The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs

September 2013

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education

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

The Effectiveness of Secondary Math Teachers from Teach For America and the Teaching Fellows Programs

September 2013

Melissa A. Clark
Hanley S. Chiang
Tim Silva
Sheena McConnell
Kathy Sonnenfeld
Anastasia Erbe
Mathematica Policy Research

Michael Puma
Chesapeake Research Associates

Elizabeth Warner
Project Officer, July 2012–October 2013
Institute of Education Sciences

Stefanie Schmidt
Project Officer, September 2008–June 2012
Institute of Education Sciences

NCEE 2013-4015
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Ruth Curran Neild
Commissioner

September 2013

The report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0112/0009. The project officer is Elizabeth Warner in the National Center for Education Evaluation and Regional Assistance.

Clark, Melissa A., Hanley S. Chiang, Tim Silva, Sheena McConnell, Kathy Sonnenfeld, Anastasia Erbe, and Michael Puma. (2013). *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.

This report is available on the IES website at http://ies.ed.gov/ncee.

ACKNOWLEDGMENTS

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 national and regional staff at Teach For America and TNTP, who provided essential information about their programs over the course of the study.

We are grateful for input provided by members of the study's technical working group, which included Julie Greenberg, Paul Holland, Tim Sass, Jeff Smith, Suzanne Wilson, and Jim Wyckoff. The study has benefited greatly from their expertise.

The study also benefited from the contributions of many individuals at Mathematica and its research partners, Chesapeake Research Associates and Branch Associates. Daniel Player and Alison Wellington were heavily involved in the early stages of the study and provided invaluable input on the study design, development of data collection instruments, and random assignment. Eric Zeidman and Charlotte Cabili led our large team of dedicated staff who recruited districts and schools into the study. Nancy Duda and Barbara Kennen were responsible for the collection and preparation of much of the data used in the study. Mason DeCamillis and Matt Jacobus provided excellent programming assistance in cleaning and analyzing the data, with expert guidance from Mary Grider. Phil Gleason provided thoughtful, critical reviews of the study design and final report. John Kennedy provided editorial support, and Jennifer Baskwell and Allie Clifford provided production support.

This page has been left blank for double–sided copying

CONTENTS

EXECUTIVE SUMMARY xix

I INTRODUCTION 1

A. Background 2

1. Concerns about Teacher Quality in Hard-to-Staff Schools and Subjects 2
2. Policies Targeting the Teacher Pipeline 2
3. TFA and the Teaching Fellows Programs 3
4. Past Research on Teachers from TFA and the Teaching Fellows Programs 4

B. Our Evaluation 5

C. Looking Ahead 7

II STUDY DESIGN, DATA, AND METHODS 9

A. Types of Teachers in the Study 9

1. Eligible Routes to Certification 9
2. No Limits on Prior Teaching Experience 10

B. Experimental Design 11

C. Recruitment of Districts, Schools, and Teachers 12

1. Recruitment of Districts 13
2. Recruitment of Schools 14
3. Classroom Matches and Teachers in the Final Study Sample 15
4. Differences between the TFA and Teaching Fellows Study Samples 15

D. Selection and Assignment of Students 16

E. Data Used in the Study 18

1. Data on Students 18
2. Data on Teachers 20
3. Data on Schools 21
4. Data on TFA and Teaching Fellows Programs 21

F. Overview of the Analytic Approach 21

1. Experimental Analyses 21
2. Nonexperimental Analyses 23

III SELECTION, TRAINING, AND SUPPORT OF TFA PROGRAM PARTICIPANTS 27

A. Application and Selection 27

1. Application Form and Eligibility Rules 27
2. Selection Process 28
3. Program Selectivity and Application Yields 31

B. Preservice Training for New Corps Members 31

1. Independent Study and Regional Induction 32
2. Summer Institute 33
3. Post-Institute Activities 34

C. Placement and Retention of TFA Corps Members 34

D. Instruction Provided to TFA Corps Members Through Local Teacher Certification Programs 35

1. Preservice Instruction 36
2. Instruction During and After the First Year of Teaching 36
3. Total Amount of Instruction in Certification Programs 37

E. Training and Support for New Corps Members After They Begin Teaching 37

1. Instruction, Training, and Professional Development 37
2. Other Support Services 37

F. Financial Support for Corps Members 38

IV TEACH FOR AMERICA STUDY TEACHERS: THEIR CHARACTERISTICS, THE SCHOOLS THEY TEACH IN, AND THE STUDENTS THEY TEACH 39

A. Schools in the TFA Study Sample 39

B. TFA Teachers in the Study Sample Compared with Comparison Teachers 41

1. Demographic Characteristics 42
2. Educational Background and Math Content Knowledge 42
3. Teaching and Nonteaching Work Experience 45
4. Teacher Training, Professional Development, and Support 47
5. Views Toward Teaching 50

C. Students in the TFA Study Sample 51

V TEACH FOR AMERICA: IMPACTS ON SECONDARY STUDENT MATH ACHIEVEMENT 55

A. Impacts of TFA Teachers Relative to All Comparison Teachers 55

1. Average Impact 55
2. Variation in Impacts 57

B. Impacts of TFA Teachers Relative to Comparison Teachers from Traditional and Alternative Routes to Certification 58

C. Impacts Within Other Teacher Subgroups 58

1. Subgroups Defined by Teaching Experience 58
2. Subgroups Defined by Middle and High School 59

D. Conclusions 61

VI SELECTION, TRAINING, AND SUPPORT OF TEACHING FELLOWS PROGRAM PARTICIPANTS 63

A. Application and Selection 63

1. Application Form and Eligibility Rules 63
2. Selection Process 64
3. Program Selectivity and Application Yields 66

B. Preservice Training for New Teaching Fellows 67

1. Independent Study and Orientation 67
2. Summer Institute 67
3. Post-Institute Activities 70

C. Placement and Retention of Teaching Fellows 70

D. Instruction Provided to Teaching Fellows Through Local Teacher Certification Programs 71

1. Preservice Instruction 72
2. Instruction During and After the First Year of Teaching 72
3. Total Amount of Instruction in Certification Programs 72

E. Training and Support for New Teaching Fellows After They Begin Teaching 73

1. Instruction, Training, and Professional Development 73
2. Other Support Services 73

F. Financial Support for Teaching Fellows 74

VII TEACHING FELLOWS STUDY TEACHERS: THEIR CHARACTERISTICS, THE SCHOOLS THEY TEACH IN, AND THE STUDENTS THEY TEACH 75

A. Schools in the Teaching Fellows Study Sample 75

B. Teaching Fellows in the Study Sample Compared with Comparison Teachers 77

1. Demographic Characteristics 78
2. Educational Background and Content Knowledge 78
3. Teaching and Nonteaching Work Experience 83
4. Teacher Training, Professional Development, and Support 84
5. Views Toward Teaching 86

C. Students in the Teaching Fellows Study Sample 87

VIII TEACHING FELLOWS: IMPACTS ON SECONDARY STUDENT MATH ACHIEVEMENT 91

A. Impacts of Teaching Fellows Relative to All Comparison Teachers 91

1. Average Impact 91
2. Variation in Impacts 93

B. Impacts of Teaching Fellows Relative to Comparison Teachers from Traditional and Alternative Routes to Certification 94

C. Impacts Within Other Teacher Subgroups 95

1. Subgroups Defined by Specific Levels of Teaching Experience 95
2. Subgroup of Teaching Fellows and Comparison Teachers with Similar Levels of Teaching Experience 96
3. Subgroups Defined by Middle and High School 97

D. Conclusions 97

IX RELATIONSHIPS BETWEEN TEACHER CHARACTERISTICS AND TEACHER EFFECTIVENESS 101

A. Analysis of Characteristics That Predict Teacher Effectiveness 101

1. General Academic Ability 102
2. Exposure to and Knowledge of Math 104
3. Instructional Training and Experience 105

B. Accounting for the Impact of TFA Teachers 106

1. Predictions Based on Each Individual Characteristic 107
2. Predictions Based on All Observed Characteristics 109

REFERENCES 111

APPENDIX A: SUPPLEMENTARY TECHNICAL INFORMATION ON STUDY DESIGN AND DATA COLLECTION A.1

APPENDIX B: SUPPLEMENTARY INFORMATION ON ANALYTIC METHODS B.1

APPENDIX C: SUPPLEMENTARY INFORMATION ON TEACH FOR AMERICA AND TEACHING FELLOWS PROGRAMS C.1

APPENDIX D: TEACH FOR AMERICA AND TEACHING FELLOWS TEACHERS COMPARED WITH COMPARISON TEACHERS BY ENTRY ROUTE (ALTERNATIVE OR TRADITIONAL) D.1

APPENDIX E: SUPPLEMENTARY INFORMATION ON TEACH FOR AMERICA AND TEACHING FELLOWS TEACHERS COMPARED WITH COMPARISON TEACHERS E.1

APPENDIX F: SUPPLEMENTARY ANALYSES OF THE IMPACTS OF TEACH FOR AMERICA AND TEACHING FELLOWS TEACHERS F.1

APPENDIX G: SUPPLEMENTARY FINDINGS ON FACTORS ASSOCIATED WITH TEACHER EFFECTIVENESS G.1

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

TABLES

II.1 Number of Districts, Schools, Classroom Matches, Teachers, and Classes in the Study 13

II.2 Distribution of Sample Sizes Across Districts in the Study 14

II.3 Number of Classroom Matches and Study Classes, by Subject 15

II.4 Number of Students in the Study 17

II.5 Changes in Composition of Study Classes During the School Year 18

II.6 Data Sources for the Evaluation 19

II.7 Teacher Characteristics Examined in the Nonexperimental Analysis 24

III.1 Estimated Percentage of New TFA Corps Members Who Received Teaching Placements, on Average, Across the 2007–2008, 2008–2009, and 2009–2010 School Years 35

III.2 Hours of Instruction Toward Certification Reportedly Required in Focal AC Programs for TFA Secondary Math Teachers 36

IV.1 Characteristics of Study Schools with TFA Teachers Compared with All Secondary Schools with TFA Teachers and All Secondary Schools Nationwide 40

IV.2 Demographic Characteristics of TFA and Comparison Teachers in the Study and All Secondary Teachers Nationwide 43

IV.3 Educational Background of TFA and Comparison Teachers in the Study 44

IV.4 Average Number of College-Level Math Courses Taken by TFA and Comparison Teachers in the Study 45

