teachers who completed no training, 100 for those who completed more than 100 hours, and the midpoint of the range for all other categories.

[b]Teachers were asked, "Did your teacher education/preparation program require you to do any student teaching or practice teaching in which you went to an elementary or secondary school and taught one or more lessons to a whole classroom of students?" If so, "On approximately how many days, in total, did you teach at least one full lesson to a whole classroom of students during your teacher education/preparation program?" Possible responses were fewer than 5, 6 to 10, 11 to 15, 16 to 20, 21 to 40, 41 to 60, 61 to 80, and more than 80. To construct average days of student teaching, we created a continuous variable equal to zero for teachers who did not do any student teaching, 80 for those who did more than 80 days, and the midpoint of the range for all other categories.

[c]Teachers were asked, "Did your teacher education/preparation program require you to do any student teaching or practice teaching in which you went to an elementary or secondary school and taught one or more lessons to a whole classroom of students?" If so, "On the days on which you taught at least one full lesson to a whole classroom of students as part of your teacher education/preparation program, how long did you typically teach?" Possible responses were fewer than 20 minutes, 20 to 30 minutes, 31 to 40 minutes, 41 to 50 minutes, and more than 50 minutes. To construct average minutes per day of student teaching, we created a continuous variable equal to zero for teachers who did not do any student teaching, 50 for those who did more than 50 minutes, and the midpoint of the range for all other categories.

[d]Possible responses were extremely prepared, very prepared, somewhat prepared, slightly prepared, and not at all prepared.

[e]Possible responses were extremely helpful, very helpful, somewhat helpful, slightly helpful, and not at all helpful.

**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

E. Coursework, support, and professional development during the school year

Because almost all TFA teachers in the sample were in their first or second year of teaching, many were still fulfilling coursework requirements for certification or obtaining an advanced degree. Relative to comparison teachers, TFA teachers were significantly more likely to have taken coursework during the school year, and they spent more total hours attending classes, although this difference was not statistically significant (Table IV.5). Of those who took

coursework, TFA teachers and comparison teachers reported doing so for similar reasons, with the highest percentage reporting that they were obtaining an advanced or master's degree not required for state certification. Among those who took coursework, TFA teachers were less likely than comparison teachers to feel that the coursework they took during the school year was very or extremely helpful.

Table IV.5. Coursework taken during the school year by TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Took coursework related to teaching job during school year	37.3	19.7	17.6*	0.023
Total hours spent during school year attending classes (average)[a]	80.2	56.3	23.9	0.388
Hours spent out of class during school year on coursework (average)[a]	39.5	35.1	4.3	0.794
Reason for coursework				
Maintain current professional state teacher certification	4.5	21.4	-16.9	
Obtain state teacher certification without advanced or master's degree	27.3	7.1	20.1	
Obtain advanced or master's degree required for state teacher certification	13.6	14.3	-0.6	
Obtain advanced or master's degree not required for state teacher certification	50.0	42.9	7.1	
Other	4.5	14.3	-9.7	
Chi-squared test of difference in distributions				0.283
Helpfulness of coursework				
Felt coursework was very or extremely helpful[b]	22.7	80.0	-57.3**	0.000
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Calculations are based on all teachers, regardless of whether they took coursework related to their teaching job during the school year.

[b]Possible responses were extremely helpful, very helpful, somewhat helpful, slightly helpful, not at all helpful.

*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

TFA teachers were significantly more likely than comparison teachers to have had a formal mentor during the school year, but similar percentages in both groups reported having an informal mentor (Table IV.6). The TFA teachers were significantly less likely than comparison teachers to report that their mentors were other teachers or administrators and more likely to report that their mentors were faculty or staff members affiliated with their teacher preparation program. Although less than 40 percent of both groups thought their formal mentor was helpful, more than 80 percent in both groups thought their informal mentor was very or extremely helpful.

Table IV.6. Mentoring received during the school year by TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Had a formal mentor during school year	72.9	15.8	57.1**	0.000
Type of formal mentor				
Teacher from school	39.5	41.7	-2.1	0.896
Administrator from school	14.0	41.7	-27.7*	0.034
Teacher or administrator from outside school assigned by district	2.3	33.3	-31.0**	0.001
Faculty member or staff member affiliated with teacher preparation program	79.1	0.0	79.1**	0.000
Some other type of mentor	2.3	0.0	2.3	0.602
Type of support received from formal mentor[a]				
Average time spent being observed by mentors (minutes)	122.1	96.8	25.3	0.726
Average time spent observing mentor (minutes)	8.4	33.7	-25.3	0.426
Average time spent in formal meetings with mentors (minutes)	181.7	62.5	119.2**	0.009
Average time spent in informal meetings with mentors (minutes)	121.5	52.8	68.8	0.155
Average number of times received written feedback on teaching performance	2.9	1.5	1.4	0.142
Average number of times received written feedback on materials developed for classroom	2.0	1.6	0.4	0.720
Average number of times received resources to use in classroom	4.7	1.4	3.3*	0.011
Felt formal mentoring was very or extremely helpful[a]	39.5	33.3	6.2	0.702
Had informal mentor during school year	61.0	51.3	9.7	0.264
Type of informal mentor				
Teacher from school	77.8	74.4	3.4	0.733
Administrator from school	13.9	33.3	-19.4*	0.050
Faculty member or staff member affiliated with teacher preparation program	44.4	17.9	26.5*	0.013
Some other type of mentor (other)	8.3	15.4	-7.1	0.355
Felt that informal mentoring was very or extremely helpful[b]	80.6	82.1	-1.5	0.870
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Calculations are based on all teachers, regardless of whether they had a formal mentor during the school year.

[b]Possible responses were extremely helpful, very helpful, somewhat helpful, slightly helpful, and not at all helpful.

*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

When combining professional development provided both by the school or school district and the teacher preparation program, TFA teachers reported spending more time in professional development during the school year, on average, than comparison teachers (Table IV.7).[20] TFA

[20] The teacher survey asked teachers about time they spent in both coursework and professional development. Coursework included university-based classes taken to maintain or obtain certification or an advanced degree, whereas professional development included classes, workshops, or seminars provided by their school, school district, or teacher preparation program.

teachers spent slightly less time in professional development provided by their school or district than comparison teachers (13.3 versus 16.2 hours), but they spent significantly more time in professional development provided by their teacher preparation program (15.3 versus 1.9 hours).[21]

Table IV.7. Professional development and other support activities for TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Time spent in professional development classes, workshops, or seminars during school year				
Provided by school or school district				
Average hours spent in classes[a]	13.3	16.2	-2.9**	0.004
Percentage of classes that took place outside normal teaching hours	53.3	53.4	-0.1	0.986
Provided by teacher preparation program				
Average hours spent in classes[b]	15.3	1.9	13.5**	0.000
Percentage of classes that took place outside normal teaching hours	96.8	50.0	46.8**	0.000
Type of support received during school year				
Reduced teaching schedule	3.4	1.4	2.0	0.450
Seminars or classes for beginning teachers	37.3	19.4	17.8*	0.023
Extra professional classroom assistance	37.3	35.6	1.7	0.844
Regular supportive communication with your principal, other administrators, or department chair	36.2	66.2	-30.0**	0.001
Opportunities to observe other teachers	40.7	37.3	3.3	0.696
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Teachers were asked, "During this school year, did you attend any professional development classes, workshops, or seminars *provided by your school or school district*?" If so, "In total, how many hours did you spend attending these professional development classes, workshops, or seminars?" Possible responses were none, 1 to 4, 5 to 10, 11 to 20, and more than 20. To construct average hours of professional development, we created a continuous variable equal to zero for teachers who did no professional development, 20 for those who did more than 20 hours, and the midpoint of the range for all other categories.

[b]Teachers were asked, "During this school year, did you attend any professional development classes, workshops, or seminars *provided by your teacher preparation program*?" If so, "In total, how many hours did you spend attending these professional development classes, workshops, or seminars?" Possible responses were none, 1 to 4, 5 to 10, 11 to 20, and more than 20. To construct average hours of professional development, we created a continuous variable equal to zero for teachers who did no professional development, 20 for those who did more than 20 hours, and the midpoint of the range for all other categories.

*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

TFA teachers reported that almost all of the professional development provided by their preparation program took place outside of normal teaching hours, whereas comparison teachers

[21] Professional development opportunities offered by the school or district may differ by teachers' years of experience. Therefore, the difference in average years of experience between TFA and comparison teachers might explain the difference in the reported amount of time spent in professional development provided by the school or district.

reported that only half of the classes provided by their preparation programs took place outside of normal teaching hours. TFA teachers more commonly reported being offered seminars or classes for beginning teachers than comparison teachers but less commonly reported receiving regular supportive communication from their school administrators or department chair.

F. Classroom experiences

There were a few differences in the ways TFA and comparison teachers allocated their work time (Table IV.8). When asked about how they spend their non-classroom time during a typical week, both groups of teachers reported spending similar amounts of time working with students, interacting with parents, and attending faculty meetings. However, TFA teachers reported spending significantly less time grading, reviewing, or providing feedback on student work and on reviewing and analyzing student performance on assessments than comparison teachers. They spent significantly more time than comparison teachers planning and preparing for classroom instruction, but less time helping other teachers plan instruction for their classes. When asked how they spend their classroom time during a typical day, teachers in both groups reported

Table IV.8. How TFA and comparison teachers spend their time during a typical week and day

	TFA teachers	Comparison teachers	Difference	*p*-Value
Time spent during typical week (average hours)				
Grading, reviewing, or providing feedback on student work	2.6	4.4	-1.8**	0.005
Planning and preparing for classroom instruction	7.7	5.7	2.0*	0.015
Reviewing and analyzing student performance on assessments	1.9	2.6	-0.7*	0.037
Working with students outside of normal classroom hours	2.5	1.6	0.8	0.274
Interacting with parents	1.6	1.6	0.0	0.936
Attending faculty meetings	1.2	1.4	-0.2	0.215
Accessing online or hard-copy resources to help plan instruction	2.7	2.8	-0.1	0.799
Consulting other teachers or experts to help plan instruction for own class	2.0	1.5	0.4	0.486
Helping other teachers plan instruction for their classes	0.8	1.3	-0.5*	0.046
Time spent during typical day of teaching (average hours)				
Instructional activities				
Teacher-directed whole class activities	2.1	2.1	0.0	0.989
Teacher-directed small-group activities	1.4	1.7	-0.3	0.170
Students working independently in pairs/teams/small groups	1.5	1.5	-0.1	0.676
Students working individually on class assignments	0.9	1.0	-0.2	0.300
Other instructional activities	0.1	0.3	-0.2	0.119
Noninstructional activities				
Daily routines	0.9	0.9	0.0	0.905
Behavior management	0.6	0.8	-0.2	0.271
Free play	0.5	0.6	-0.1	0.244
Other noninstructional activities	0.1	0.1	-0.1	0.209
Number of teachers	**59**	**76**		

Source: Teacher survey.
Note: Information in the table is based on teachers in the study classrooms at the start of the school year.
*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.
TFA = Teach For America.

spending the most time on teacher-directed whole class activities followed by other types of instructional activities including teacher-directed small-group work, student-directed small-group work, and individual work.

Teachers' perceptions of issues that hinder student learning in their classrooms can reflect the challenges they encounter, but overall TFA and comparison teachers had similar perceptions of these issues (Table IV.9). Both groups of teachers commonly reported that students' insufficient academic foundation or preparation, a lack of parental or home support, student absenteeism, and general misbehavior hindered student learning to a great or very great extent.

Table IV.9. Classroom experiences and goals of TFA and comparison teachers in the study

	Percentage of teachers who said issue hindered student learning in classroom to a great or very great extend since the start of the 2012–2013 school year[a]			
	TFA teachers	Comparison teachers	Difference	*p*-Value
Student tardiness	20.3	19.7	0.6	0.932
Student absenteeism/class cutting	39.7	38.7	1.0	0.909
Physical conflicts among students	10.2	11.0	-0.8	0.885
Verbal conflicts among students	25.4	18.7	6.8	0.349
Verbal abuse of teachers by students	8.8	8.2	0.6	0.911
General misbehavior	39.0	28.4	10.6	0.199
Students' insufficient academic foundation/preparation	55.9	44.0	11.9	0.173
Lack of student effort or motivation	27.1	26.7	0.5	0.954
Lack of adequate classroom materials or equipment	27.1	20.3	6.8	0.357
Inadequate learning space	15.3	13.3	1.9	0.754
Teacher or administrative turnover/attrition	15.5	13.7	1.8	0.771
Lack of parental/home support	39.0	53.9	-15.0[+]	0.085
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Possible responses were to a very great extent, to a great extent, to a moderate extent, to a slight extent, and not at all.
[+]Difference is statistically significant at the 0.10 level, two-tailed test.