IV.5 Praxis II Scores of TFA and Comparison Teachers in the Study 46

IV.6 Nonteaching and Teaching Work Experience of TFA and Comparison Teachers in the Study 47

IV.7 Training and Professional Development of TFA and Comparison Teachers in the Study 48

IV.8 Coursework Taken During the School Year by TFA and Comparison Teachers in the Study 49

IV.9 Mentoring Received During the School Year by TFA and Comparison Teachers in the Study 50

IV.10 Job Satisfaction of TFA and Comparison Teachers in the Study 51

IV.11 Average Baseline Characteristics of Students in the Analysis Sample Who Were Assigned to TFA Teachers or Comparison Teachers 52

V.1 Difference in Effectiveness Between TFA Teachers and All Comparison Teachers 56

VI.1 Domains and Corresponding Competencies of TNTP's Teaching for Student Achievement Framework 69

VI.2 Characteristics of Math Immersion Programs Offered by Three Teaching Fellows Programs in the Study 70

VI.3 Estimated Percentage of New Teaching Fellows Who Received Teaching Placements, on Average, Across the 2007–2008, 2008–2009, and 2009–2010 School Years 71

VI.4 Hours of Instruction Toward Certification Reportedly Required in Focal AC Programs for Secondary Math Teaching Fellows 73

VII.1 Characteristics of Study Schools with Teaching Fellows Compared with All Secondary Schools with Teaching Fellows and All Secondary Schools Nationwide 76

VII.2 Demographic Characteristics of Teaching Fellows and Comparison Teachers in the Study and All Secondary Teachers Nationwide 79

VII.3 Educational Background of Teaching Fellows and Comparison Teachers in the Study 80

VII.4 Average Number of College-Level Math Courses Taken by Teaching Fellows and Comparison Teachers in the Study 81

VII.5 Praxis II Scores of Teaching Fellows and Comparison Teachers in the Study 82

VII.6 Nonteaching and Teaching Work Experience of Teaching Fellows and Comparison Teachers in the Study 83

VII.7 Training and Professional Development of Teaching Fellows and Comparison Teachers in the Study 85

VII.8 Coursework Taken During the School Year by Teaching Fellows and Comparison Teachers in the Study 86

VII.9 Mentoring Received During the School Year by Teaching Fellows and Comparison Teachers in the Study 87

VII.10 Job Satisfaction of Teaching Fellows and Comparison Teachers in the Study 88

VII.11 Average Baseline Characteristics of Students in the Analysis Sample Who Were Assigned to Teaching Fellows or Comparison Teachers 89

VIII.1 Difference in Effectiveness Between Teaching Fellows and All Comparison Teachers....92

VIII.2 Difference in Effectiveness Between Teaching Fellows and Comparison Teachers with Similar Levels of Teaching Experience....97

IX.1 Differences in Student Math Achievement Associated with Differences in Teacher Characteristics....103

IX.2 Predicted Differences in Effectiveness Between TFA Teachers and Comparison Teachers Based on Observed Teacher Characteristics....107

A.1 Percentage of Classroom Matches in Which Students Receive Supplemental Math....A.4

A.2 Number of Districts, Schools, Classroom Matches, Teachers, and Classes in the Study, by School Year....A.7

A.3 Structure of Classroom Matches in the TFA Study Sample....A.7

A.4 Structure of Classroom Matches in the Teaching Fellows Study Sample....A.8

A.5 Average Baseline Characteristics of Students Assigned to TFA Teachers or Comparison Teachers....A.13

A.6 Average Baseline Characteristics of Students Assigned to Teaching Fellows or Comparison Teachers....A.14

A.7 Movement of Members of Research Sample During the School Year....A.15

A.8 Characteristics of Nonstudy Students on End-of-Year Rosters of Classrooms in the TFA Study Sample....A.17

A.9 Characteristics of Nonstudy Students on End-of-Year Rosters of Classrooms in the Teaching Fellows Study Sample....A.18

A.10 Realized Values for Key Precision Parameters....A.19

A.11 Reliability of the NWEA Assessments, by Subject....A.20

A.12 Invalid NWEA Scores, by Treatment Status....A.21

A.13 NWEA Test Statistics in Analysis Sample....A.22

A.14 Characteristics of State Math Assessments from which End-of-Year Scores Were Used in the Study....A.24

A.15 State Math Assessment Scores of Middle School Students in the Analysis Sample Relative to Proficiency Cut Points....A.25

A.16 Student Response Rates, by Type of Impact Estimate to Which the Student's Classroom Contributes....A.26

A.17 Characteristics of Students With and Without Valid Outcome Data....A.27

A.18 Teacher Response Rates and Sample Sizes A.28

A.19 Characteristics of Teachers, by Response to Teacher Survey A.29

A.20 Characteristics of Teachers, by Availability of Praxis II Scores A.29

B.1 Means and Standard Deviations of Covariates Included in Models for Estimating the Impacts of TFA Teachers B.8

B.2 Means and Standard Deviations of Covariates Included in Models for Estimating the Impacts of Teaching Fellows B.9

C.1 Similarities and Differences in TFA and the Teaching Fellows Programs C.6

D.1 Demographic Characteristics of Teachers in the TFA Study Sample, by Entry Route D.5

D.2 Educational Background of Teachers in the TFA Study Sample, by Entry Route D.6

D.3 Average Number of College-Level Math Courses Taken by Teachers in the TFA Study Sample, by Entry Route D.7

D.4 Praxis II Scores of Teachers in the TFA Study Sample, by Entry Route D.8

D.5 Nonteaching and Teaching Work Experience of Teachers in the TFA Study Sample, by Entry Route D.9

D.6 Training and Professional Development of Teachers in the TFA Study Sample, by Entry Route D.10

D.7 Coursework Taken During the School Year by Teachers in the TFA Study Sample, by Entry Route D.12

D.8 Mentoring Received During the School Year by Teachers in the TFA Study Sample, by Entry Route D.13

D.9 Job Satisfaction of Teachers in the TFA Study Sample, by Entry Route D.14

D.10 Demographic Characteristics of Teachers in the Teaching Fellows Study Sample, by Entry Route D.15

D.11 Educational Background of Teachers in the Teaching Fellows Study Sample, by Entry Route D.16

D.12 Average Number of College-Level Math Courses Taken by Teachers in the Teaching Fellows Study Sample, by Entry Route D.17

D.13 Praxis II Scores of Teachers in the Teaching Fellows Study Sample, by Entry Route D.18

D.14 Nonteaching and Teaching Work Experience of Teachers in the Teaching Fellows Study Sample, by Entry Route D.19

D.15 Training and Professional Development of Teachers in the Teaching Fellows Study Sample, by Entry Route D.20

D.16 Coursework Taken During the School Year by Teachers in the Teaching Fellows Study Sample, by Entry Route D.22

D.17 Job Satisfaction of Teachers in the Teaching Fellows Study Sample, by Entry Route D.24

E.1 Sensitivity of Differences in Praxis II Scores Between TFA and Comparison Teachers in the Study E.7

E.2 Sensitivity of Differences in Praxis II Scores Between Teaching Fellows and Comparison Teachers in the Study E.11

F.1 Differences in Effectiveness Between TFA Teachers and Comparison Teachers, Sensitivity Analyses F.7

F.2 Impact on Student Absences of TFA Teachers Relative to Comparison Teachers F.9

F.3 Differences in Effectiveness Between TFA Teachers and Comparison Teachers, by Subgroup F.10

F.4 Differences in Effectiveness Between Teaching Fellows and Comparison Teachers, Sensitivity Analyses F.11

F.5 Impact on Student Absences of Teaching Fellows Relative to Comparison Teachers F.13

F.6 Differences in Effectiveness Between Teaching Fellows and Comparison Teachers, by Subgroup F.14

G.1 Summary Statistics of Teacher Characteristics Examined in the Nonexperimental Analysis G.6

G.2 Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Only a Single Teacher Characteristic G.7

G.3 Estimated Associations Between Teacher Characteristics and Student Math Achievement in a Model that Includes Multiple Teacher Characteristics: Main Specification G.9

G.4 Extent to Which Teacher Characteristics Explain the Difference in Effectiveness Between TFA Teachers and Comparison Teachers G.10

G.5 Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Multiple Teacher Characteristics: Specifications with Alternative Measures of College Selectivity G.12

G.6 Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Multiple Teacher Characteristics: Specifications with Measures of Academic Degrees in Math or Math-Related Subjects.....G.13

G.7 Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Multiple Teacher Characteristics: Specifications with Alternative Measures of Praxis Achievement G.15

G.8 Estimated Associations Between Teacher Characteristics and Student Math Achievement in a Model that Includes Multiple Teacher Characteristics: Specification with Teacher Demographic Characteristics G.16

FIGURES

I.1 Conceptual Framework 6

III.1 TFA Selection Process and Applicant Flow 31

III.2 Overview of Placement and Training of TFA Corps Members 32

V.1 Differences in Effectiveness Between TFA Teachers and Comparison Teachers from Individual Classroom Matches 58

V.2 Differences in Effectiveness Between TFA Teachers and Comparison Teachers from Traditional and Alternative Routes to Certification 59

V.3 Differences in Effectiveness Between Novice TFA Teachers and Comparison Teachers, by Experience Level of Comparison Teacher 60

V.4 Differences in Effectiveness Between TFA Teachers and Comparison Teachers Within Middle Schools and High Schools 61

VI.1 Teaching Fellows Selection Process and Applicant Flow 66

VI.2 Overview of Placement and Training of Teaching Fellows 68

VIII.1 Differences in Effectiveness Between Teaching Fellows and Comparison Teachers from Individual Classroom Matches 93

VIII.2 Differences in Effectiveness Between Teaching Fellows and Comparison Teachers from Traditional and Alternative Routes to Certification 94

VIII.3 Differences in Effectiveness Between Novice Teaching Fellows and Comparison Teachers, by Experience Level of Comparison Teacher 95

VIII.4 Differences in Effectiveness Between Experienced Teaching Fellows and Comparison Teachers, by Experience Level of Comparison Teacher 96

VIII.5 Differences in Effectiveness Between Teaching Fellows and Comparison Teachers Within Middle Schools and High Schools 98

A.1 District and School Recruiting A.6

A.2 Number of Students Involved in Each Stage of Random Assignment and Data Collection A.12

C.1 Teach For America Regions, as of August 2011 ..... C.3

C.2 Teach For America's Teaching as Leadership Rubric C.4

C.3 TNTP's Teaching Fellows Programs, as of August 2011 ..... C.5

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

EXECUTIVE SUMMARY

Teach For America (TFA) and the Teaching Fellows programs are an important and growing source of teachers of hard-to-staff subjects in high-poverty schools, but comprehensive evidence of their effectiveness has been limited. This report presents findings from the first large-scale random assignment study of secondary math teachers from these programs. The study separately examined the effectiveness of TFA and Teaching Fellows teachers, comparing secondary math teachers from each program with other secondary math teachers teaching the same math courses in the same schools. The study focused on secondary math because this is a subject in which schools face particular staffing difficulties.