TFA = Teach For America.

G. Job satisfaction and career plans

Because teachers' levels of satisfaction with their jobs may influence how long they stay in teaching, we measured the job satisfaction of both groups. TFA teachers were generally less satisfied with various aspects of teaching than comparison teachers (Table IV.10). They were less satisfied with all aspects of teaching at their current school, including the level of collegiality with other teachers, professional caliber of their colleagues, sense of physical safety, availability of resources, influence over school policies, autonomy or control over classroom, support from administration, opportunities for professional development, students' behavior, principal's leadership, and the procedures for performance evaluation. When asked about satisfaction with

the teaching profession more generally, TFA teachers were significantly less likely to report being satisfied with opportunities for professional advancement, the salary, the professional prestige, and intellectual challenge than comparison teachers. However, most TFA teachers and comparison teachers were satisfied with the opportunities to help students and personal fulfillment offered by the teaching profession. Differences in reported levels of satisfaction could result in part from differences in the level of experience between TFA and comparison teachers. If the least satisfied teachers leave the profession over time, those who remain could be more satisfied with their jobs than novices, such as the TFA teachers in our sample). Alternatively, TFA teachers might be less satisfied with teaching if their opportunities outside of teaching are perceived to be more rewarding than those of comparison teachers.

Table IV.10. Job satisfaction of TFA and comparison teachers in the study

	Percentage of teachers who were somewhat or very satisfied with this aspect of job[a]			
	TFA teachers	Comparison teachers	Difference	*p*-Value
Aspect of teaching at current school				
Level of collegiality feel with other teachers at school	61.0	84.2	-23.2**	0.002
Professional caliber of colleagues	38.6	82.9	-44.3**	0.000
Sense of own physical safety at school	71.2	86.8	-15.7*	0.024
Availability of resources and materials/equipment for classroom	40.7	61.8	-21.2*	0.014
Influence over school policies and practices	11.9	51.3	-39.5**	0.000
Autonomy or control over classroom	62.7	78.9	-16.2*	0.038
Recognition and/or support from administration	28.8	65.8	-37.0**	0.000
Opportunities for professional development	44.1	81.6	-37.5**	0.000
Students' discipline and behavior	41.4	53.9	-12.6	0.151
Principal's leadership and vision	27.1	72.4	-45.2**	0.000
Support provided by assistant principal	39.7	66.2	-26.5**	0.003
Procedures for performance evaluation	25.4	64.5	-39.0**	0.000
Aspect of teaching profession				
Opportunities for professional advancement	30.5	60.5	-30.0**	0.000
Salary	22.0	38.2	-16.1*	0.045
Benefits	46.6	50.0	-3.4	0.697
Professional prestige	25.4	42.1	-16.7*	0.044
Intellectual challenge	51.7	75.0	-23.3**	0.005
Opportunities to help students achieve academically	83.1	86.8	-3.8	0.542
Opportunities to help students be successful in and outside of school	76.3	81.3	-5.1	0.478
Personal fulfillment	76.3	86.7	-10.4	0.121
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Teachers were asked about their satisfaction with each aspect of their job—possible responses were very dissatisfied, somewhat dissatisfied, somewhat satisfied, and very satisfied.

*Difference is statistically significant at the 0.05 level, two-tailed test.

**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Consistent with the fact that TFA requires its teachers to make only a two-year commitment to teaching, most TFA teachers did not plan to spend the rest of their career as a classroom teacher, whereas the opposite was true for comparison teachers (Table IV.11). More than 87 percent of TFA teachers reported that they did not plan to spend the rest of their career as a

classroom teacher, compared with only 26 percent of comparison teachers. Of those who planned to leave the profession, TFA teachers planned to teach for fewer additional years, on average, than comparison teachers, and they planned to stay at their current school for fewer years. TFA teachers who planned to leave the profession expected to pursue different types of careers than the comparison teachers who planned to leave. TFA teachers were less likely than comparison teachers to anticipate pursuing another education-related career and more likely to anticipate pursuing a non-education-related career. Some of the differences might be driven by the more extensive experience of some comparison teachers. These teachers, with an average of 14 years of experience, could have already chosen to commit to teaching as a professional career, and therefore might have had different projections about their future career plans than the novice TFA teachers in the sample.

Table IV.11. Career plans for TFA and comparison teachers in the study (percentages unless otherwise indicated)

	TFA teachers	Comparison teachers	Difference	*p*-Value
Do not plan to spend the rest of career as classroom teacher	87.5	26.3	61.2**	0.000
For those who plan to leave teaching profession				
Number of years plan to teach after 2012–2013 school year (average)[a]	1.5	2.5	-1.0*	0.046
0 years	25.0	6.7	18.3	
1 to 2 years	50.0	46.7	3.3	
3 to 5 years	14.3	26.7	-12.4	
6 or more years	0.0	6.7	-6.7	
Unsure	10.7	13.3	-2.6	
Chi-squared test of difference in distributions				0.341
Number years plan to teach at current school after 2012–2013 school year (average)[a]	0.7	2.3	-1.7**	0.001
0 years	53.6	26.7	26.9	
1 to 2 years	42.9	20.0	22.9	
3 to 5 years	0.0	33.3	-33.3	
6 or more years	0.0	6.7	-6.7	
Unsure	3.6	13.3	-9.8	
Chi-squared test of difference in distributions				0.004**
Anticipated primary career pursuit after ending classroom teaching career				
Other education-related career	42.9	80.0	-37.1	
Non-education-related career	42.9	6.7	36.2	
Undecided	14.3	13.3	1.0	
Chi-squared test of difference in distributions				0.036*
Number of teachers	**59**	**76**		

Source: Teacher survey.

Note: Information in the table is based on teachers in the study classrooms at the start of the school year.

[a]Teachers were asked, "How many more years do you think you will teach after this school year (2012-2013)?" Possible responses were none, 1 to 2 more years, 3 to 5 more years, 6 or more years, and don't know/unsure. To construct average years, we created a continuous variable equal to zero for teachers who planned to teach 0 more years, 6 for those who planned to teach for 6 or more years, and the midpoint of the range for the other two categories.

[b]Teachers were asked, "How many more years do you think you will teach at your current school after this school year (2012-2013)?" Possible responses were none, 1 to 2 more years, 3 to 5 more years, 6 or more years, and don't know/unsure. To construct average years, we created a continuous variable equal to zero for teachers who planned to teach 0 more years, 6 for those who planned to teach for 6 or more years, and the midpoint of the range for the other two categories.

*Difference is statistically significant at the 0.05 level, two-tailed test.
**Difference is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

V. TFA IMPACTS ON MATH AND READING ACHIEVEMENT

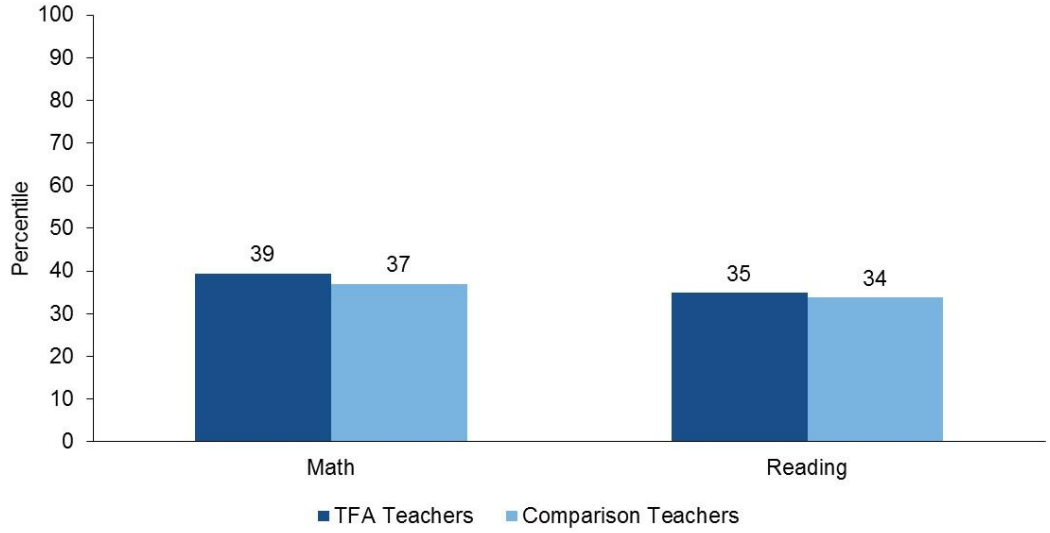
In this chapter, we examine the effectiveness of TFA teachers recruited during the first two years of the i3 scale-up, relative to comparison teachers who taught the same grade and subjects in the same schools. We focus on the core subjects of math and reading, and limit our analysis to the elementary grades, which accounted for 36 percent of TFA's placements during this period. As summarized in Chapter IV, the TFA teachers in the sample were more likely than comparison teachers to have graduated from a selective college but had far fewer years of teaching experience, on average. Among comparison teachers in the sample, 85 percent were from traditional routes into teaching.

To estimate effectiveness, we compared end-of-year math and reading scores of students assigned to TFA teachers with those of students assigned to comparison teachers. Because we randomly assigned students to teachers, both sets of students were similar at the start of the school year. Thus, comparing the achievement of the two groups of students at the end of the school year provides a rigorous measure of the relative effectiveness of TFA teachers.

A. Impacts of TFA teachers relative to comparison teachers

On average, the TFA teachers in our sample were equally as effective as comparison teachers in both reading and math (Figure V.1). In both subjects, the students assigned to TFA teachers scored slightly higher, on average, than those assigned to comparison teachers; however, these differences were not statistically significant. In math, students of TFA teachers scored at the 39th percentile and students of comparison teachers scored at the 37th percentile among all students statewide or nationwide who took the same test. In reading, students of TFA teachers scored at the 35th percentile and students of comparison teachers scored at the 34th percentile.

Figure V.1. No significant differences in achievement



Source: District administrative records and study-administered Woodcock-Johnson assessments.

Note: Average test scores, in z-score units, were regression-adjusted for classroom match fixed effects and all covariates listed in Appendix Table A.9 and then converted to percentiles based on a normal distribution. Neither difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.

TFA = Teach For America.

Our finding that TFA and comparison teachers were equally effective is robust to multiple sensitivity analyses (Appendix Tables B.1 and B.2). We estimated models that (1) excluded matches in which a high proportion of students was exempted from random assignment, (2) excluded students who took the tests in Spanish, (3) modified the way we standardized end-of-year test scores, (4) allowed the relationship between student background characteristics and end-of-year achievement to vary across lower elementary and upper elementary school students, (5) changed our strategy for handling missing data, (6) used alternative approaches to weighting classroom matches, and (7) accounted for students who switched to a different type of teacher (TFA or comparison) from their originally assigned teacher. In no cases were the differences in the effectiveness of TFA and comparison teachers statistically significant at conventional levels. See Appendix B for details.

B. Impacts among subgroups of TFA and comparison teachers

We also estimated TFA impacts for particular subgroups of students and teachers. This allowed us to examine whether TFA teachers' impacts varied across grade level, when they were compared only with novice comparison teachers, and when they were compared only with traditionally certified comparison teachers.

1. Impacts by grade level

We estimated impacts for subgroups based on grade level. For math, we estimated impacts for two grade-level subgroups: (1) lower elementary (grades 1 and 2) and (2) upper elementary (grades 3 through 5). For reading, we estimated impacts for three grade-level subgroups: (1) early childhood (prekindergarten and kindergarten), (2) lower elementary (prekindergarten through grade 2), and (3) upper elementary (grades 3 through 5).[22] Impacts might vary by grade level for a variety of reasons—for instance, TFA's training could be more effective for particular grade levels or the quality of comparison teachers could vary by grade level. We found some evidence that TFA teachers in lower elementary grades had positive effects on student achievement in both reading and math. In math, students assigned to TFA teachers in grades 1 and 2 outscored their peers assigned to comparison teachers by 0.16 standard deviations (Table V.1, middle panel). This difference that was almost statistically significant at conventional levels (p-value = 0.054). This effect is equal to about 15 percent of an average year of learning for students who took the same assessments in these grades nationwide—that is, about 1.5 months of learning in a 10-month school year.[23] Similarly, in reading, students assigned to TFA teachers in grades prekindergarten through 2 outscored their peers assigned to comparison teachers by a statistically significant 0.12 standard deviations (Table V.2, middle panel). This effect is equal to about 13 percent of an average year of learning for students who took the same assessments in

[22] As discussed in Chapter II, we analyzed impacts for prekindergarten and kindergarten both on their own and as part of the lower elementary subgroup for reading because there was no prior rigorous evidence of TFA teachers' effectiveness at the prekindergarten and kindergarten levels, and sample sizes were too small for us to analyze first and second grade students as a distinct subgroup. We intentionally oversampled prekindergarten and kindergarten students so that we could conduct this subgroup analysis. Although we also assessed students in prekindergarten and kindergarten in math, we were unable to include these scores in the analysis due to errors in test administration.

[23] To translate the effect into years of learning, we divided the impact estimate in *W* score units by the average annual gain in *W* scores for the relevant Woodcock-Johnson assessments for students ages 4 to 7, available from McGrew et al. (2007).

these grades nationwide—that is, about 1.3 months of learning in a 10-month school year. We did not find statistically significant effects on reading or math scores for any of the other grade-level subgroups we examined.

2. Impacts relative to novice comparison teachers

We also examined novice TFA teachers' effectiveness relative to other novice teachers (those in their first two years of teaching). Given that almost all TFA teachers were in their first or second year of teaching at the time of the study, and that teacher effectiveness typically improves with experience (Hanushek et al. 2005; Boyd et al. 2006; Kane et al. 2008; Papay and Kraft 2013), we might expect TFA teachers to perform better when compared with other novice teachers. For both math and reading, the impact estimate when we compare novice TFA and comparison teachers is positive but not statistically significant (Tables V.1 and V.2, bottom panel). These estimates are based on very small samples and may not be reliable.[24]

Table V.1. Differences in effectiveness between TFA and comparison teachers by subgroup, math

	Impact estimates			Sample sizes	
	Effect size	Standard error	*p*-Value	Students	Teachers
(1) Benchmark (all students)	**0.07**	**0.05**	**0.197**	**1,182**	**83**
(2) Lower elementary school students (1 and 2)	0.16+	0.08	0.054	770	56
(3) Upper elementary school students (3 to 5)	0.01	0.07	0.885	412	27
(4) Novice comparison teachers[a]	0.07	0.17	0.692	170	13
(5) Traditionally certified comparison teachers	0.10+	0.06	0.098	1,078	74

Source: District administrative records and study-administered Woodcock-Johnson assessments.

Note: The sample sizes presented are for the subgroup of interest only. The model sample size consists of all students in the benchmark model.

+Significantly different from zero at the .10 level, two-tailed test.

[a]We define novice teachers as those in their first or second year of teaching. This estimate excludes the single TFA teacher in the sample who had taught for two years before entering TFA and thus had taught for three years in total.

K = kindergarten; pre-K = prekindergarten; TFA = Teach For America.

[24] Because our sample of TFA teachers was limited primarily to those with just one or two years of experience, we defined novice teachers as those with fewer than three years of experience, so that the TFA and comparison teachers in this analysis would have comparable amounts of experience. Other studies (Decker et al. 2004; Clark et al. 2013) have defined novice teachers as those in their first three years of teaching. Using this alternative definition of novice, we also find no statistically significant effects of TFA teachers on reading or math.

Table V.2. Differences in effectiveness between TFA and comparison teachers by subgroup, reading

	Impact estimates			Sample sizes	
	Effect size	Standard error	*p*-Value	Students	Teachers
(1) Benchmark (all students)	**0.03**	**0.05**	**0.570**	**2,123**	**154**
(2) Early childhood students (pre-K and K)	0.15	0.12	0.214	878	67
(3) Lower elementary school students (pre-K to 2)	0.12*	0.06	0.035	1,653	123
(4) Upper elementary school students (3 to 5)	-0.07	0.08	0.398	470	31
(5) Novice comparison teachers[a]	0.13	0.12	0.263	313	23
(6) Traditionally-certified comparison teachers	0.03	0.05	0.640	1,884	132

Source: District administrative records and study-administered Woodcock-Johnson assessments.

Note: The sample sizes presented are for the subgroup of interest only. The model sample size consists of all students in the benchmark model.

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

[a]We define novice teachers as those in their first or second year of teaching. This estimate excludes the single TFA teacher in the sample who had taught for two years before entering TFA and thus had taught for three years in total.

K = kindergarten; pre-K = prekindergarten; TFA = Teach For America.

3. Impacts relative to traditionally certified comparison teachers

We also estimated impacts of TFA teachers relative to traditionally certified comparison teachers. Critics of TFA have raised concerns that corps members are underprepared for teaching relative to teachers who completed traditional university-based teacher certification programs (Darling-Hammond 2011; Ravitch 2013), and this analysis allows us to examine that concern. We found that for both math and reading, TFA teachers were equally as effective as traditionally certified comparison teachers (including both novices and veterans).

VI. DISCUSSION

In this report, we examined the effectiveness of TFA teachers recruited during the first two years of TFA's efforts to scale up its program under an i3 grant from the U.S. Department of Education. Under the scale-up, TFA planned to increase the size of its teacher corps by more than 80 percent over four years. Our study used a rigorous random assignment design to estimate the effects of TFA corps members recruited under the scale-up on student achievement in reading and math, focusing on first- and second-year corps members teaching in prekindergarten through grade 5 in the 2012–13 school year. This was the second year of the scale-up, by which time TFA had expanded its placements by 25 percent from the pre-scale-up year, from 8,206 to 10,255 first- and second-year corps members.

We found that the first- and second-year TFA teachers in our sample were equally as effective as other teachers in the same high-poverty schools in both reading and math. On average, students assigned to TFA teachers scored slightly above students assigned to non-TFA teachers, but these differences were small and not statistically significant. However, we found that TFA teachers in lower elementary grades (prekindergarten through grade 2) had a positive, statistically significant effect on student reading achievement of 0.12 standard deviations, or about 1.3 additional months of learning for the average student in these grades nationwide. Similarly, we found that TFA teacher in grades 1 and 2 had a positive effect on student math achievement of 0.16 standard deviations, or about 1.5 months of additional learning. This difference was almost statistically significant at conventional levels (p-value = 0.054).

Our findings differ from the first experimental study of TFA elementary school teachers, which found that TFA teachers were more effective than colleagues of any experience level in teaching math and equally effective in teaching reading (Decker et al. 2004). Grade-level results in Decker et al. (2004) do not show any particular pattern for math. For reading, they found a positive and statistically significant result for fifth grade teachers but not for other grades. The second experimental study of TFA, which focused on secondary math teachers, found that TFA teachers were more effective than their colleagues at teaching math in secondary grades (Clark et al. 2013). By contrast, in the current study we find a difference not by subject but by grade level—TFA teachers are more effective than other teachers in the same schools in prekindergarten to grade 2 and as effective in grades three to five.

Our study provides a snapshot of TFA's effectiveness at the elementary school level in the second year of the i3 scale-up. It is possible that the effectiveness of TFA's teachers could either increase or decrease as the program continues to strive to meet the needs of schools with many high-poverty students. However, the findings suggest that TFA can provide high poverty schools with teachers who are, on average, as effective as other teachers in these same schools, and potentially more effective at lower grade levels.

This page has been left blank for double-sided copying.

REFERENCES

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, vol. 91, no. 434, 1996, pp. 444–455.

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.

Chiang, Hanley S., Melissa A. Clark, and Sheena McConnell. "Supplying Disadvantaged Schools with Effective Teachers: Experimental Evidence on Secondary Math Teachers from Teach For America." Mathematica Policy Research Working Paper. Princeton, NJ: Mathematica Policy Research, May 2014.

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." NCEE 2013-4015. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2013.

Darling-Hammond, Linda. "Teacher Preparation is Essential to TFA's Future." *Education Week*, March 14, 2011. Available at [http://www.edweek.org/ew/articles/2011/03/16/24darling-hammond.h30.html]. Accessed July 23, 2014.

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.

Goldring, Rebecca, Lucinda Gray, and Amy Bitterman. "Characteristics of Public and Private Elementary and Secondary School Teachers in the United States: Results from the 2011–12 Schools and Staffing Survey." NCES 2013-314. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2013.

Hansen, Michael, Ben Backes, Victoria Brady, and Zeyu Xu. "Examining Spillover Effects from Teach For America Corps Members in Miami-Dade County Public Schools." National Center for Analysis of Longitudinal Data in Education Research Working Paper 113. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, June 2014.

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.

Hedges, Larry V. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics*, vol. 6, no. 2, 1981, pp. 107–128.

Henry, Gary T., Kevin C. Bastian, C. Kevin Fortner, David C. Kershaw, Kelly M. Purtell, Charles L. Thompson, and Rebecca A. Zulli. "Teacher Preparation Policies and Their Effects on Student Achievement." *Education Finance and Policy*, vol. 9, no. 3, 2014, pp. 264–303.

Huber, Peter J. "The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, edited by L.M. LeCam and J. Neyman. Berkeley, CA: University of California Press, 1967.

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.

Liang, Kung-Yee, and Scott L. Zeger. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, vol. 73, 1986, pp. 13–22.

McGrew, Kevin S., Fredrick A. Schrank, and Richard W. Woodcock. *Woodcock-Johnson® III Normative Update: Technical Manual*. Rolling Meadows, IL: Riverside Publishing, 2007.

Mead, Sara, Carolyn Chuong, and Caroline Goodson. "Exponential Growth, Unexpected Challenges: How Teach For America Grew in Scale and Impact." Sudbury, MA: Bellwether Education Partners, 2015. Available at http://bellwethereducation.org/sites/default/files/Bellwther_TFA_Growth.pdf.

Papay, John, and Matthew Kraft. "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement." Working Paper. Cambridge, MA: Harvard University, May 2013.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. "What to Do When Data Are Missing in Group Randomized Controlled Trials." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, October 2009.

Raghunathan, Trivellore E., James M. Lepkowski, John Van Hoewyk, and Peter Solenberger. "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models." *Survey Methodology*, vol. 27, no. 1, 2001, pp. 85–95.

Ravitch, Diane. *Reign of Error: The Hoax of the Privatization Movement and the Danger to America's Public Schools*. New York: Alfred A. Knopf, 2013.

Rotherham, Andrew J. "Teach For America Makes the Grade at Challenged Schools, Criticism Aside." *U.S. News & World Report*, February 9, 2009. Available at [http://www.usnews.com/opinion/articles/2009/02/09/teach-for-america-makes-the-grade-at-challenged-schools-criticism-aside]. Accessed July 23, 2014.

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

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. *The Psychology of Survey Response*. New York: Cambridge University Press, 2000.

U.S. Department of Education. *What Works Clearinghouse Procedures and Standards Handbook, Version 3.0.* Available at [http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf]. Accessed July 30, 2014.

White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* vol. 48, 1980, pp. 817–830.

Xu, Zeyu, Jane Hannaway, and Colin Taylor. "Making a Difference? The Effects of Teach For America in High School." Washington, DC: Urban Institute, March 2008.

Zukiewicz, Marykate, Melissa A. Clark, and Libby Makowsky. "Implementation of the Teach For America Investing in Innovation Scale-Up." Princeton, NJ: Mathematica Policy Research, March 2015.

This page has been left blank for double-sided copying.

APPENDIX A: STUDY DESIGN, DATA COLLECTION, AND ANALYTIC METHODS

This page has been left blank for double-sided copying.

In this appendix we provide additional detail on the design, data collection, and analytic methods used for the impact evaluation, including recruitment of districts, schools, and classroom matches; selection and assignment of students; response rates for data collection; statistical power of the impact analysis; sample weights; and analytic methods for the contextual and impact analyses.

A. Recruitment of districts/partners, schools, and classroom matches

As discussed in Chapter II, we focused recruitment efforts on districts and other placement partners with large concentrations of elementary TFA teachers. Figure A.1 illustrates the recruitment of districts or placement partners into the sample and Figure A.2 illustrates the recruitment of schools into the sample.

Out of TFA's 394 placement partners for the 2012–2013 school year, we contacted 70. Of these 70, 28 allowed us to contact their schools directly to assess interest and eligibility, and 42 either declined to participate or were unresponsive to our requests to discuss the study with them. Of the 28 that agreed to participate, 15 had at least one school that (1) was interested in participating, (2) had at least one eligible classroom match, and (3) allowed us to conduct random assignment. All matches we randomly assigned in two districts dropped out of the study, leaving 13 districts or placement partners in the sample (11 public school districts, one charter school district, and one community-based organization that runs an early childhood education program).