The study had two main findings, one for each program studied:

1. **TFA teachers were more effective than the teachers with whom they were compared.** On average, students assigned to TFA teachers scored 0.07 standard deviations higher on end-of-year math assessments than students assigned to comparison teachers, a statistically significant difference. This impact is equivalent to an additional 2.6 months of school for the average student nationwide.

2. **Teaching Fellows were neither more nor less effective than the teachers with whom they were compared.** On average, students of Teaching Fellows and students of comparison teachers had similar scores on end-of-year math assessments.

By providing rigorous evidence on the effectiveness of secondary math teachers from TFA and the Teaching Fellows programs, the study can shed light on potential approaches for improving teacher effectiveness in hard-to-staff schools and subjects. The study findings can provide guidance to school principals faced with the choice of hiring teachers who have entered the profession via different routes to certification. The findings can also aid policymakers and funders of teacher preparation programs by providing information on the effectiveness of teachers from various routes to certification that use different methods to identify, attract, train, and support their teachers.

Background

High-quality, effective teachers are critical to students' success in the classroom (Aaronson et al. 2003; Chetty et al. 2011), yet little is known about how best to identify, attract, train, and support such teachers. The need for effective teachers is especially acute in schools serving low-income students who already face numerous disadvantages (Monk 2007; Jacob 2007). These schools face particular difficulty attracting qualified teachers to teach secondary math and science classes (Ingersoll and Perda 2009; Ingersoll and May 2012).

TFA and The New Teacher Project (TNTP) Teaching Fellows programs take a distinctive approach to addressing the need for high-quality teachers of hard-to-staff subjects in high-poverty schools.[1] Like other programs that offer alternative routes to teacher certification, both TFA and the Teaching Fellows programs aim to lower the barriers to entering the teaching profession; both programs recruit individuals without prior teaching experience and enable them to begin teaching before completing all of the training requirements for certification. However, unlike most programs providing alternative routes to certification that do not have restrictive selection criteria and admit most applicants (Walsh and Jacobs 2007; Mayer et al. 2003), TFA and the Teaching Fellows programs have highly selective admissions criteria designed to admit only applicants who have demonstrated a high level of achievement in academics or other endeavors and who possess characteristics that the programs view as being associated with effective teaching.

Research Questions and Study Design

The study addresses two analogous primary research questions, one for each program studied:

1. How effective are **TFA teachers** at teaching secondary math compared with other teachers teaching the same math courses in the same schools?
2. How effective are **Teaching Fellows** at teaching secondary math compared with other teachers teaching the same math courses in the same schools?

The study used an experimental design to learn about the effectiveness of teachers from each of the two programs studied. At the beginning of the school year, the study team assigned students who enrolled in an eligible middle or high school math course to either a class taught by a math teacher from the program being studied (TFA or Teaching Fellows) or to a similar math class in the same school taught by a teacher who did not participate in either of the programs studied. This latter teacher is referred to as the comparison teacher. The comparison teacher could have entered teaching after completing all certification requirements (the traditional route to certification) or through a less selective alternative route to certification. Students were assigned to teachers randomly—that is, completely by chance—to ensure that there would be no systematic differences between the types of students assigned to the different types of teachers. To estimate the effectiveness of the TFA or Teaching Fellows teachers relative to the comparison teachers, we compared the end-of-year math test scores of students assigned to the different types of teachers. The study design is summarized in the box below.

[1] TNTP, formerly known as The New Teacher Project, is a national nonprofit organization that partners with urban school districts to establish and operate local Teaching Fellows programs that recruit and train new teachers.

Study Design

Experimental design. We used the same experimental design to separately examine the effectiveness of teachers from each of two programs—TFA and the Teaching Fellows programs. In each participating school, we identified "classroom matches"—two or more classes covering the same middle or high school math course at the same level, with at least one class taught by a teacher from the program being studied (TFA or Teaching Fellows) and at least one class taught by another teacher, referred to as a comparison teacher, who did not enter teaching through a highly selective alternative route. The comparison teacher could have entered teaching through either a traditional or less selective alternative route to certification. In each classroom match, students were randomly assigned at the beginning of the school year to a class taught by a teacher from the program being studied or a class taught by a comparison teacher.

Sample. In order to obtain a large enough sample of teachers and students, we recruited participants into the study in both the 2009–2010 and 2010–2011 school years. The final TFA study sample consisted of 4,573 students, 111 classroom matches, 136 math teachers, 45 schools, and 11 districts in 8 states. The final Teaching Fellows study sample consisted of 4,116 students, 118 classroom matches, 153 math teachers, 44 schools, and 9 districts in 8 states. There was limited overlap between the two samples—only one-third of study districts, 9 percent of study schools, and less than three classroom matches provided study teachers from both the TFA and Teaching Fellows programs. Most (75 percent) of the classroom matches in the TFA study sample were in middle schools, while most (69 percent) of the classroom matches in the Teaching Fellows study sample were in high schools. Forty-one percent of the comparison teachers in the TFA sample entered teaching through less selective alternative routes to certification, compared with 27 percent of the comparison teachers in the Teaching Fellows sample.

Data on teacher characteristics. To collect information on the characteristics of teachers in the study, we administered a survey to all teachers in the study in the spring of each of the study school years. We also collected teachers' scores from either the Praxis II Middle School Mathematics assessment (for middle school math teachers) or the Praxis II Mathematics Content Knowledge assessment (for high school math teachers), administering the test to teachers who had not taken it previously and gathering existing scores from those who had.

Data on outcomes. The math achievement of students assigned to teachers in the study was the key outcome used to evaluate the teachers' effectiveness. Student math achievement was measured by math assessments administered at the end of the school year in which the students were randomly assigned. For students in grades 6 to 8, we obtained scores on state-required assessments. For students in grades 9 to 12, we administered end-of-course math assessments developed by the Northwest Evaluation Association (NWEA). The state assessments were expected to align more closely with state curriculum standards, whereas the NWEA assessments were expected to be less prone to floor or ceiling effects for high- and low-achieving students.

Analysis. To estimate the effectiveness of TFA or Teaching Fellows teachers relative to the comparison teachers, we compared end-of-year test scores of the students assigned to the TFA or Teaching Fellows teachers and the students assigned to the comparison teachers. We estimated the effectiveness of TFA and Teaching Fellows teachers separately.

This experimental design ensured that any differences in math achievement between the students of TFA or Teaching Fellows teachers and the students of comparison teachers could be attributed to differences in the teachers' effectiveness rather than differences in the characteristics of the students they taught, the subject they taught, or the school in which they taught.

Although students in the study were randomly assigned among teachers from different routes to certification, teachers were not randomly assigned to these routes. TFA and the Teaching Fellows programs may attract different types of candidates than other routes to certification—these differences can arise both from the programs' approaches to recruitment and selection and from the teachers' decisions on which programs to apply to and attend. Therefore, differences in effectiveness between TFA teachers and comparison teachers, and between Teaching Fellows and comparison teachers, do not reflect the effectiveness of the programs alone. Instead, they reflect the influence of *both* differences in the types of individuals who choose to enter teaching through TFA or a Teaching Fellows program versus some other training program and differences in the recruitment and selection procedures and training and support the programs offer. The study cannot rigorously disentangle these components.

No restriction was placed on the experience levels of the TFA teachers, Teaching Fellows, or comparison teachers. Hence, in some comparisons, novice TFA (or Teaching Fellows) teachers were compared with more experienced comparison teachers. However, the study also explored the effectiveness of the TFA (or Teaching Fellows) teachers when they were compared with teachers with similar levels of experience.

The study was not designed to compare the effectiveness of TFA teachers with that of Teaching Fellows. There are some substantive differences between the programs; for instance, TFA requires its teachers to commit to two years in teaching, whereas Teaching Fellows programs expect their teachers to teach for many years. In addition, TFA focuses more on recruiting new college graduates, whereas the Teaching Fellows programs focus more on recruiting professionals who want to switch careers. However, there are also many reasons not directly related to the effectiveness of the programs that might lead estimates of the effectiveness of their teachers to differ. Students were not randomly assigned between TFA teachers and Teaching Fellows, and thus any differences in the estimated effectiveness of TFA teachers and Teaching Fellows could be due to differences in the students they teach, the teachers with whom they were compared, or the schools or districts in which they teach.

In fact, TFA and Teaching Fellows teachers in the study largely taught in different schools and districts—only one-third of study districts, 9 percent of study schools, and less than three classroom matches provided study teachers from both the TFA and Teaching Fellows programs. Moreover, TFA teachers in the study were more likely to teach in middle schools, whereas Teaching Fellows were more likely to teach in high schools, consistent with the fact that, nationwide, secondary math teachers from TFA are more likely to be placed in middle schools and those from Teaching Fellows programs are more likely to be placed in high schools. Because the challenges of teaching may vary across grade levels and because the study used different student assessments at the middle and high school levels, it could be misleading to directly compare the effectiveness of the TFA and Teaching Fellows teachers in the study. Finally, TFA and Teaching Fellows teachers were also compared with different types of teachers; 41 percent of the teachers with whom TFA teachers were compared entered teaching through less selective alternative routes to certification, while 27 percent of the teachers with whom Teaching Fellows

were compared entered teaching through less selective alternative routes. For all of these reasons, the study findings cannot be used to compare the effectiveness of TFA and Teaching Fellows teachers.

Teach For America

Selection, Training, and Support of TFA Program Participants

TFA is a national program that places teachers in numerous regions across the country—43 as of August 2011. TFA recruits, selects, trains, and supports its teachers. Each of these elements of the program could influence the effectiveness of its secondary math teachers, so information on TFA's operations can provide important context for interpreting the study's findings.

TFA focuses its recruitment efforts on new college graduates, targeting seniors at many of the nation's most distinguished colleges and universities. Individuals apply to TFA as a whole, not to a regional program; they may specify regional preferences, but the program will assign them to whatever region is deemed the best fit. Furthermore, TFA requires its teachers to commit to two years in teaching. The selection process includes a written application, a telephone interview, an analysis and writing exercise, an in-person interview, a monitored group discussion, and a five-minute sample teaching lesson. Using a computer-based algorithm to help identify applicants with the greatest potential to be effective teachers, TFA accepts about 12 percent of applicants nationwide.