We randomly assigned at least one classroom match in each of 48 schools with a total of 82 matches. Thirty-six of these 48 schools (75 percent, comprising 57 matches) properly implemented random assignment, maintained viable classroom matches, and cooperated with data collection activities—these schools and matches formed the study's sample. The remaining 12 schools (25 percent) were dropped from the study sample. Ten of these schools were dropped because they failed to implement random assignment—the rosters they sent to the study team after random assignment did not correspond to the assignments we had given them, and they failed to make the requested changes.[25] The other two schools were dropped after random assignment because there were personnel changes or the school decided to departmentalize instruction by having all students within a match go to one teacher for reading and the other teacher for math.[26]

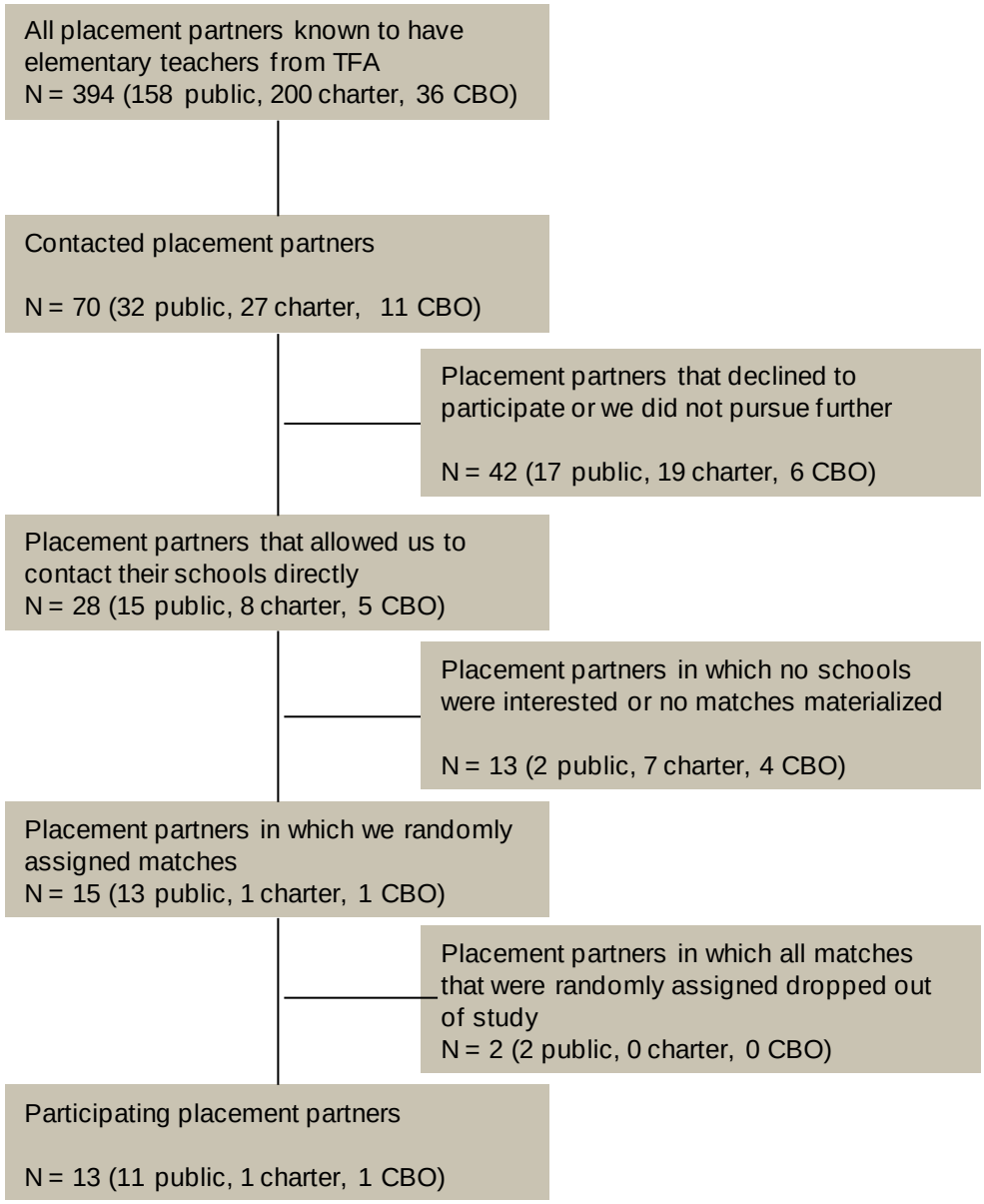
More than 50 percent of classroom matches consisted of one class taught by a TFA teacher and one class taught by a comparison teacher (Table A.1). Almost 30 percent included three teachers, and all but one of these matches included one TFA teacher and two comparison teachers. The remaining matches included more than three teachers: one match included multiple TFA teachers and one comparison teacher, five matches included one TFA teacher and multiple

[25] Nineteen matches were dropped at this point; 18 matches in the 10 schools that were dropped and one match in a school that stayed in the study with other viable matches.

[26] Six matches were dropped at this point; three matches in the two schools that were dropped and three matches in schools that stayed in the study with other viable matches.

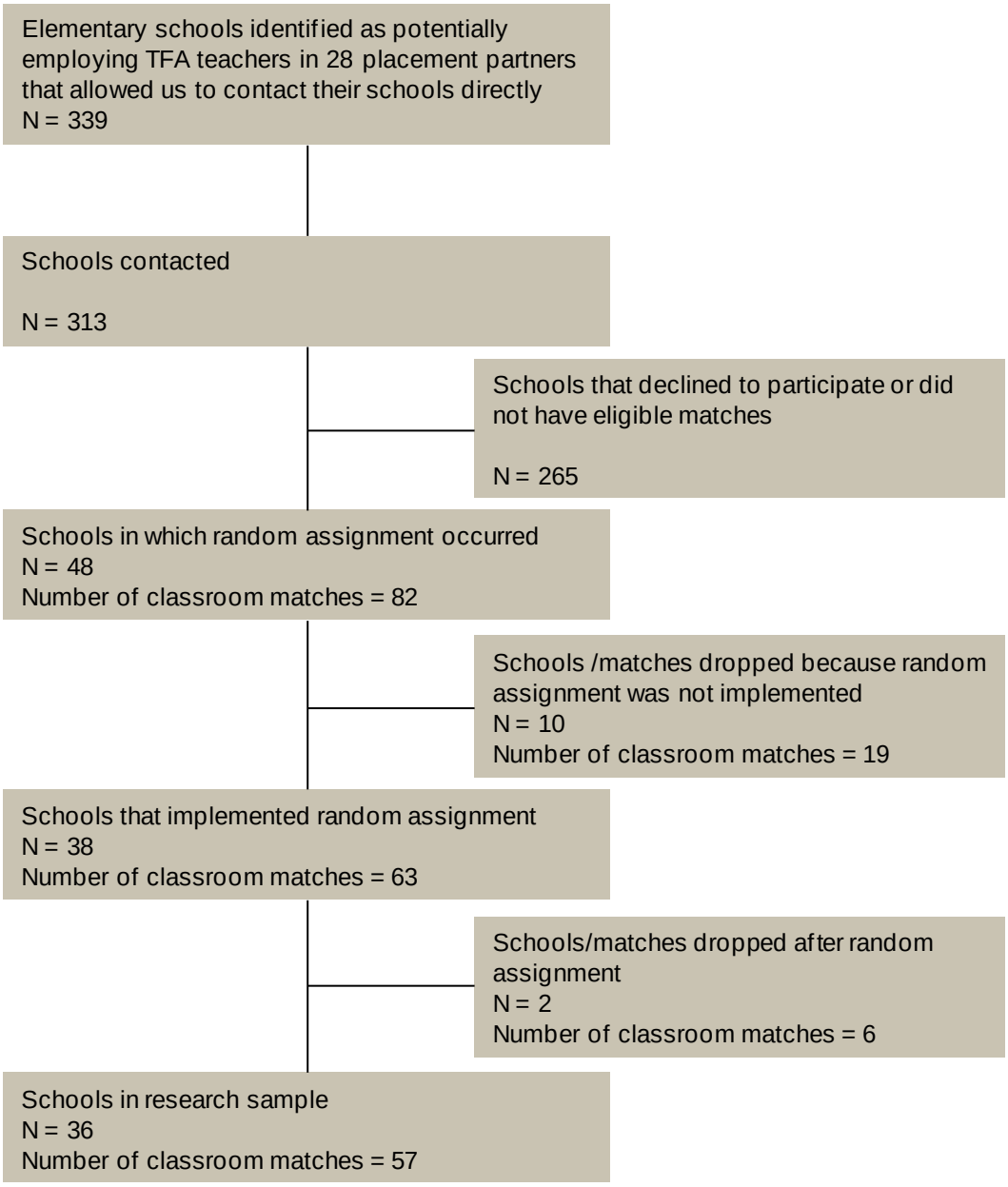
comparison teachers, and four matches included both multiple TFA teachers and multiple comparison teachers.

Figure A.1. District recruiting



CBO = Community-based organization; TFA = Teach For America.

Figure A.2. School recruiting



TFA = Teach For America.

Table A.1. Structure of classroom matches in the sample

Number of TFA and comparison classes in the classroom match	Number of classroom matches
1 TFA class, 1 comparison class	31
1 TFA class, 2 comparison classes	15
2 TFA classes, 1 comparison class	1
Other structures	10
Multiple TFA classes, 1 comparison class	1
1 TFA class, multiple comparison classes	5
Multiple TFA classes, multiple comparison classes	4
Total number of classroom matches	**57**

Source: Mathematica evaluation tracking system.

TFA = Teach For America.

B. Selection and assignment of students

All students who enrolled in a study class before the start of the school year or in the first two weeks of school were potentially eligible for random assignment and inclusion in the study sample. We conducted initial random assignment in summer 2012, which was the summer preceding the study school year, as soon as schools were able to provide student lists for assignment. After this initial random assignment, we assigned any additional students who needed to enroll in a study class through a process we referred to as rolling random assignment. Eighty-four percent of randomly assigned students were assigned via initial random assignment and 16 percent via rolling random assignment.[27] Below we describe these two random assignment procedures, our process for verifying that schools properly implemented the assignments, and the final student sample.

1. Initial random assignment

We conducted initial random assignment using the study's sample management system. To accommodate schools' needs to ensure balance in particular student characteristics across classes, we allowed them to specify up to three categorical variables on which to stratify the assignments. If the school did not request any stratifiers, we stratified on gender. The range of variables on which schools requested stratification included gender, race, ethnicity, academic ability, special education status, ELL status, age, and behavior. We also accommodated a limited number of special requests from the school—113 in all—to exempt particular students from random assignment and place them in a particular class.

If there were no exemptions from random assignment in a match, students assigned during initial random assignment had equal probabilities of assignment to each class in a match. The probability of assignment to a particular group (treatment or control) was thus equal to the number of classes in that group divided by the total number of classes in the match. For example, in a match with one class taught by a TFA teacher and two classes taught by comparison

[27] Because assignment probabilities to the treatment and control groups in a given match might have varied for students assigned via either procedure, we developed sample weights to adjust for differential assignment probabilities in the analysis, as discussed in Section E of this appendix.

teachers, a given student would have a 1/3 = 0.33 probability of being assigned to the TFA teacher (the treatment group) and a 2/3 = 0.67 probability of being assigned to the comparison teachers (the control group).

The only exceptions to the simple scenario described here occurred when a school required that a particular student or students be placed with a particular teacher. In these cases, the excluded students were placed in the required classes and then the remaining students in each stratum were randomly assigned to the remaining slots in the match. Within a given stratum, randomly assigned students' probabilities of assignment to a given group (treatment or control) were equal to the number of available slots for that stratum in that group (after the excluded students had been placed) divided by the total number of slots for that stratum in the match (again after the excluded students had been placed). For example, if a given match had one treatment and two control classes and no stratification, with a total of 60 students to be assigned to the classes, two of whom had to be placed in the treatment class, the probability of assignment to the treatment group for randomly assigned students would have been (20-2)/(60-2) = 0.31, and the probability of assignment to the control group would have been (40)/(60-2) = 0.69.

The probability of assignment to the treatment group in a given match and stratum is summarized by the following formula, with the probability of assignment to the control group determined in a parallel manner:

$$(A.1) \qquad pr(T_s) = \left[\left[\left(\frac{N_t}{N}\right) * \left(n_s + f_{t,s} + f_{c,s}\right)\right] - f_{t,s}\right] * \left(\frac{1}{n_s}\right)$$

where $pr(T_s)$ is the probability of assignment to the treatment group for a student in stratum s, N_t is the number of treatment group classes in the match, N in the total number of classes in the match, n_s is the number of students in the stratum to be randomly assigned in that match, $f_{t,s}$ is the number of students in the stratum forced to the treatment group, and $f_{c,s}$ is the number of students in the stratum forced to the control group. In the simple case in which no students are nonrandomly placed into a particular class, the formula reduces to the number of treatment classes divided by the total number of classes in the match.

2. Rolling random assignment

After we conducted initial random assignment, we assigned any late-enrolling students, either individually or in small batches, in a process we referred to as rolling random assignment. We gave school staff a hotline number to call for each new student's class assignment. Study staff entered information on newly enrolling students into an Excel form; students were then randomly assigned via an embedded Visual Basic program. We did not stratify these late assignments. We conducted rolling random assignment through the first two weeks of classes; after that time, we allowed schools to assign new students to classes as they chose. We excluded students who enrolled after the first two weeks of school from the study sample.

Because rolling random assignment occurred in the first two weeks of school, at a time when there was movement into and out of classes, class sizes were often not perfectly equal. To correct for any class size imbalances that existed at the time of rolling random assignment, we

constructed the rolling random assignment program to give students a greater probability of being assigned to smaller classes. Our approach was as follows:

- If the number of students to be assigned was greater than or equal to the number needed to equalize class sizes, all classes with fewer than the maximum number of students would be given the number of slots required to bring the class size to the maximum class size in the match, plus one. The largest class(es) in the match would (each) be given one slot. If the number of students to be assigned exceeded this number of slots, additional slots would be evenly distributed between all matches until there were enough slots for all students. The students would then be randomly assigned between these slots. For example, if a match had three classes—TFA class A with 20 students, control class B with 22 students, and control class C with 25 students—and there were 8 students to be assigned, class A would be given 6 slots, class B would be given 4 slots, and class C would be given one slot. The newly enrolling student or students would be randomly assigned between the available slots with equal probability of being assigned to a given slot (because there were fewer students than slots in this example, not all slots would be filled). Thus, the probability of assignment to the TFA class (class A) would be 6/(6+4+1) = 6/11 = 0.55, and the probability of assignment to the control group (class B or C) would be 4/11 +1/11 = 5/11= 0.45.
- If the number of students to be assigned was less than the number needed to equalize class sizes, we increased the probability of assignment to the smaller classes. Specifically, all classes with fewer than the maximum number of students would be given the number of slots required to bring the class size to the maximum class size in the match, plus one, and then this number would be multiplied by three (a factor that was chosen arbitrarily to increase the probability of assignment to the smaller classes). The largest class(es) in the match would (each) be given one slot. Then students would be randomly assigned between these slots. For example, if a match had three classes—TFA class A with 20 students, control class B with 22 students, and control class C with 25 students—and there were two students to be assigned, class A would be given 6*3 = 18 slots, class B would be given 4*3 = 12 slots, and class C would be given one slot. The newly enrolling student or students would be randomly assigned between the available slots with equal probability of being assigned to a given slot. Thus, the probability of assignment to the TFA class (class A) would be 18/(18+12+1) = 18/31 = 0.58, and the probability of assignment to the control group (class B or C) would be 12/31 +1/31 = 13/31= 0.42.

3. Roster verification

Immediately after we conducted initial random assignment, we asked schools to send us updated rosters so we could verify that they had properly implemented the assignments. If we identified students who were not in their assigned classes, we followed up with the school and asked them to move the students to the correct classes. In some cases, schools moved misplaced students to their study-assigned classes (and confirmed this move with an updated roster); in other cases, they failed to move the students. We considered random assignment to have been properly implemented in a match if at least 75 percent of randomly assigned students were in their assigned classes at the time of the initial roster verification. If more than 25 percent of students were not in their assigned classes at the time of initial verification, we classified the match as having refused to implement the randomly assigned rosters and dropped it from the study sample. We dropped 19 of the 82 matches (10 of the 48 schools) in which we conducted

random assignment because the school failed to implement the assignments. After the initial roster verification, we requested updated rosters at three other points during the study school years—in the fall, in the first week of classes in the spring, and then toward the end of the spring semester. We used these rosters to monitor the integrity of random assignment and the extent to which students left or were added to classes as well as to help locate study students for assessment in the spring.

4. Student sample

We randomly assigned 3,724 students to study classes in the 57 classroom matches, during either initial or rolling random assignment (Figure A.3). An additional 113 students enrolled in study classes during the random assignment period but were exempted from random assignment and placed in a specific class at the school's request. About 41 percent of randomly assigned students, or 1,544 students, were assigned to classes taught by Teach For America teachers, and about 59 percent, or 2,180 students, were assigned to classes taught by comparison teachers. Because classes in some study schools were departmentalized, in 4 of the 57 classroom matches, teachers taught either math or reading, but not both. We only included students in the impact analysis for the subject covered in their classroom match, resulting in a sample of 3,590 students in math and 3,679 students in reading.

Most schools sent their rosters for random assignment before enrollment was completely finalized, so some students (15 percent of those randomly assigned to the treatment group and 18 percent of those randomly assigned to the control group) never enrolled in the study school. Consistent with the research review standards used by the U.S. Department of Education's What Works Clearinghouse (U.S. Department of Education 2014) to calculate attrition rates, we included all randomly assigned students in the denominator, whether or not they actually enrolled at the study school.

We attempted to obtain parental consent for enrolled students to participate in the study in all districts. Nine of the school districts in the study required us to obtain active parental consent to assess students or obtain their school records data, meaning parents had to return a signed form providing consent for their child to participate.[28] In the remaining four districts, we sent parents a letter describing the study and providing them the opportunity to decline their child's participation. Across all districts, we obtained parental consent for 76 percent of randomly assigned students who enrolled in study schools. Consent rates were similar for students in the treatment and control groups (74 and 78 percent, respectively). The consent form did not indicate whether the child had been assigned to the treatment or the control group, although some parents may have known whether or not their child had been assigned to a TFA teacher at the time they signed the form.

We attempted to collect school records and outcome test score data for all 2,363 students (971 treatment and 1,392 control) whose parents consented for them to participate in the study.

[28] Although federal law, including the Family Educational Rights and Privacy Act, did not require parental consent for participation in this study, many school districts had policies that required us to obtain active parent consent to assess students, obtain their school records, or both.

We administered the Woodcock-Johnson assessments to students in prekindergarten through grade 2 and obtained test score data from district records from students in grades 3 through 5. We successfully obtained outcome test score data in reading for 2,123 students (90 percent of students with parental consent). We obtained test score data in math for 1,182 students (50 percent of students with parental consent). We included all students for whom we obtained data in our impact analysis.[29]

As expected because of random assignment, baseline characteristics were similar for treatment and control group students who had been randomly assigned. Of the 12 characteristics we examined in Table A.2, only one differs between the two groups by a statistically significant margin, and this difference is relatively small: 1.0 percent of the treatment group students were Asian, compared with 2.6 percent of control group students. As shown in Table II.5, baseline characteristics of treatment and control group students included in the analysis (that is, randomly assigned students with outcome test score data) were also similar, again with a statistically significant difference only for the percentage of Asian students. This suggests that differential attrition did not result in any apparent bias—even though we were not able to obtain test score data for all students who had been randomly assigned, students in the treatment and control groups in the final analysis remained balanced in terms of their baseline characteristics.

Not all students remained in the class to which they were originally assigned. Most of the randomly assigned students (68 percent of the treatment group and 66 percent of the control group) stayed in their originally assigned class for the full study year (Table A.3). A small percentage of students (about 3 percent) "crossed over" to a class taught by the opposite type of teacher (TFA or comparison) or moved to a class in the same match taught by the same type of teacher (about 1 percent). About 1 percent moved to a nonstudy class in the same school, and the remaining 27 percent left the school entirely or never enrolled.

[29] The main reason for the differences in rates of test score data collection between math and reading was an error in the administration of the Woodcock-Johnson tests in math (described below), which caused us to be without math outcome data for students in prekindergarten and kindergarten.

Figure A.3. Number of students involved in each stage of random assignment and data collection

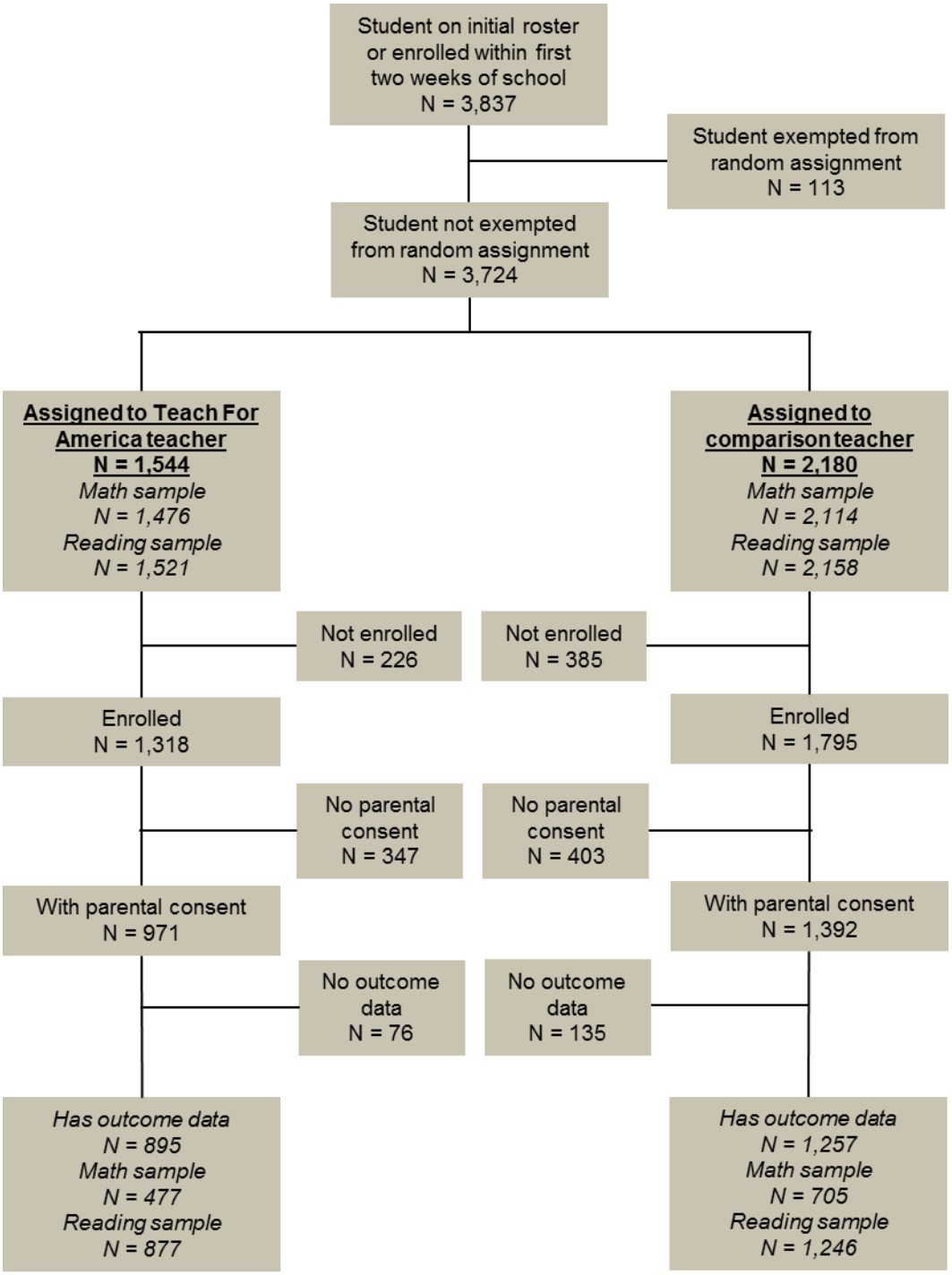


Table A.2. Average baseline characteristics of students assigned to TFA teachers or comparison teachers (percentages unless otherwise indicated), math and reading samples

Characteristic	All students	Assigned to TFA teachers	Assigned to comparison teachers	Difference between TFA and comparison	*p*-Value
Baseline math score (average z-score)	-0.1	-0.2	0.0	-0.2	0.219
Baseline reading score (average z-score)	-0.3	-0.3	-0.2	0.0	0.769
Age (average years)	6.8	6.8	6.8	0.0	0.613
Female	47.4	47.3	47.6	-0.3	0.908
Race					
Asian, non-Hispanic	1.8	1.0	2.6	-1.6**	0.003
Black, non-Hispanic	46.7	47.2	46.2	1.0	0.542
Hispanic	40.7	41.4	40.0	1.4	0.413
White, non-Hispanic	7.8	7.9	7.7	0.2	0.870
Other, non-Hispanic	3.0	2.5	3.5	-1.0	0.180
Eligible for free/reduced-price lunch	83.8	84.5	83.1	1.4	0.328
Limited English proficiency	32.7	32.6	32.8	-0.2	0.908
Individualized education plan	7.3	8.3	6.2	2.2	0.075
Number of students	**3,724**	**1,544**	**2,180**		
Number of teachers	**156**	**66**	**90**		
Number of classroom matches	**57**	**57**	**57**		
Number of schools	**36**	**36**	**36**		

Sources: District administrative records and study-administered Woodcock-Johnson assessments.

Note: Means and percentages are weighted with sample weights and adjusted for classroom match fixed effects; *p*-values are based on a regression of the specified characteristic on a TFA indicator and classroom match indicators, accounting for sample weights and clustering at the teacher level.

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

TFA = Teach For America.