TFA teachers receive training and support both before and after they start teaching. The main training experience before participants become classroom teachers is an intensive five-week summer institute in which they receive formal instruction and engage in practice teaching. Most TFA regional programs do not provide all of the instruction necessary for certification, so participants typically must also enroll in a separate state-authorized alternative certification program in order to complete coursework toward their certification during their first years of teaching. Staff from participating TFA programs reported that the alternative certification programs where most of their secondary math teachers had enrolled in the three preceding years involved an average of 137 hours of coursework. In addition, TFA provides information, feedback, and mentoring to its teachers in their first one to two years of teaching.

TFA works with districts to place teachers in high-need schools. Compared with the average U.S. secondary school, secondary schools with new TFA teachers are more likely to be located in urban areas and to be eligible for Title I funds for schools with high percentages of low-income students. In addition, they serve a higher percentage of students from racial and ethnic minorities and students eligible for free or reduced-price lunch.

Study Schools with TFA Teachers

Because the study's findings about TFA teachers pertain to the types of schools represented in the TFA study sample, it is important to understand the characteristics of these schools.

Schools with TFA teachers in the study served disadvantaged populations. Eighty-two percent of schools with TFA teachers in the study were eligible for Title I funds. Eighty-nine percent of students in schools with TFA study teachers were Hispanic or black. Students in the

TFA study sample scored below average for their states in both reading and math before participating in the study.

Schools with TFA teachers in the study were similar to schools employing secondary math teachers from TFA nationwide. Over 80 percent of students at both sets of schools were from racial and ethnic minority groups, and close to 80 percent were eligible for free or reduced-price lunch. However, there were some differences—for example, urban schools were overrepresented and charter schools were underrepresented in the TFA study sample.

Teachers in the TFA Sample

Interpreting the study estimates of teacher effectiveness requires understanding the characteristics of both TFA teachers and the teachers with whom they were compared. TFA has unique procedures for recruiting and selecting individuals with particular characteristics that it believes are associated with effectiveness in the classroom. Hence, it is not surprising that we found many differences in the characteristics and experiences of the TFA teachers in the study and the teachers with whom they were compared.

TFA teachers had less teaching experience than comparison teachers. Consistent with TFA requiring a two-year commitment to teaching, TFA teachers in our study had significantly less teaching experience than comparison teachers. Eighty-three percent of the TFA teachers were in either their first or second year of teaching during the study school year, compared with 10 percent of the comparison teachers. TFA teachers in the study also had less nonteaching work experience than the comparison teachers.

TFA teachers were less likely than comparison teachers to be from racial or ethnic minorities and less likely to be female. Eighty-nine percent of TFA teachers in the study were white and non-Hispanic, compared with 30 percent of the comparison teachers. Sixty-one percent of TFA teachers in the study were female, compared with 79 percent of the comparison teachers. In terms of both race and gender distribution, TFA teachers were more similar to secondary teachers nationwide than the comparison teachers—84 percent of secondary teachers nationwide were white and non-Hispanic, and 59 percent were female.

TFA teachers had graduated from more selective colleges than comparison teachers. Eighty-one percent of TFA teachers in the study had graduated from a selective college or university, compared with 23 percent of the comparison teachers.[2]

TFA teachers were less likely than comparison teachers to have majored in math, but scored higher on a test of math content knowledge. Relative to the comparison teachers, TFA teachers in the study were less likely to have a major or minor in math or secondary math education; they had also taken significantly fewer college-level math courses. Despite these differences, on average, TFA teachers scored significantly higher on the Praxis II math exam than the comparison teachers. TFA teachers who took the Mathematics Content Knowledge Test scored 22 points (or 0.93 standard deviations) higher than their counterparts from other routes;

[2] College selectivity was defined based on *Barron's Profiles of American Colleges 2003*. Selective colleges were defined as those ranked by Barron's as very competitive, highly competitive, or most competitive.

TFA teachers who took the Middle School Mathematics Test outscored their counterparts by 22 points (or 1.19 standard deviations).

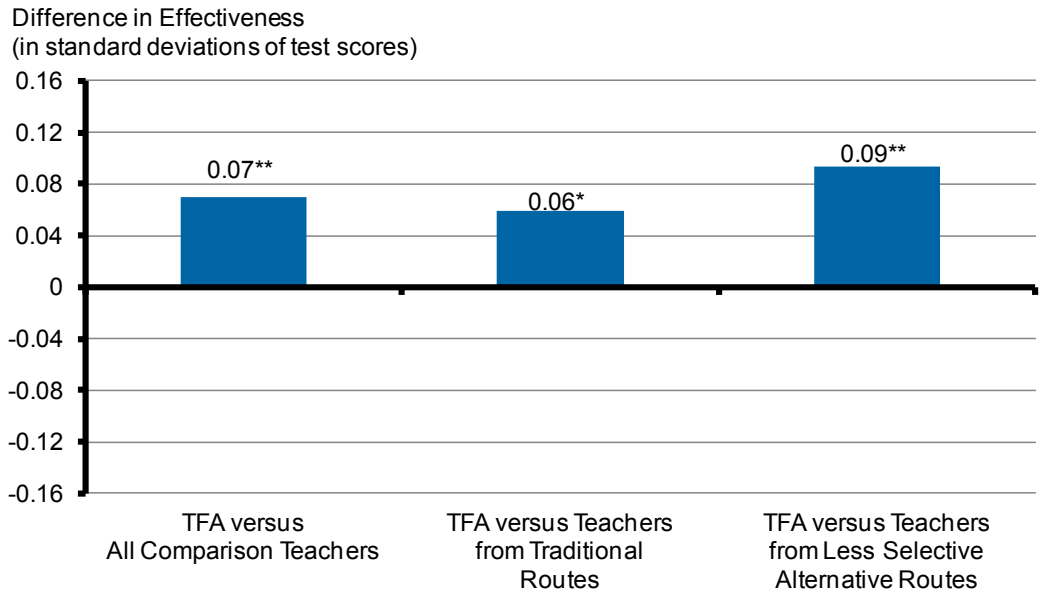
TFA Teachers Were More Effective Than Comparison Teachers

On average, students assigned to TFA teachers scored higher (by 0.07 standard deviations) on end-of-year math assessments than students assigned to comparison teachers (Figure ES.1). This difference in math scores was equivalent to an increase in student achievement from the 27th to the 30th percentile. This difference also translated into an additional 2.6 months of school for the average student nationwide.

In addition, TFA teachers were more effective than the teachers with whom they were compared across several key teacher subgroups we examined.

- **TFA teachers were more effective than their counterparts from both traditional and less selective alternative routes to certification.** One criticism of alternative routes to certification, both less and highly selective routes, is that, because they provide less preparation for teaching, their teachers are less effective than teachers from traditional routes to certification (Darling-Hammond 1990, 2000; Darling-Hammond et al. 2005). To explore the validity of this criticism as it applies to TFA teachers, we conducted a subgroup analysis estimating the effectiveness of TFA teachers relative to their counterparts from traditional routes. We found that students of TFA teachers outperformed those of teachers from traditional routes by 0.06 standard deviations. We also examined the effectiveness of TFA teachers relative to their counterparts from less selective alternative routes—the hard-to-staff schools in which TFA teachers are placed may often be faced with the choice of hiring a TFA teacher or a teacher from these other alternative routes. We found that students of TFA teachers scored higher than students of teachers from less selective alternative routes by 0.09 standard deviations.
- **Novice TFA teachers were more effective than both novice and experienced comparison teachers.** Another common criticism of TFA is that it seeks teachers willing to make a two-year rather than a longer-term commitment to teaching. Critics claim that too many TFA teachers leave teaching before they accumulate the experience needed to be as effective as their counterparts from other routes (Heilig and Jez 2010). To examine this concern, we compared novice TFA teachers (defined as those in their first three years of teaching) with both novice and experienced comparison teachers. We found that novice TFA teachers were more effective than comparison teachers regardless of the comparison teachers' experience. Students of TFA teachers in their first three years of teaching scored higher on math assessments (by 0.08 standard deviations) than students of their counterparts in their first three years of teaching. Students of TFA teachers in their first three years of teaching also had higher student math achievement (by 0.07 standard deviations) compared with students of comparison teachers with more than three years of experience.

Figure ES.1. Differences in Effectiveness Between Teach For America Teachers and Comparison Teachers, Overall and by Entry Route into Teaching



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

*Estimate is statistically significant at the 0.05 level based on a two-tailed test after adjusting for multiple hypothesis testing.
**Estimate is statistically significant at the 0.01 level based on a two-tailed test after adjusting for multiple hypothesis testing.

TFA = Teach For America.

- **TFA teachers were more effective than comparison teachers in both middle and high schools.** Because the effective teaching of math might require different knowledge or skills depending on the level of math being taught, we examined TFA teachers' effectiveness separately in middle and high schools. We found that TFA teachers were more effective than comparison teachers at both the middle and high school levels. At the middle school level, students of TFA teachers outperformed students of comparison teachers by 0.06 standard deviations. At the high school level, students of TFA teachers outperformed students of comparison teachers by 0.13 standard deviations.

Teaching Fellows

Selection, Training, and Support of Teaching Fellows Program Participants

TNTP works with districts—19 as of August 2011—to establish locally run Teaching Fellows programs. The local programs have distinct identities, but TNTP guidance ensures basic consistency in their approach to recruiting, selecting, training, and supporting their teachers.

Teaching Fellows programs accept new college graduates but focus their recruitment efforts more on professionals who want to switch careers. Individuals apply to a specific Teaching Fellows program in the locality in which they want to teach. Teaching Fellows programs expect their participants to make an open-ended commitment to teaching rather than a short-term

commitment. The selection process includes a written application, an analysis and writing exercise, an in-person interview, a monitored group discussion, and a five-minute sample teaching lesson. Trained staff apply specific selection criteria and use their judgment to identify applicants with the greatest potential to be effective teachers. Teaching Fellows programs nationwide accept about 13 percent of applicants.