Table A.3. Movement of randomly assigned students during the school year (percentages unless otherwise indicated)

Mobility status	All students in research sample	Assigned to TFA teachers	Assigned to comparison teachers
Stayed in originally assigned class through end of year	67.2	68.4	66.3
Crossed over to study class with opposite teacher type	3.3	4.2	2.6
Switched to another study class with same teacher type before end of year	1.2	0.3	1.9
Switched to nonstudy class in same school before end of year	1.5	1.4	1.6
Left study school before end of year	26.8	25.8	27.5
Number of students	**3,724**	**1,544**	**2,180**

Source: Mathematica evaluation tracking system.

TFA = Teach For America.

Because we allowed schools to place newly enrolling students in the study classes without random assignment after the first two weeks of school, about 25 percent of the students in the study classes at the end of the year had not been randomly assigned. We examined the baseline characteristics of students enrolled in study classes at the end of the school year who were not randomly assigned to see whether schools had systematically placed particular types of students with either TFA or comparison teachers. We found no statistically significant differences between the two sets of students (Table A.4).

Table A.4. Characteristics of nonstudy students on end-of-year rosters of classrooms in the TFA study sample (percentages unless otherwise indicated), math and reading samples

Characteristic	TFA classes	Comparison classes	Difference between TFA and comparison	*p*-Value
Baseline math score (average z-score)	-0.4	-1.1	0.6	0.113
Baseline reading score (average z-score)	-0.8	-0.7	-0.1	0.889
Age (average years)	6.9	6.9	0.0	0.721
Female	53.2	49.9	3.3	0.574
Race				
Asian, non-Hispanic	3.0	2.0	0.9	0.606
Black, non-Hispanic	46.5	50.3	-3.8	0.423
Hispanic	34.8	26.7	8.1	0.097
White, non-Hispanic	10.6	14.8	-4.2	0.284
Other, non-Hispanic	5.1	6.1	-1.0	0.690
Eligible for free/reduced-price lunch	81.7	82.7	-1.0	0.820
Limited English proficiency	30.2	22.6	7.6	0.112
Individualized education plan	4.3	8.9	-4.6	0.113
Number of students	**105**	**116**		
Number of teachers	**41**	**51**		
Number of classroom matches	**19**	**23**		
Number of schools	**13**	**15**		

Source: District administrative records.
Note: Means and percentages are adjusted for classroom match fixed effects. None of the differences is statistically significant at the 0.05 level, two-tailed test.

C. Response rates

1. Response rates for students

On average, we had valid outcome test score data (from either state assessments or the Woodcock-Johnson tests) for 33 percent of randomly assigned students in math and 58 percent of randomly assigned students in reading (Table A.5). The difference between math and reading was driven by the error in administering the Woodcock-Johnson Applied Problems assessment to prekindergarten and kindergarten students. Because of this error, we have no outcome data for either the treatment or control group in math for these students. Apart from this, response rates were similar at different grade levels. Within each subject, average response rates for the treatment and control groups were also similar.

Table A.5. Student response rates, by subject and grade level (percentages unless otherwise indicated)

Type of impact estimate to which the student's classroom contributes	Assigned to TFA teachers	Assigned to comparison teachers	Total
Math	32.3	33.3	32.9
Early childhood students (pre-K and K)	0	0	0
Lower elementary (grades 1 and 2)	58.2	57.2	57.5
Upper elementary (grades 3–5)	59.3	59.8	59.6
Reading	57.7	57.7	57.7
Early childhood students (pre-K and K)	57.1	55.6	56.2
Lower elementary (grades prekindergarten–2)	57.6	56.7	57.0
Upper elementary (grades 3–5)	58.0	62.1	60.3

Source: Mathematica evaluation tracking system.

TFA = Teach For America.

As shown earlier in Figure A.3, overall student response rates depended on whether students who were randomly assigned actually enrolled in the study school, whether their parents consented for them to participate in the study, and whether we were able to obtain their outcome test score data. About 84 percent of randomly assigned students (85 percent of the treatment group and 82 percent of the control group) enrolled in the study schools. Overall, on average, we obtained parental consent for 64 percent of students who had been randomly assigned (63 percent for the treatment group and 64 percent for the control group). For reading we obtained parental consent and valid outcome test score data for 58 percent of randomly assigned students, for both treatment and control groups. For math, we obtained parental consent and valid outcome test score data for 32 percent of randomly assigned students (31 percent of the treatment group and 32 percent of the control group).

Randomly assigned students without valid outcome data differed from students with valid outcome data in a few ways (Table A.6). On average, those with valid outcome data had higher baseline test scores, although these differences were not statistically significant. Students with valid outcome data were less likely to be white, non-Hispanic and more likely to be Hispanic and to have limited English proficiency relative to students without valid outcome data. Students without valid outcome data were more likely to have an IEP, although this difference was not statistically significant at the 5 percent level.

2. Response rates for teachers

Response rates for the teacher survey were slightly higher for TFA teachers than for comparison teachers. Ninety percent of TFA teachers and 85 percent of comparison teachers completed the survey, for an overall response rate of 87 percent.

Table A.6. Characteristics of randomly assigned students with and without outcome data (percentages unless otherwise indicated), math and reading samples

Characteristic	Students with outcome data	Students without outcome data	Difference between students with and without outcome data	*p*-Value
Baseline math score (average z-score)	-0.1	-0.9	0.8	0.136
Baseline reading score (average z-score)	-0.2	-0.8	0.5	0.262
Age (average years)	6.8	6.8	0.0	0.601
Female	47.2	49.9	-2.7	0.466
Race				
Asian, non-Hispanic	1.8	2.9	-1.2	0.240
Black, non-Hispanic	46.1	46.5	-0.4	0.866
Hispanic	41.9	31.9	10.0**	0.002
White, non-Hispanic	7.3	13.3	-5.9**	0.002
Other, non-Hispanic	2.9	5.3	-2.4	0.051
Eligible for free/reduced-price lunch	83.7	83.1	0.6	0.790
Limited English proficiency	33.8	23.2	10.5**	0.000
Individualized education plan	6.6	9.9	-3.3	0.075
Number of students	**2,152**	**1,572**		
Number of teachers	**156**	**156**		
Number of classroom matches	**57**	**57**		
Number of schools	**36**	**36**		

Source: District administrative records.

Note: Means and percentages are adjusted for classroom match fixed effects.

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

D. Statistical power

To examine the statistical power of our sample to detect impacts, we computed minimum detectable effects based on the standard error of the treatment effects we obtained. The minimum detectable effect is the smallest true impact for which there would be an 80 percent probability of obtaining a statistically significant estimate. The minimum detectable effect for the full sample was 0.15 standard deviations for math and 0.14 standard deviations for reading (Table A.7). That is, if students truly scored at least 0.15 standard deviations higher in math because of being assigned to a TFA teacher rather than a comparison teacher, then any study with the same design and the same population of teachers would have at least an 80 percent probability of obtaining a statistically significant impact estimate. These minimum detectable effects are about the same as the 0.15 standard deviation impact that the first random assignment study of TFA elementary school teachers (Decker et al. 2004) found for TFA teachers' effectiveness in math. Minimum detectable effect for impacts within subgroups were higher because sample sizes were smaller and ranged from 0.16 to 0.47 standard deviations for math and 0.15 to 0.34 standard deviations for reading.

Table A.7. Minimum detectable effects

Sample	Math		Reading	
	Analysis sample size	Minimum detectable effect	Analysis sample size	Minimum detectable effect
Full sample	1,182	0.15	2,123	0.14
Early childhood	0	n.a.	878	0.34
Lower elementary	770	0.22	1,653	0.16
Upper elementary	412	0.20	470	0.23
Novice comparison teachers	170	0.47	313	0.33
Comparison teachers with traditional certification	1,078	0.16	1,884	0.15

Source: District administrative records and study-administered Woodcock-Johnson assessments.

Note: Minimum detectable effects are expressed in standard deviations of outcome test scores within the reference population of the student's assessment. Minimum detectable effect = $2.802 * \text{standard error of treatment effect}$. Early childhood includes prekindergarten and kindergarten—we do not have math test scores for these students. Lower elementary includes grades 1–2 for math and grades prekindergarten through 2 for reading. Upper elementary includes grades 3–5.

E. Sample weights

We weighted the impact estimates to account for two issues: (1) different random assignment probabilities within each classroom match and (2) discrepancies between the characteristics of TFA teachers in our sample and the overall population of TFA teachers.

Probability of assignment to the treatment group or control group was generally equal for all students in a classroom match (for instance, in a two-classroom match, students typically had a 0.5 probability of assignment to the treatment group) but was adjusted for students who were assigned after school began to help balance class sizes. For instance, as described in Section B of this appendix, if a late-enrolling student could be assigned to a treatment classroom with 18 students or a control classroom with 22 students, we increased the probability of assignment to the treatment classroom above 0.5, and the sample weight for that student reflected his or her higher probability of assignment to the treatment group.

To calculate these weights, we first constructed a raw weight, equal to the inverse of the probability of assignment to the group (treatment or control) to which each student was actually assigned:

$$\text{(A.2)} \qquad raw\_weight_{igk} = \frac{1}{p_{igk}},$$

where raw_weight_{igk} is the raw weight for student i in group (treatment or control) g and match k and p_{igk} is the student's *ex ante* probability of being assigned to the group g to which he or she was actually assigned.

For math and reading separately, we then normalized the raw weights so that the sum of the normalized weights within a match equaled the total number of randomly assigned students in

the match, with the sum of the weights among treatment group students equal to the sum of the weights among control group students:

$$\text{(A.3)} \qquad sample\_weight_{igk} = \left(\frac{raw\_weight_{igk}}{\sum_{i=1}^{N_{gk}} raw\_weight_{igk}} \right) * \left(\frac{N_k}{2} \right)$$

where $sample\_weight_{igk}$ is the final sample weight for student i in group g and match k, N_{gk} is the total number of randomly assigned students assigned to group g in match k, and N_k is the total number of randomly assigned students in match k.

We also established poststratification weights to rescale each classroom match such that the proportion of students of TFA teachers in the weighted sample equaled the proportion of total students taught by TFA teachers nationally in the 2012–2013 school year, by TFA cohort and grade span. There were two cohorts of TFA teachers in the study (those who started teaching in fall 2011 and those who started in fall 2012) and three grade spans (prekindergarten to kindergarten, grades 1 to 2, and grades 3 to 5). To create the poststratification weights, we first created 2 (cohorts) x 3 (grade spans) = 6 cells and then weighted them up to their population counterparts by dividing the population percentage by the sample percentage within each cell. For example, if the percentage of upper elementary, 2012 cohort TFA teachers in the population was 30 percent, and the corresponding percentage in the sample was 10 percent, we would create a poststratification weight of.30/.10 = 3 for these teachers. We created separate poststratification weights for math and reading. Students of comparison teachers received the same weight as students of the TFA teacher within the same classroom match. The final weight for each student was the product of the sample weight and the poststratification weight. We also conducted two sensitivity analyses using alternative weights, as explained in Appendix B.

F. Contextual analysis

To provide context for the impact analysis, we examined TFA's program model and implementation of the i3 scale-up as well as the schools, teachers, and students in the study sample.

1. TFA's program and implementation of the scale-up

To describe TFA's program model and its implementation of the i3 scale-up, we conducted semi-structured interviews with 17 members of TFA's senior staff that we summarized in narrative form. We also analyzed quantitative data provided by TFA, including admissions, training, and placement data, along with data from surveys it administered to all its corps members, to describe and examine changes over time in key elements of the program. The study's implementation report (Zukiewicz et al. 2015) provides more detail on this analysis.

2. Schools in the study

To describe the schools in the study, we compared the average characteristics of study schools to the average characteristics of all elementary schools with TFA teachers and all elementary schools nationwide using the Common Core of Data, Public Elementary/Secondary School Universe Survey, 2011–2012. For each comparison, we calculated the difference between

the groups and tested the statistical significance of the differences, using t-tests for binary and continuous variables and chi-squared tests for categorical variables.

3. Teachers in the study

To describe the teachers in the study, we documented and compared the characteristics of TFA and comparison teachers in the sample. We examined the teachers' background characteristics, teaching experience, preparation for teaching, support received throughout the school year, and attitudes toward teaching. For each characteristic, we calculated the difference in mean values between the two groups and tested the statistical significance of the differences.

4. Students in the study

We examined the characteristics of students in the study sample to document their demographic characteristics and to assess the integrity of random assignment. To assess the integrity of random assignment, we estimated treatment-control differences in several baseline student characteristics and tested the statistical significance of the differences.

G. Impact analysis

1. Main estimation model

The main model we estimated, separately for reading and math test scores, was

$$y_{ijk} = \alpha_{jk} + \lambda_k w_{ijk} + \beta X_{ijk} + \delta T_{ijk} + \varepsilon_{ijk}, \quad \text{(A.4)}$$

where y_{ijk} is the reading or math test score of student i in classroom match j taking baseline test k (the Woodcock-Johnson test or a particular state test); α_{jk} is a vector of classroom match fixed effects, w_{ijk} is the baseline test score for student i in classroom match j on test k; X_{ijk} is a vector of student characteristics; T_{ijk} is an indicator equal to one if the student was assigned to the treatment group and zero otherwise; ε_{ijk} is a student-level error term; and λ_k, β, and δ are parameters or vectors of parameters to be estimated. We allowed the coefficient on the baseline test score, λ_k, to vary by baseline test. The impact estimation model also included a set of binary variables indicating whether the value of a particular covariate was missing for a given observation. We estimated heteroskedasticity-robust standard errors (Huber 1967, White 1980) and adjusted for clustering at the teacher level (Liang and Zeger 1986). The estimate of δ is the estimated impact of TFA teachers on student achievement.

2. Outcomes

As described in Chapter II, we used a combination of state administrative tests of math and reading for students in grades 3 to 5 and administered Woodcock-Johnson tests to students in prekindergarten to grade 2.

We chose Woodcock-Johnson tests that were appropriate for the grade level of a given student. In reading, we administered the Letter-Word Identification subtest to students in prekindergarten to grade 2 and the Passage Comprehension subtest to students in kindergarten to grade 2. In math, we administered the Applied Problems subtest to students in prekindergarten to

grade two and the Calculation subtest to students in grades 1 and 2. Table A.8 shows the tests and subtests taken by students in the study at various grade levels.

Table A.8. Achievement tests by grade level

Test	Prekindergarten	Kindergarten	Grades 1–2	Grades 3–5
Reading				
Woodcock-Johnson				
Letter-word identification	X	X	X	
Passage comprehension		X	X	
State reading assessments				X
Math				
Woodcock-Johnson				
Applied problems	X[a]	X[a]	X[a]	
Calculation			X	
State math assessments				X

[a]Due to an error in the test administration procedures for the Woodcock-Johnson Applied Problems assessment, we were unable to use those scores in the analysis.

We administered the Woodcock-Johnson tests during students' regular class time in the last four weeks of the school year. We tested each child individually, with each subtest taking about five minutes to complete. To ensure comparable testing conditions among treatment and control classes, we tried to test all classes in a match at the same time on the same day. Testing staff were not aware of the teacher's route to certification (TFA or non-TFA).

We attempted to assess all early childhood and lower elementary school students in the sample, irrespective of whether they moved to other classes at the school, were absent on the day of testing, or transferred to other schools in the same school district. The only students we did not attempt to test, because of logistical challenges, were those who had transferred to schools in other districts. We invited students who switched classes within a school to attend the regularly scheduled test session and scheduled additional testing sessions as needed for students who were unable to attend the initial session. Mathematica staff also contacted other schools in the district where sample members had transferred to arrange to test to these students. In matches in which primary instruction in reading or math was in Spanish as of the end of the school year, we administered the Spanish-language versions of the tests in the relevant subject(s).

To scale the outcome variable comparably across all classroom matches, we converted the original scale scores to z-scores (original scores minus the mean score divided by the standard deviation of the scores). Student reading achievement was measured by the broad reading *W* score determined by the Woodcock-Johnson III Letter-Word Identification subtest for prekindergarten students and by the Letter-Word Identification and Passage Comprehension subtests for students in kindergarten to grade 2. Student math achievement was measured by the broad math *W* score determined by the Woodcock-Johnson III Calculation subtests for students in grades one and two. To create a population mean of broad *W* scores in grades in which students took two subtests, we calculated the overall mean by averaging the published means of the subtests. To calculate the standard deviation, we needed to know the correlation between subtests. Because published data on the correlation between the components were not available,

we substituted the observed correlation between subtests among students in the analysis. We combined this information with published standard deviations for each subtest.

Error in administering Woodcock-Johnson Applied Problems assessment. An error occurred when the assessors were administering the Woodcock-Johnson Applied Problems assessment. This error caused scores on the assessment to be inappropriately constrained, which may have prevented us from reliably measuring impacts on the assessment. As a result, we have excluded results from the Applied Problems assessment and rely only on other available assessment data for the evaluation.

The Woodcock-Johnson Applied Problems assessment, as correctly administered, contains 63 questions which increase in difficulty. Under established test administration procedures, the assessor begins the assessment with a pre-specified question that varies based on the student's grade level, with students at higher grade levels beginning with more difficult questions. The assessor then progresses through the remaining questions until the child answers six questions in a row incorrectly, at which point the assessment ends and a score can be assigned.

For the TFA-i3 evaluation, assessors administered the Applied Problems and other Woodcock-Johnson assessments to students individually. The assessment was programmed into laptop computers from which assessors read the questions and entered the students' responses. However, due to an error in programming specifications, the Applied Problems assessment stopped at the 29th item instead of allowing administration of the full 63. As a result, many students' scores were inappropriately constrained—they reached the end of the assessment before answering six items in a row incorrectly, and thus the score was not a valid estimate of their ability. We administered the assessment to students in prekindergarten through grade 2. The scores of 66 percent of these students (24 percent of prekindergarten students, 57 percent of kindergarteners, 88 percent of first graders, and 96 percent of second graders) were inappropriately constrained by the administration error, meaning that the students answered the 29th question and finished the assessment without having answered 6 questions in a row incorrectly.

3. Covariates

In the impact estimation, we controlled for several baseline student characteristics:

- Prior achievement in reading and math (regardless of whether the outcome test score was for reading or math)
- Eligibility for a free or reduced-price lunch
- Special education status or whether the student had an IEP
- Limited English proficiency status
- Gender
- Whether a student is black, non-Hispanic
- Whether a student is Hispanic

We accounted for prior achievement only when data were available from participating school districts. These test scores were available only for students in grades 4 and 5. Table A.9 shows a list of the coefficients from the baseline regression models for the full sample.

Table A.9. Coefficients on covariates in impact analysis, math and reading

Variable	Math	Reading
Assignment to TFA		
Teacher was TFA teacher	0.07 (0.05)	0.03 (0.05)
Pretest scores (average coefficients)		
Same-subject pretest score	0.55** (0.17)	0.48 (0.26)
Opposite-subject pretest score	0.34** (0.13)	0.40 (0.28)
Individual student background characteristics		
Eligible for free or reduced-price lunch	0.01 (0.13)	0.21 (0.11)
Special education	-0.42** (0.15)	-0.17 (0.41)
Limited English proficiency	-0.42** (0.13)	-0.46** (0.12)
Female	0.12 (0.07)	0.23** (0.07)
Asian, non-Hispanic	0.36 (0.20)	-0.13 (0.27)
Black, non-Hispanic	-0.32** (0.11)	-0.47** (0.12)
Hispanic	0.09 (0.13)	-0.12 (0.12)

Source: District administrative records and study-administered Woodcock-Johnson assessments.

Notes: Standard errors in parentheses. The table excludes coefficients for classroom match fixed effects and indicators for imputed data.

TFA = Teach For America.

**Coefficient is statistically significant at the 0.01 level, two-tailed test.

4. Missing data

We accounted for missing values of prior test scores and other baseline covariates using dummy variable adjustment (Puma et al. 2009). Under this approach, we set missing values of each covariate to the mean of that covariate within each classroom match; otherwise, if the variable was missing for all students within a classroom match (for instance, prior test scores in a state and grade in which there was no testing in the previous year), we set missing values equal to the sample mean. For each variable with missing values, we included in the impact estimation model an indicator variable equal to one if the value of the variable was missing for a given observation and zero otherwise.

As a sensitivity analysis, we imputed missing values of covariates using the multiple imputation by chained equation method (Raghunathan et al. 2001). The imputation model included all covariates included in the impact estimation model as well as the treatment

indicator, classroom match fixed effects, and outcome test score variables. We combined the estimates using the approach recommended by Rubin (1987) to account for the variability between imputations. We implemented multiple imputation for (1) missing student demographic data for students at any grade level and (2) missing test score data for students in classroom matches for which a majority of students had pretest data. For students in classrooms that lacked pretest data, we used the dummy variable adjustment approach outlined above.

5. Subgroup analyses

As noted in Chapter II, we estimated the impact of TFA teachers for five subgroups: (1) early childhood students (prekindergarten and kindergarten) in reading, (2) lower elementary students (prekindergarten to grade 2 in reading and grades 1 to 2 in math); (3) upper elementary students (grades 3 to 5); (4) TFA teachers compared with other novice teachers, defined as teachers in their first two years of teaching; and (5) TFA teachers compared with traditionally certified comparison teachers. To estimate subgroup impacts, we estimated Equation A.5:

$$(A.5) \qquad y_{ijk} = \alpha_{jk} + \lambda_k w_{ijk} + \beta_1 X_{ijk} + \beta_2 C_{ijk} + \delta_1 T_{ijk} + \delta_2 T_{ijk} * C_{ijk} + \varepsilon_{ijk},$$

where C_{ijk} is an indicator equal to one if the student's teacher was a member of subgroup C and zero otherwise. In the first subgroup analysis, this indicator represented teachers of prekindergarten and kindergarten. For the second and third subgroup analyses, the indicator represented students in prekindergarten to grade 2. In the fourth subgroup analysis, it represented novice comparison teachers and their TFA counterparts in the same classroom match, and in the fifth subgroup analysis it represented traditionally certified comparison teachers and their TFA counterparts in the same classroom match.[30] By summing the overall treatment effect δ_1 with the effect for the subgroup δ_2, we estimated the total treatment effect of members of the subgroup and tested its statistical significance.[31]

6. Adjusting for noncompliance with random assignment

Our estimates of the relative effectiveness of TFA teachers might have been understated because some students initially placed with TFA teachers transferred out of their class during the year, meaning that they did not receive a full year's worth of the "treatment." Table A.3 documents the number of students of TFA and comparison teaches who moved between and out of study classes. Our main impact estimates, known as "intent-to-treat" estimates, reflect the

[30] For the estimation of effects by grade level, we were unable to estimate β_2 because we were not be able to distinguish grade effects from classroom match effects (represented by α_k) because in this case the category defining the subgroup was assigned at the school-grade level. In addition, although in many cases there was only one novice non-TFA teacher matched with one TFA teacher, in other cases there were multiple non-TFA teachers in a classroom match, some of whom were novices and others of whom were experienced teachers. Therefore, for estimating the effect of novice TFA teachers relative to novice comparison teachers, we estimated β_2 alongside α_k because the category was defined at the teacher level instead of the school-grade level.

[31] When the indicator represents students in prekindergarten and kindergarten, the treatment effect equals $\delta_1 + \delta_2$ for early childhood TFA teachers. When the indicator represents students in prekindergarten to grade 2, the treatment effect equals $\delta_1 + \delta_2$ for lower elementary school TFA teachers and δ_1 for upper elementary school TFA teachers. In the novice and traditionally certified teacher cases, the treatment effect is $\delta_1 + \delta_2$.

impact of being *assigned* to a TFA teacher's class (whether or not the student actually complied with that assignment).

As a sensitivity analysis, we estimated the impact of being *taught* by a TFA teacher for a full year. To do this, we estimated a "complier average causal effect" by adjusting the estimates for student movement out of their assigned classes using instrumental variables estimation (Angrist et al. 1996). An instrumental variable predicts the variable of interest but is not otherwise related to the final outcome. In this case, whether a student was randomly assigned to a TFA teacher is an instrumental variable for being taught by a TFA teacher for the full year.

For students who left the entire set of study classes before we collected spring rosters, we did not know the type of teacher that they had at the time of testing. Therefore, we made two alternative sets of assumptions that led to lower- and upper-bound estimates for the complier average causal effect. First, we assumed that all students who left the study classes moved to a class taught by the same type of teacher (TFA or non-TFA) with which they were last observed before they left. Second, we assumed that all students who left the study classes were subsequently taught by the opposite type of teacher to their original assignment.

Formally, we estimated this system of equations:

(A.6) $$F_{ijk} = \alpha_{1jk} + \tau_{1k} w_{ijk} + \tau_2 X_{ijk} + \tau_3 T_{ijk} + \mu_{ijk}$$

(A.7) $$y_{ijk} = \alpha_{2jk} + \lambda_{2k} w_{ijk} + \beta_2 X_{ijk} + \delta_2 \hat{F}_{ijk} + \nu_{ijk}$$

In the first-stage equation (A.6), we regressed F_{ijk}, which represents being taught by a TFA teacher, on all of the other independent variables from the outcome equation (A.7) plus T_{ijk}, which represents being assigned to a TFA teacher. T_{ijk} is the instrumental variable in this system. In the second-stage (outcome) equation (A.7), we use the predicted value of F_{ijk}, which is generated from equation (A.6) by setting the error term μ_{ijk}, to zero. The results of this analysis are shown in Appendix B.