Teaching Fellows teachers receive training and support both before and after they start teaching. The main training experience before participants become classroom teachers is an intensive five- to seven-week summer institute in which they receive formal instruction and engage in practice teaching. Most Teaching Fellows programs do not provide all of the instruction necessary for certification, so participants typically must also enroll in a separate state-authorized alternative certification program in order to complete coursework toward their certification during their first years of teaching. Staff from participating Teaching Fellows programs reported that the alternative certification programs where most of their secondary math teachers had enrolled in the three preceding years involved an average of 137 hours of coursework. In addition, Teaching Fellows programs provide information, feedback, and mentoring to their teachers in their first one to two years of teaching.

Teaching Fellows programs work with districts to place teachers in high-need schools. Compared with the average U.S. secondary school, secondary schools with new Teaching Fellows teachers are more likely to be located in urban areas and to be eligible for Title I funds. In addition, they serve a higher percentage of students from racial and ethnic minorities and students eligible for free or reduced-price lunch.

Study Schools with Teaching Fellows

Because the study's findings about Teaching Fellows pertain to the types of schools represented in the Teaching Fellows study sample, it is important to understand the characteristics of these schools.

Schools with Teaching Fellows in the study served disadvantaged populations. Eighty-two percent of all study schools with Teaching Fellows were eligible for Title I funds. More than 80 percent of students in study schools with Teaching Fellows were Hispanic or black. Students in the Teaching Fellows study sample scored below average for their states in both reading and math before participating in the study.

Schools with Teaching Fellows in the study were similar to schools employing secondary math teachers from Teaching Fellows programs nationwide. Over 80 percent of students at both sets of schools were from racial or ethnic minority groups, and close to 80 percent were eligible for free or reduced-price lunch. However, as in the TFA study sample, urban schools were overrepresented and charter schools were underrepresented in the Teaching Fellows study sample.

Teachers in the Teaching Fellows Sample

Teaching Fellows programs have unique procedures for recruiting and selecting individuals with particular characteristics that they believe are associated with effectiveness in the classroom. Hence, as with TFA teachers and their counterparts, we found many differences in

the characteristics and experiences of the Teaching Fellows in the study and the teachers with whom they were compared.

Teaching Fellows had less teaching experience than comparison teachers. Consistent with the Teaching Fellows programs' expectations that their teachers remain in teaching for many years, 76 percent of Teaching Fellows had three or more years of teaching experience. However, on average, they still had less experience than comparison teachers. Twenty-three percent of Teaching Fellows were in their first or second year of teaching, compared with 7 percent of the comparison teachers. This may reflect the fact that the Teaching Fellows programs were only launched in the year 2000 or after, and hence teachers from these programs have not had time to gain decades of experience. Consistent with the focus of Teaching Fellows programs on recruiting professionals to transition into teaching, Teaching Fellows had more nonteaching work experience than did the comparison teachers.

Teaching Fellows were less likely than comparison teachers to be from racial or ethnic minorities. Seventy-one percent of Teaching Fellows in the study were white and non-Hispanic, compared with 43 percent of the comparison teachers. Teaching Fellows were more similar to secondary teachers nationwide than the comparison teachers in terms of racial and ethnic distribution—84 percent of secondary teachers nationwide were white and non-Hispanic. Both Teaching Fellows and comparison teachers were similar to the average secondary teacher nationwide in terms of gender distribution. Fifty-four percent of Teaching Fellows and 57 percent of comparison teachers were female, compared with 59 percent of secondary teachers nationwide—these differences were not statistically significant.

Teaching Fellows had graduated from more selective colleges than comparison teachers. Seventy-two percent of Teaching Fellows in the study had graduated from a selective college or university, compared with 34 percent of the comparison teachers.[3]

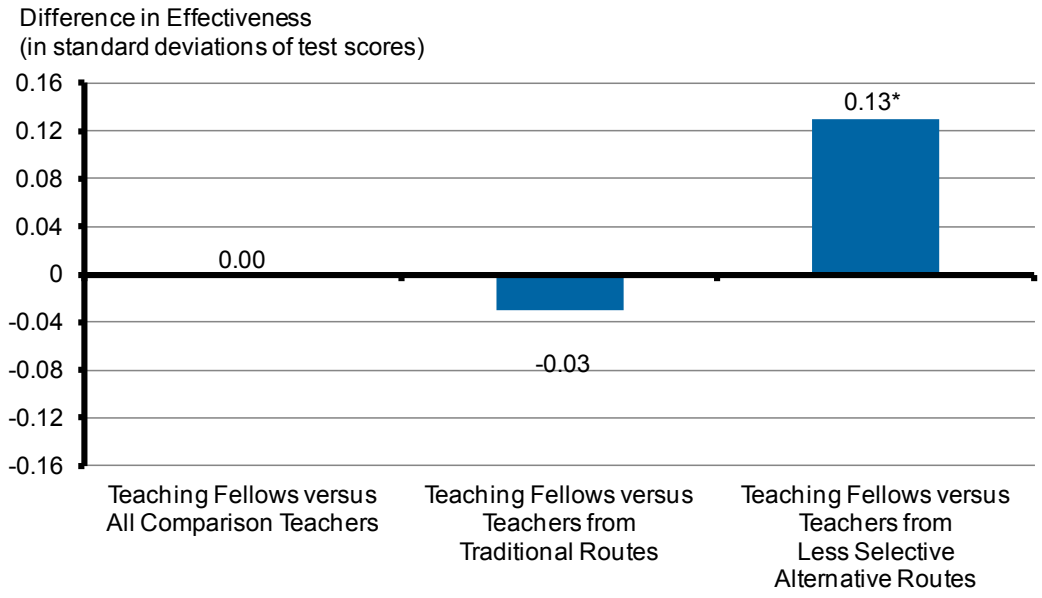
Teaching Fellows were less likely than comparison teachers to have majored in math, but scored higher on a test of math content knowledge. Relative to the comparison teachers, Teaching Fellows were less likely to have majored in math or secondary math education; nonetheless, they had taken about the same number of math courses. Despite being less likely to have majored in math, Teaching Fellows scored significantly higher on the Praxis II math exam than the comparison teachers. Teaching Fellows who took the Mathematics Content Knowledge Test scored 19 points (or 0.80 standard deviations) higher than their counterparts from other routes, whereas Teaching Fellows who took the Middle School Mathematics Test outscored their counterparts by 17 points (or 0.92 standard deviations).

Teaching Fellows Did Not Differ in Effectiveness from Comparison Teachers Overall, but the Study Found Differences for Some Subgroups

On average, students of Teaching Fellows and students of comparison teachers had similar scores on end-of-year math assessments (Figure ES.2). However, we found that effectiveness varied for specific subgroups of Teaching Fellows and comparison teachers:

[3] College selectivity was defined based on *Barron's Profiles of American Colleges 2003*. Selective colleges were defined as those ranked by Barron's as very competitive, highly competitive, or most competitive.

Figure ES.2. Differences in Effectiveness Between Teaching Fellows and Other Math Teachers in the Same Schools, Overall and by Entry Route into Teaching



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

*Estimate is statistically significant at the 0.05 level based on a two-tailed test after adjusting for multiple hypothesis testing.
**Estimate is statistically significant at the 0.01 level based on a two-tailed test after adjusting for multiple hypothesis testing.

- **Teaching Fellows were more effective than teachers from less selective alternative routes to certification, and neither more nor less effective than teachers from traditional routes to certification.** As an alternative route to teacher certification, Teaching Fellows programs are subject to the same criticisms mentioned earlier regarding TFA—specifically, that they, like other alternative routes to certification, provide insufficient preparation for teaching relative to traditional routes. However, we found that students of Teaching Fellows did not perform significantly better or worse than students of teachers from traditional routes. We also compared the effectiveness of Teaching Fellows with that of teachers from less selective alternative routes, as many schools may be faced with the choice of hiring a Teaching Fellow or a teacher from these other alternative routes. We found that students of Teaching Fellows outperformed students of teachers from less selective alternative routes by 0.13 standard deviations.
- **Novice Teaching Fellows were more effective than novice comparison teachers, but this difference dissipated among more experienced teachers.** Schools may often be faced with the choice of hiring a new Teaching Fellow or a new teacher from some other route. We found that novice Teaching Fellows (those in their first three years of teaching) were more effective than novice comparison teachers, increasing student achievement in math by 0.13 standard deviations. However, when Teaching Fellows from all levels of experience were compared with comparison teachers whose experience differed by no more than two years, there was no statistically

significant difference in math achievement between the students of these two groups of teachers.

- **Teaching Fellows were neither more nor less effective than comparison teachers in both middle and high schools.** We examined Teaching Fellows' effectiveness separately at the middle and high school levels. We found that math achievement did not differ significantly between students of Teaching Fellows and students of comparison teachers at either the middle or high school level.

Observed Characteristics and Teacher Effectiveness

Because teachers in the study varied substantially in their effectiveness, the study provided an opportunity to examine the associations between specific teacher characteristics and student math achievement. This analysis used the full sample of teachers in our study—TFA teachers, Teaching Fellows, and comparison teachers. Understanding the relationship between teacher characteristics and effectiveness can inform policies to improve teacher effectiveness and principals' hiring decisions. Because this analysis relied on nonexperimental methods, the results can generate hypotheses about the characteristics of effective math teachers, but cannot provide conclusive evidence about whether these characteristics *caused* the higher student math achievement.

Consistent with previous research (Kane et al. 2008), we found that different teachers who took the same route into teaching varied considerably in their effectiveness in the classroom. Yet, few observed characteristics appeared to explain these differences. We examined eight teacher characteristics: (1) the selectivity of the teacher's undergraduate institution, (2) amount of college-level math coursework completed, (3) use of college-level math in a prior nonteaching position, (4) performance on Praxis II math content knowledge assessments for teachers, (5) amount of pedagogy instruction received, (6) extent of student teaching completed, (7) amount of coursework taken during the school year, and (8) years of teaching experience. Of these characteristics, two were consistently associated with student achievement, and a third was associated with student achievement in some but not all analyses conducted:

1. **Teacher effectiveness increased with teacher experience.** Students assigned to a second-year teacher were predicted to score 0.08 standard deviations higher on math assessments than students assigned to a first-year teacher. Among teachers with at least five years of teaching experience, each additional year of teaching experience was associated with an increase of 0.005 standard deviations in student achievement.
2. **Teacher effectiveness declined with increasing amounts of teacher coursework during the school year.** For each additional 10 hours of coursework that teachers took during the school year, the math achievement of their students was predicted to drop by 0.002 standard deviations. These findings imply that a teacher who took an average amount of coursework during the school year, whether for initial certification or any other certification or degree, decreased student math achievement by 0.04 standard deviations relative to a teacher who took no coursework while teaching. Although we cannot directly examine why coursework was negatively related to teacher effectiveness, the findings are consistent with the hypothesis that coursework taken during the school year diverts teachers' energy and attention from the classes they are teaching.

3. **In some of the analyses we conducted, greater math content knowledge was positively associated with teacher effectiveness in high schools, but not in middle schools.** In high schools, we found that teachers' math content knowledge, as measured by their scores on the Praxis II math assessments, was positively associated with student achievement in some of our sensitivity analyses, but not in our main analysis. Our main analysis indicates that a one standard deviation increase in a high school teacher's Praxis II score—equivalent to rising from the 50th to the 84th percentile of Praxis II scores—was associated with an increase in student math achievement of 0.04 standard deviations, but this difference was not statistically significant at the 5 percent level (p-value = 0.051). Our sensitivity analyses found that high school students taught by teachers in the top half of the Praxis II score distribution had higher math achievement, by a statistically significant degree, than those taught by teachers in the bottom half of the distribution. In middle schools, we found no association between teachers' scores on the Praxis II and student achievement.

Collectively, the teacher characteristics we examined could not account for any portion of the difference in effectiveness between TFA teachers and the teachers with whom they were compared. Although TFA teachers would have been predicted to be *less* effective than comparison teachers on the basis of their characteristics—primarily because of the TFA teachers' relative inexperience in the classroom—TFA teachers were, in fact, more effective than their counterparts. The nonexperimental methods used for this analysis cannot provide conclusive evidence about the causal relationship between the characteristics examined and student math achievement, but the results suggest that TFA teachers' greater effectiveness might have been driven by teacher characteristics that we did not observe.

Conclusions

Teachers in the study who entered teaching through either TFA or the Teaching Fellows programs were at least as effective at teaching secondary math as those who entered via traditional or less selective alternative routes to certification, and in some cases they were more effective. Therefore, TFA and the Teaching Fellows programs represent two examples of program models for recruiting, selecting, training, and supporting teachers that can address teacher shortages in secondary math within high-need schools without decreasing student achievement.

The study findings can provide guidance to school principals considering hiring decisions. Although a specific teacher from TFA or a Teaching Fellows program might be more or less effective than a teacher from a traditional or less selective alternative route, our findings can shed light on the average effectiveness of the teachers from TFA relative to teachers from another route and on the average effectiveness of Teaching Fellows relative to teachers from another route.

Our study suggests that, on average, principals of the secondary schools in the study would raise student math achievement by hiring a TFA teacher rather than a teacher from a traditional or less selective alternative route to teach the math classes examined in the study. Although the TFA teachers in the study were less experienced, on average, than the comparison teachers, students of TFA teachers outperformed students of other teachers in the same grades and schools by a statistically significant margin. This result held true whether the comparison teachers were

from traditional routes or less selective alternative routes. Similarly, students of TFA teachers in their first three years of teaching outperformed students of other novice teachers in the same grades and schools as well as students of more experienced teachers. This latter finding is particularly important given the fact that TFA requires its teachers to make only a two-year commitment to teaching.

Our main impact findings for Teaching Fellows suggest that a secondary school in the study would experience neither higher nor lower student math achievement if its principal hired Teaching Fellows math teachers rather than math teachers from traditional or less selective alternative routes. Nevertheless, a principal faced with a more specific choice between a novice Teaching Fellow and a novice teacher from another route or a choice between a Teaching Fellow and a teacher from a less selective alternative route should expect higher student achievement, on average, from hiring the Teaching Fellow. If comparing a Teaching Fellow with another teacher with the same years of experience, on average, the principal would do just as well hiring either teacher.

I. INTRODUCTION

High-quality, effective teachers are critical to students' success in the classroom (Aaronson et al. 2003; Rockoff 2004; Hanushek et al. 2005; Rivkin et al. 2005; Chetty et al. 2011), yet little is known about how best to identify, attract, train, and support such teachers. The need for effective teachers is especially acute in schools serving low-income students who already face numerous disadvantages (Monk 2007; Jacob 2007). These schools face particular difficulty attracting qualified teachers to teach secondary math and science classes (Ingersoll and Perda 2009; Ingersoll and May 2012).

Two programs—Teach For America (TFA) and the TNTP Teaching Fellows programs—take a distinctive approach to addressing the need for high-quality teachers of hard-to-staff subjects in high-poverty schools.[4] Like other programs that offer alternative routes to teacher certification (AC routes), both TFA and the Teaching Fellows programs aim to lower the barriers to entering the teaching profession—both programs recruit new college graduates and professionals without prior teaching experience and enable them to begin teaching before completing all of the training requirements for certification. However, unlike most programs providing alternative routes to certification that do not have restrictive selection criteria and admit most applicants (Walsh and Jacobs 2007; Mayer et al. 2003), TFA and the Teaching Fellows programs have highly selective admissions criteria designed to admit only applicants who have demonstrated a high level of achievement in academics or other endeavors and who possess characteristics that the programs view as being associated with effective teaching. Both programs provide their teachers with intensive training the summer before they begin teaching and provide ongoing support and professional development after their teachers begin teaching.

To learn about the effectiveness of teachers from these programs, the U.S. Department of Education's (ED's) Institute of Education Sciences (IES) commissioned an evaluation of the impact on student achievement of middle and high school math teachers from the TFA and Teaching Fellows programs. The evaluation followed an experimental design: students were randomly assigned to a class taught by a teacher from either TFA or a Teaching Fellows program or to a class taught by a comparison teacher, a teacher who entered the teaching profession through a traditional or less selective alternative route to certification. Because the students were randomly assigned, any difference in the outcomes of students taught by a TFA or Teaching Fellows teacher and the outcomes of students taught by a comparison teacher can be attributed to differences in the teachers' effectiveness.[5] The study was conducted by Mathematica Policy Research, Branch Associates, and Chesapeake Research Associates.

[4] TNTP, formerly known as The New Teacher Project, is a national nonprofit organization that partners with urban school districts to operate local Teaching Fellows programs that recruit and train new teachers.

[5] Because teachers were not randomly assigned between TFA, Teaching Fellows programs, less selective alternative routes, and traditional routes, estimates of effectiveness reflect the combined effect of the characteristics and motivations of teachers who choose to enter teaching through each route and the recruitment and selection procedures and training and support provided by the programs themselves.

A. Background

1. Concerns about Teacher Quality in Hard-to-Staff Schools and Subjects

ED has a significant interest in programs and policies designed to address teacher shortages in high-need schools and hard-to-staff subjects, and in approaches for improving teacher quality more generally. In recognition of the difficulties schools face filling certain vacancies with qualified teachers, Title II of the No Child Left Behind Act of 2001 (NCLB) allows state funds to be used for "carrying out programs that establish, expand, or improve alternative routes for state certification of teachers and principals, especially in the areas of mathematics and science." Title II funds also support research on issues of teacher preparation and teacher quality. Because TFA and the Teaching Fellows represent a distinctive approach to addressing issues of teacher quality and teacher shortages in hard-to-staff schools and subjects, research on the success of their approach could assist ED in developing policies to address these issues.

Teacher quality is of particular concern in schools serving high concentrations of low-income students. On average, low-income students are taught by teachers with weaker qualifications—less teaching experience and lower certification examination scores—than students who are not economically disadvantaged (Clotfelter et al. 2007; Lankford et al. 2002; Carroll et al. 2000). Although the evidence is mixed, some studies have also found that teachers in high-poverty schools are, on average, less effective than those elsewhere (Glazerman and Max 2011; Sass et al. 2010). Taken together, the evidence suggests that high-poverty schools tend to have greater difficulty in attracting and retaining qualified, effective teachers (Monk 2007; Jacob 2007).

Math and science are among the subjects for which secondary schools—especially those serving lower-income populations—have the greatest difficulty retaining qualified teachers. Relative to the new supply of teachers in a given subject, rates of departure from teaching are highest among math and science teachers. Consequently, school principals report greater difficulty filling vacancies in math and science than in other subjects (Ingersoll and Perda 2009). For high-poverty schools in particular, these challenges are compounded by a net tendency for math and science teachers to transfer from high- to low-poverty schools (Ingersoll and May 2012).

2. Policies Targeting the Teacher Pipeline

One approach for improving teacher quality seeks to expand the supply of qualified, capable individuals who enter teaching. One of the most common policies to expand the teacher supply is the use of AC routes. Instead of requiring teachers to complete all certification-related coursework and student teaching before becoming full-time teachers—the traditional route to teaching certification (TC route)—these AC routes enable participants to begin full-time teaching before completing all the requirements for certification. Participants in a typical AC program, to the extent one exists, take a few weeks of coursework in the summer, possibly supplemented with some fieldwork—such as observing and helping in a summer school classroom—and then begin regular, full-time teaching in the fall. During their first (and sometimes second) year of teaching, they complete the remaining required coursework and then receive their teaching certificates. As of 2010, 45 states and the District of Columbia reported that they had approved AC routes (U.S. Department of Education 2011).

Most AC programs do not have very selective entrance requirements and consequently admit nearly all applicants (Mayer et al. 2003; Walsh and Jacobs 2007). A large-scale, random assignment evaluation of elementary school teachers from less selective AC programs—those that did not require applicants to have a grade point average above 3.0—found that teachers from these programs were neither more nor less effective than teachers in the same schools who entered teaching through a TC program and had about the same amount of teaching experience (Constantine et al. 2009).

3. TFA and the Teaching Fellows Programs

TFA, founded in 1989, and TNTP, founded by a TFA alumna in 1997, both seek to provide high-quality teachers to low-income schools. While TFA directly provides teachers to school districts, TNTP instead established locally run Teaching Fellows programs that recruit and train teachers, also known as Teaching Fellows, to teach in low-income schools. In 2011, TFA placed approximately 5,200 new teachers, or "corps members," in classrooms in 43 regions across the country; TNTP's 19 Teaching Fellows programs placed approximately 1,900 new Teaching Fellows in classrooms.

Both TFA and the Teaching Fellows programs are highly selective, admitting less than 15 percent of applicants nationwide. The programs have a multi-stage application process that includes in-person interviews, group discussions, writing exercises, and sample teaching lessons. Both programs apply specific selection criteria based on factors they believe to be associated with effective teaching.

Before entering the classroom, TFA and Teaching Fellows programs require participants to attend an intensive summer training program. These five- to seven-week summer institutes include both instruction and practice teaching. The programs typically do not provide certification; hence, TFA and Teaching Fellows participants must also enroll in a separate, state-authorized AC program. Both programs provide information, feedback, and mentoring to the teachers in their first one to two years of teaching.