This page has been left blank for double-sided copying.

APPENDIX B: SENSITIVITY ANALYSES

This page has been left blank for double-sided copying.

In this appendix, we explore the sensitivity of our main impact estimates, presented in Chapter V, to various statistical assumptions. We refer to the main model we used to generate the results in Chapter V as our benchmark model. To explore the sensitivity of results from the benchmark model, we (a) estimated models that excluded matches in which a high proportion of students were exempted from random assignment, (b) excluded students who took the tests in Spanish, (c) modified the way we standardized end-of-year test scores, (d) allowed the relationship between student background characteristics and end-of-year achievement to vary across lower elementary and upper elementary school students, (e) changed our strategy for handling missing data, (f) used alternative approaches to weighting classroom matches, (g) estimated models that did not cluster standard errors at the teacher level, (h) dropped classes in which the original teacher left midyear and was replaced by a teacher of the opposite type (TFA or comparison) and (i) accounted for students who switched to a different type of teacher (TFA or comparison) from their originally assigned teacher. Below we describe each of these sensitivity analyses in more detail. We find that none of the sensitivity analyses alter our basic finding that TFA teachers hired during the first two years of the i3 scale-up are neither more nor less effective than comparison teachers in teaching both reading and math.

A. Excluding matches in which a high proportion of students was exempted from random assignment

In our benchmark model, we included all 57 classroom matches that assigned students to classes based on the results of the random assignment we provided at the start of the school year. As a sensitivity test, we excluded classes in which a high proportion of students (more than 20 percent) enrolled at the end of the school year had not been randomly assigned. As discussed in Appendix A, we allowed schools to request a limited number of exemptions from random assignments, for students who needed to be placed in a particular class, as long as the number of exemptions per class was less than 10 percent of the total class size. However, the percentage that was not randomly assigned could have increased after the start of the school year if schools failed to contact us to determine student assignments during the first two weeks of school or if students continued to enroll after the first two weeks, when the random assignment period had ended. Even though we excluded students who were not randomly assigned from the research sample, these students could have potentially affected their peers in ways that influenced our estimates of TFA teachers' effectiveness. For example, if particularly unruly students were placed in the classrooms of TFA teachers, this might depress the measured effectiveness of these teachers. To explore the sensitivity of our benchmark model to these potential peer effects, we reestimated the model excluding the classrooms in which 20 percent or more of students at the end of the school year had not been randomly assigned. The results, shown in row 2 of Table B.1 for math and Table B.2 for reading, indicate that the exclusion of these matches does not affect our main finding that TFA teachers had no statistically significant impact on student achievement in either subject.

B. Excluding students who were tested in Spanish

In our benchmark model, we included all randomly assigned students with outcome test score data as long as they took the test in the same language as the majority of the students in the classroom match (ensuring that both treatment and control students in each match all took the same test in the same language). Although the majority of students in the analysis sample were

tested in English, 4 percent were tested in Spanish in both reading and math on either the study-administered Woodcock-Johnson assessments or their state assessments. To explore the sensitivity of our findings to this decision, we reestimated the model without students who were tested in Spanish. Results (shown in row 3 of Tables B.1 and B.2) are similar to those from our benchmark model.

C. Changing our approach for standardizing end-of-year test scores

We measured teacher effectiveness based on students' end-of-year math and reading test scores. However, students took different tests depending on their grade (for students in prekindergarten through grade 2 who took the study-administered Woodcock-Johnson assessments) or grade and state (for students in grades 3 through 5 who took their state assessments). To standardize scores across all students in our sample, in our benchmark model we converted all test scores to a common metric known as a z-score, which measures the number of standard deviation units a student was above or below the average student in his or her grade, as described in Chapter II. Impacts on z-scores can be interpreted as effect sizes, a common metric used in education evaluations. To construct the z-scores for our benchmark model, we used the broadest possible reference groups—national norms for students taking the Woodcock-Johnson tests and all students in the same grade in the state for students taking state tests.

As an alternative method for constructing z-scores, we standardized by the means and standard deviations for students in the control group sample. This approach may be more appropriate if the distribution of achievement among the students served by TFA is systematically different from that of the broader reference population. The downside of this approach is that the estimated standard deviations based on the control group may be imprecise in cases where there are few test takers for a particular assessment, biasing the effect sizes (Hedges 1981). When we reestimated the results using z-scores based on the control group means and standard deviations, we saw no overall difference in math (row 4 of Table B.1) but in reading the impact increased to 0.08 and was marginally significant (p-value = 0.077), the only such finding across the entire range of sensitivity analyses. However, with so many tests, a single finding of marginal significance is not unusual (row 4 of Table B.2).

As another alternative for standardizing test scores, we avoided z-scores altogether and used a different metric known as the *W* score. A potential concern with using z-scores is that a unit of student learning represented by a standard deviation gain in one grade may not be equivalent to a unit of learning represented by a standard deviation gain in another grade. The *W* score is a measure from the Woodcock-Johnson assessment, which is designed to measure student learning in increments that are common across grade levels (vertically aligned test scores). We already had *W* scores for students in prekindergarten through grade 2, whom we assessed with the Woodcock-Johnson. To incorporate the tests of students in grades 3 to 5, we created pseudo-*W* scores using the following approach: (1) we collected data on the mean and standard deviation of *W* scores in math and reading for students whose age matched that of the modal student in each grade 3 to 5; then (2) we translated the z-score of students on state tests to an equivalent *W* score based on the same z-score but using the mean and standard deviation of the Woodcock-Johnson test for their subject and grade. This approach assumes that the variability of student achievement in the states in which participating districts were located was the same as the variability of student achievement of test takers in the national Woodcock-Johnson sample. Once

all scores had been put on the *W* score scale, we created z-scores using all students in the sample so that the impact estimate could be interpreted as an effect size. Results using this approach to standardizing student test scores (row 5 of Tables B.1 and B.2) are consistent with the results from our benchmark model.

Table B.1. Difference in effectiveness between TFA teachers and comparison teachers, alternative model specifications, math

Model	Impact (effect size)	*p*-Value	Sample sizes		
			Students	Teachers	Classroom matches
(1) Benchmark	**0.07**	**0.197**	**1,182**	**83**	**32**
(2) Excludes matches with many exemptions	-0.08	0.437	375	26	16
(3) Excludes Spanish-language test takers	0.07	0.207	1,169	81	31
(4) Uses control group norms for z-scores	0.04	0.572	1,182	83	32
(5) Uses pseudo-*W* scores as outcome	0.05	0.183	1,182	83	32
(6) Demographic relationships vary by grade range	0.07	0.211	1,182	83	32
(7) Uses multiple imputation	0.07	0.229	1,182	83	32
(8) Uses only random assignment probability weights	0.07	0.217	1,182	83	32
(9) Does not use any weights	0.07	0.233	1,182	83	32
(10) Does not use clustered standard errors	0.07	0.212	1,182	83	32
(11) Excludes classes with changes in teacher type	0.08	0.164	1,166	82	32
(12) Uses IV to estimate complier average causal effect	0.10	0.470	1,182	83	32

Source: District administrative records and study-administered Woodcock-Johnson assessments.

Note: None of the impact estimates is statistically significant at the 0.05 level, two-tailed test.

IV = instrumental variables estimation; TFA = Teach For America.

Table B.2. Difference in effectiveness between TFA teachers and comparison teachers, alternative model specifications, reading

Model	Impact (effect size)	*p*-Value	Sample sizes		
			Students	Teachers	Classroom matches
(1) Benchmark	**0.03**	**0.570**	**2,123**	**154**	**56**
(2) Excludes matches with many exemptions	-0.07	0.537	776	55	29
(3) Excludes Spanish-language test takers	0.02	0.765	2,041	148	53
(4) Uses control group norms for z-scores	0.08[+]	0.077	2,123	154	56
(5) Uses pseudo-*W* scores as outcome	0.03	0.256	2,123	154	56
(6) Demographic relationships vary by grade range	0.03	0.523	2,123	154	56
(7) Uses multiple imputation	0.03	0.513	2,123	154	56
(8) Uses only random assignment probability weights	0.06	0.142	2,123	154	56
(9) Does not use any weights	0.07	0.123	2,123	154	56
(10) Does not use clustered standard errors	0.03	0.677	2,123	154	56
(11) Excludes classes with changes in teacher type	0.02	0.704	2,091	152	56
(12) Uses IV to estimate complier average causal effect	0.04	0.668	2,123	154	56

Source: District administrative records and study-administered Woodcock-Johnson assessments.

[+]Difference is statistically significant at the 0.10 level, two-tailed test.

IV = instrumental variables estimation; TFA = Teach For America.

D. Allowing relationships between student achievement and student characteristics to vary by grade range

As discussed in Chapter II, because students were randomly assigned to classes, we do not need to adjust for their baseline characteristics to estimate unbiased impacts of TFA teachers; however, including covariates in the estimation model increases the precision of the estimates. In our benchmark model, we controlled for students' baseline characteristics and test scores but did not allow the relationship between these characteristics and the outcome test scores to vary by students' grade level. As an alternative approach, we allowed the relationships between student achievement and baseline variables to vary by grade range, with separate relationships estimated for lower elementary and upper elementary students. This approach could produce more accurate estimates of the relationships between baseline variables and outcome test scores, if there are systematic differences across the two grade ranges, but could provide less precise estimates if there are not systematic differences. When we followed this alternative approach (row 6 of Tables B.1 and B.2), we found the same general results as in the benchmark model.

E. Changing the strategy for addressing missing data

For our benchmark model, when student baseline data provided by participating school districts were incomplete, we set missing values of covariates to the mean value in the classroom match and included dummy variables indicating whether data were missing for each covariate, an approach recommended by Puma et al. (2009). However, this approach may overestimate the precision of the model because we have not accounted for the uncertainty of the imputation approach. An alternative strategy, known as multiple imputation, accounts for the uncertainty in imputation so as not to overstate the precision of the results, as explained in Appendix A. However, when we implemented multiple imputation (row 7 of Tables B.1 and B.2), results did not change appreciably from the benchmark model.

F. Changing the weight given to individual students in the sample

As discussed in Appendix A, in our benchmark model we used sample weights that adjusted for the probability that a student was assigned to a particular teacher and then rescaled the observations to better reflect the national distribution of TFA elementary school teachers in terms of corps year and grade level taught during the 2012–2013 school year. A drawback of using sample weights is that that they tend to reduce the precision of the impact estimates. To gain more precise results, first we reestimated the model using weights that adjusted for assignment probabilities but did not rescale observations to reflect the national distribution of TFA teachers. The results (row 8 of Tables B.1 and B.2), which reflect the effectiveness of TFA teachers in our sample without generalizing to some broader population, are similar to those from the benchmark model. We also estimated the model with no weights (row 9 of Tables B.1 and B.2); results from the unweighted model are also similar to the benchmark model.

G. Not clustering standard errors at the teacher level

In our benchmark model we estimated standard errors that accounted for clustering of student characteristics at the teacher level (Liang and Zeger 1986). Clustering adjusts for the fact that our sample of TFA teachers was drawn from the larger population of TFA corps members teaching in the study school year and is consistent with our use of poststratification weights to

adjust for the overrepresentation of second-year corps members and early childhood teachers in the sample. However, because the sample was not randomly drawn from the broader population of TFA teachers, clustering is not necessarily required. To examine how clustering affected the statistical significance of the results, we reestimated the model without clustering and found that the estimated impacts were still not statistically significant (row 10 of Tables B.1 and B.2).

H. Accounting for teacher turnover

Our benchmark model includes all study classes, classified according to the TFA status of the original teacher, including two classes in which the original teacher left midyear and was replaced by a teacher of the opposite type (one class in which a TFA teacher was replaced by a non-TFA teacher, and one class in which a non-TFA teacher was replaced by a TFA teacher). To examine the sensitivity of our findings to this decision, we reestimated the model without these two classes. Results from this approach (row 11 of Tables B.1 and B.2) are similar to those from the benchmark model.

I. Accounting for student mobility and crossover

Our benchmark model estimates the effect of being assigned to a TFA teacher, regardless of whether the student remained with that teacher for the full school year or transferred to a class taught by a non-TFA teacher—this is known as an intent-to-treat analysis. To examine the effect of being taught by a TFA teacher for the full school year, we estimated complier average causal effects, as described in Appendix A. Results from this approach (row 12 of Tables B.1 and B.2) are similar to those from the benchmark model.[32]

[32] We estimated this model two ways, to provide upper and lower bound estimates, making different assumptions about how to assign students when data on teacher assignments at the end of the year were unavailable. For both subjects, we obtained the same point estimate to two decimal places and *p*-value to three decimal places, regardless of which assumption we made.

This page has been left blank for double-sided copying.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA
Policy Research