Despite their many similarities, TFA and Teaching Fellows programs differ in three main ways:

1. **Recruitment focus.** Although both TFA and Teaching Fellows programs recruit new college graduates and more seasoned professionals, TFA focuses more on recruiting new college graduates, whereas the Teaching Fellows programs focus more on recruiting professionals who want to switch careers.
2. **Teaching commitment.** TFA requires its teachers to commit to two years in teaching. Teaching Fellows programs do not ask their teachers for a minimum time commitment but expect them to teach for many years.
3. **Choice of teaching location.** Individuals apply to a specific Teaching Fellows program in the locality in which they want to teach, but applicants to TFA apply to the national program and are assigned by the program to a region.

Appendix C provides a detailed description of the similarities and differences between the two programs.

Proponents of TFA and the Teaching Fellows programs argue that they are important sources of new teachers for disadvantaged schools, whereas critics of the programs argue that teachers who enter the profession through these programs and other AC routes are less well prepared than those who enter through TC programs and thus could be less effective (Darling-Hammond 1990). An additional criticism of TFA is that because it expects a commitment of only two years, TFA teachers are less likely than other teachers to remain in the profession and become more effective with experience (Heilig and Jez 2010).

4. Past Research on Teachers from TFA and the Teaching Fellows Programs

Despite the controversy over the effectiveness of teachers from TFA and the Teaching Fellows programs, the most rigorous available experimental evidence suggests that TFA teachers at the elementary school level are more effective than their non-TFA counterparts at teaching math, and neither more nor less effective at teaching reading. Research on the Teaching Fellows programs is more limited, with mixed results. No rigorous experimental study of TFA and Teaching Fellows programs has been conducted at the secondary level, a gap this study seeks to fill.

Experimental evidence. To date, there have been no large-scale experimental studies of the effectiveness of Teaching Fellows and only one experimental study of the effectiveness of TFA teachers, focused at the elementary school level.[6] Decker et al. (2004) randomly assigned nearly 1,800 students to either TFA teachers or teachers who received their certification through other routes within 17 elementary schools across seven districts. No limit was placed on the years of experience for either group of teachers, but 93 percent of the TFA teachers were "novices" (with no more than three years of experience), compared with 31 percent of the comparison teachers. The results showed that students of TFA teachers performed as well as students of non-TFA teachers in reading and scored statistically better in math (by approximately 0.15 standard deviations). A subgroup analysis of novice TFA teachers compared with novice non-TFA teachers found the difference in math scores was even greater (0.26 standard deviations) and still statistically significant; no discernible difference was found for this subgroup comparison in reading.

Nonexperimental evidence. Three recent nonexperimental studies have examined the student achievement of TFA and Teaching Fellows teachers based on a variety of analytic methods to control for nonrandom sorting of students to teachers. Boyd et al. (2006) and Kane et al. (2008) examined the achievement of students taught by TFA teachers in New York City public schools. Both used longitudinal data on students and compared students with teachers who entered the profession via different routes within the same school, controlling for the student's prior achievement. Boyd et al. (2006) found that, in math, students of TFA teachers scored higher than those of TC teachers at the middle school level, but lower at the elementary school level. In language arts, they found that students of TFA teachers scored lower than those of TC teachers at both the elementary and middle school levels. Kane et al. (2008) found that students

[6] Mathematica is currently conducting another large-scale, experimental evaluation of TFA elementary school teachers, as TFA attempts to scale up its program under an Investing in Innovation grant from ED's Office of Innovation and Improvement (Clark et al. 2012). This study will be completed by December 2014.

of middle school TFA teachers scored higher in math than students of TC teachers, and neither higher nor lower in reading. The third study examined TFA high school teachers in North Carolina (Xu et al. 2008). Because they were unable to control for prior student achievement, the authors instead compared the performance of students within the same subject with different types of teachers, controlling for their academic performance in other subjects. They found that students of TFA teachers generally scored higher than those of non-TFA teachers, and that differences in scores were most pronounced in math and science courses. Although all three of these studies attempt to adjust for unobserved differences in student characteristics, their nonexperimental designs leave open the possibility that observed differences in student achievement are due to unobserved student characteristics rather than differences in teacher effectiveness.

B. Our Evaluation

This evaluation seeks to address the gaps in the existing literature with experimental evidence on the effectiveness of secondary math teachers from TFA and the Teaching Fellows programs. Teacher effectiveness is measured as the difference in end-of-year math scores between students taught by TFA or Teaching Fellows teachers and comparable students taught by teachers from other routes. The study was *not* designed to compare the effectiveness of TFA teachers with that of Teaching Fellows. Because students were not randomly assigned between TFA teachers and Teaching Fellows teachers, any observed difference in their impacts could be due to differences in the students they teach, the teachers to whom they are compared, or the schools or districts in which they teach.

The study addresses two primary research questions:

1. How effective are TFA teachers at teaching secondary math compared with other teachers teaching the same math courses in the same schools?
2. How effective are Teaching Fellows at teaching secondary math compared with other teachers teaching the same math courses in the same schools?

To answer these questions, TFA teachers and Teaching Fellows were each compared with *comparison teachers* teaching the same subject in the same school. Comparison teachers could have entered teaching via a TC or a less selective AC route, but may not have entered through TFA or a Teaching Fellows program, or any other AC program with a similarly selective admissions process. No restriction was placed on the experience levels of the TFA teachers, Teaching Fellows, or comparison teachers. To explore possible reasons for differences in effectiveness, the study also examines differences in characteristics between TFA teachers and comparison teachers and between Teaching Fellows and comparison teachers. These characteristics include demographic characteristics, educational background, teaching and nonteaching work experience, and differences in their knowledge of mathematics. It also documents how TFA and Teaching Fellows programs select, train, place, and support their teachers.

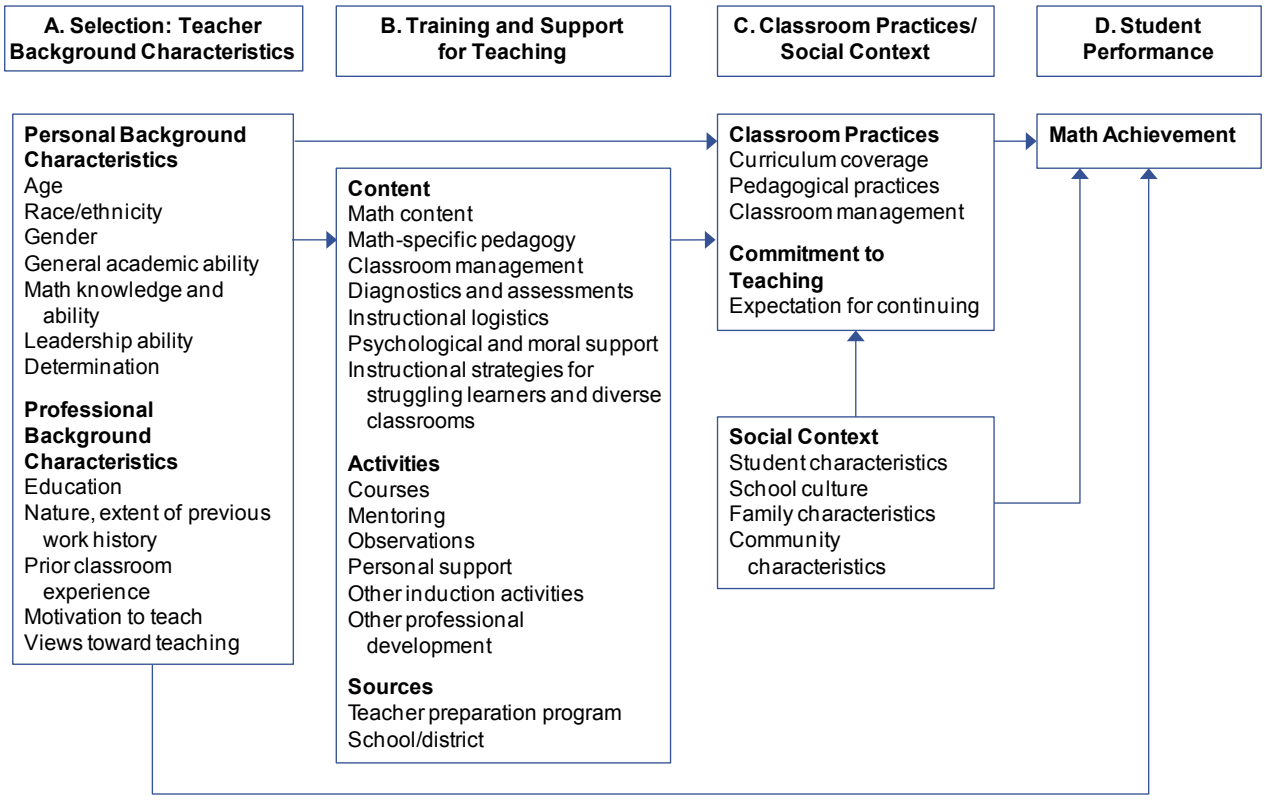
Finally, the study explores the relationship between student math achievement and observable teacher characteristics, such as educational background, teaching experience, and math content knowledge, as hiring authorities could potentially use these types of characteristics to predict an applicant's future effectiveness in the classroom. It also explores the extent to which these characteristics might explain any differences in teacher effectiveness that are found

between TFA teachers or Teaching Fellows and teachers from other routes to certification. That section of the study is based on nonexperimental methods and hence is less conclusive than the other components of the study.

The evaluation was guided by a conceptual framework that illustrates the mechanisms through which secondary math teachers, and the programs that select and train them, can influence students' achievement (Figure I.1). Students' achievement might be affected by the personal and professional background characteristics of their teachers (column A). These teacher characteristics might also influence the type of preparation program into which the teachers are selected. The preparation and support they receive through their teacher preparation programs and school districts (column B) may influence teachers' classroom practices (column C), which, in conjunction with social context, could in turn influence students' math achievement (column D).

Teachers who enter teaching via TFA and the Teaching Fellows programs can differ from those who enter teaching via other preparation programs in two main ways: (1) the professional and personal background characteristics they possess before entering the programs (column A) and (2) the training and support they receive from their programs (column B). Differences in background characteristics can arise from the programs' approaches to recruitment and selection as well as the teachers' decisions on which programs to attend. This study thus provides estimates of differences in effectiveness that reflect *both* differences in the types of individuals who enter TFA or Teaching Fellows programs versus other training programs and differences in the training and support their programs offer. It cannot rigorously disentangle these components.

Figure I.1. Conceptual Framework



Student math achievement is the ultimate outcome of interest that we used to estimate teacher effectiveness. Classroom practices and commitment to teaching are intermediate outcomes that we did not measure in this evaluation. Commitment to teaching may differ between TFA and other teachers due to the fact that TFA teachers are expected to make only a two-year commitment to the profession—this could negatively influence their effectiveness if they leave teaching before accumulating valuable experience or positively influence their effectiveness if, for instance, the short-term nature of their commitment increases their motivation and enthusiasm for teaching.

As shown in Figure I.1, both the intermediate and final outcomes can be influenced by the social context that is, the characteristics of the students and their schools, families, and communities. In this study, the random assignment of students ensured that similar students in the same schools were assigned to the TFA or Teaching Fellows teachers' classes and the comparison teachers' classes and thus enabled us to control for the social context that might influence student achievement.

In addition to highlighting potential approaches for improving teacher quality in hard-to-staff subjects within high-need schools, the findings from this evaluation are informative for policymakers, funders of teacher preparation programs, school district administrators, and school principals. They can aid policymakers and funders of teacher preparation programs by providing information on the effectiveness of teachers from routes to certification with differing methods of identifying, attracting, training, and supporting their teachers. They can also inform district officials and school principals faced with the choice between hiring secondary math teachers from TFA or teachers who entered teaching through other routes, and with the choice between hiring teachers secondary math from Teaching Fellows programs or teachers who entered through other routes.

C. Looking Ahead

The rest of this report describes in detail the study's design and its findings. Chapter II explains the experimental design, data collection, and analytic approaches. Chapters III, IV, and V focus on TFA. Chapter III describes the regional TFA programs represented in the study, Chapter IV describes the participating TFA teachers and the teachers with whom they were compared, and Chapter V presents the estimates of the impact of TFA teachers on student math achievement. Chapters VI through VIII then present similar analyses for the Teaching Fellows programs and their teachers. Chapter IX presents the nonexperimental analysis of the teacher characteristics that are correlated with student achievement.

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

II. STUDY DESIGN, DATA, AND METHODS

This chapter describes the research design, data, and analytic approaches used in the study. Section A describes the groups of teachers included in the study. Section B discusses the study's experimental design. Section C describes our strategies for recruiting and selecting districts, schools, and teachers into the study, and Section D explains the procedures by which students were randomly assigned to the participating teachers' classrooms. Section E specifies the data that were collected, and Section F describes the methods used to analyze these data to address the study's research questions.

A. Types of Teachers in the Study

This study separately compared the effectiveness of Teach For America (TFA) teachers and Teaching Fellows with the effectiveness of teachers from other certification routes. Because the effectiveness of the latter group served as the point of comparison for the effectiveness of TFA teachers and Teaching Fellows, we refer to teachers from certification routes other than TFA and the Teaching Fellows programs as *comparison teachers*. In designing the study and deciding which teachers to include in the sample, our goal was to ensure that the comparison teachers represented a meaningful and appropriate counterfactual—the types of teachers that would have been available had teachers from TFA or the Teaching Fellows programs not been available in a particular school. Below we describe our eligibility criteria for the teacher sample, for both TFA/Teaching Fellows teachers and comparison teachers. In particular, we discuss (1) eligible routes to certification and (2) our decision not to limit the teacher sample by years of teaching experience.

1. Eligible Routes to Certification

All secondary math teachers who entered teaching through TFA or a Teaching Fellows program were potentially eligible to be included in the study sample. In particular, the TFA sample included teachers who were still fulfilling their two-year commitment to the program (TFA corps members) and those who remained in teaching after completing their two-year commitment (TFA alumni).

To be eligible to serve as comparison teachers in the study, teachers could have entered teaching through any certification routes other than TFA, a Teaching Fellows program, or another alternative route to certification that was similar to TFA and the Teaching Fellows programs in terms of its selectivity. The routes that were excluded from the study due to their similarity to TFA and the Teaching Fellows programs were small in both number and size; we excluded five programs that, collectively, trained only 14 math teachers nationwide in 2007 (Clark et al. 2008).[7] Teachers from all other routes to certification were potentially eligible to be included in the sample of comparison teachers—in particular, teachers from both traditional

[7] The five excluded programs were the Mississippi Teacher Corps, the Massachusetts Initiative for New Teachers, Los Angeles' TeachLA program, the St. Louis Career Transition Certification Program, and New York City's Teaching Opportunity Program.

routes to certification (TC teachers) and less selective alternative routes to certification that accepted the majority of their applicants (AC teachers).[8]

These comparison group eligibility criteria satisfied two objectives. First, by excluding a small number of routes that were similar to TFA and the Teaching Fellows programs, we ensured that the teachers being compared in the study entered teaching through meaningfully different routes. Second, because there were no other restrictions on the comparison teachers' routes to certification, the comparison teachers in the study reflected the pool of teachers who could have taught study students had the TFA or Teaching Fellows teachers not been available.

2. No Limits on Prior Teaching Experience

We did not impose any restrictions on the amount of prior teaching experience that teachers in the study could possess. Therefore, any disparities in teaching experience between TFA or Teaching Fellows teachers and comparison teachers could contribute to the differences in effectiveness that we estimated in this study. The benefit of this approach is that the study's findings on differences in effectiveness reflect, among other factors, the real experience gaps between secondary math teachers from TFA, Teaching Fellows programs, and other routes to certification in the same schools. Below, we elaborate on the implications of this approach separately for TFA and Teaching Fellows teachers.

A common criticism of TFA is that it seeks teachers willing to make a two-year rather than a longer-term commitment to teaching. Critics claim that too many TFA teachers leave teaching before they accumulate the experience needed to be as effective as their counterparts from other routes (Heilig and Jez 2010). As expected, the TFA teachers in the study were predominantly in their first two years of teaching and had significantly less experience than comparison teachers (see Chapter IV). The study's comparison of less experienced TFA teachers with more experienced non-TFA teachers therefore highlights the choice that a hiring authority faces when selecting what type of teachers should fill vacancies, given that the TFA teachers may be more likely to leave after only two years.

Because the Teaching Fellows programs are relatively new, impact findings for Teaching Fellows must be interpreted with more caution. Unlike TFA, the Teaching Fellows programs encourage a long-term commitment to teaching. Accordingly, the Teaching Fellows in the study included a diverse mix of less experienced and more experienced teachers (see Chapter VII), reflecting the current distribution of experience among Teaching Fellows in the study schools. Although Teaching Fellows in the study had less experience than comparison teachers, it is unclear whether this gap would persist in the long run. The relatively lower experience levels of the Teaching Fellows may be due to the fact that no Teaching Fellows program existed before 2000, effectively limiting the amount of experience that Teaching Fellows teachers in the study could have had. In that case, it may not be appropriate to generalize the findings from this study to future periods in which experience gaps between Teaching Fellows and comparison teachers could be smaller.

[8] Throughout this report, the term "AC teachers" refers only to comparison teachers who entered teaching through a less selective AC route that accepted the majority of its applicants.

B. Experimental Design

To assess the impacts of TFA and Teaching Fellows teachers compared with teachers from other certification routes, the study used an experimental design based on the random assignment of students to teachers. Students in the same school who enrolled in the same math course were randomly assigned to a math class taught by a TFA or Teaching Fellows teacher or to a comparable class taught by a comparison teacher. Random assignment ensured that there were no systematic differences at the start of the school year between students assigned to TFA or Teaching Fellows teachers and those assigned to teachers from other certification routes. Therefore, differences in end-of-year math scores between these groups capture the causal impact on students' math achievement of being assigned to a TFA teacher or Teaching Fellow rather than a teacher who entered teaching through some other route. This impact represents the difference in effectiveness between (or the relative effectiveness of) the teachers being compared.

We refer to a group of classes between which students could be randomly assigned as a *classroom match*. The study included two main sets of classroom matches—those that included TFA and comparison teachers, and those that included Teaching Fellows and comparison teachers. Within both types of matches (TFA or Teaching Fellows), students who were randomly assigned to a TFA or Teaching Fellows teacher constituted the treatment group, and those who were randomly assigned to a comparison teacher constituted the control group. As discussed in more detail in Appendix A, classes of TFA teachers or Teaching Fellows and those of comparison teachers in the same school could form an eligible classroom match only if the classes met the following criteria:

- **The classes covered a math course that was deemed eligible for the study.** Any middle school math course was eligible. At the high school level, the eligible courses were general math (such as pre-algebra or remedial math), Algebra I, Geometry, and Algebra II. The study did not include more advanced courses, such as trigonometry or calculus, because appropriate assessments for these subjects were not available. Full-year classes were eligible, as well as classes that covered a full year's worth of material in a single semester.
- **The classes were similar in subject taught and class conditions.** In particular, the classes needed to cover the same general topic (for instance, Algebra I), be taught at the same level (for instance, honors or regular), be of similar sizes, have the same number of teacher's aides with the same roles (for instance one teacher aide who assisted all students), be the same length (one or two semesters), and have the same arrangements for inclusion of English language learners and special education students.
- **The classes were in the same period.** In most cases, the classes needed to be in the same period in order for the study team to be able to randomly assign students to classes in the match without disrupting the students' schedules in other periods. For example, it would not be possible to randomly assign students to either math in period 1 or math in period 2 if all of the Spanish classes were in period 2. There were a few exceptions to this rule, which we describe in Appendix A.

- **Students in the classes did not receive supplemental math instruction in a way that would systematically inflate or dampen the effects of their regular math teachers.** Many schools provide supplemental math classes to students to reinforce the content taught in the main classes. In some circumstances, this instruction could cause the impacts of the regular math teachers to be systematically overestimated or underestimated. For example, if students were assigned to supplemental math instruction during the school year to compensate for poor performance in their main math class, then ineffective teachers would appear to be better—due to the extra help provided to their students—than they actually are. The study included classes in schools with supplemental instruction only if the supplemental instruction would not lead to systematic errors in the impact estimates (see Appendix A).

Although most classroom matches consisted of only two classes, some contained more than two. Ninety-three percent of the matches in the TFA study sample and 83 percent in the Teaching Fellows study sample consisted of exactly one class taught by a TFA or Teaching Fellows teacher and one class taught by a comparison teacher (Appendix Tables A.3 and A.4). The remaining matches consisted of more than two classes. This could occur if, for example, students were randomly assigned among three comparable math classes—such as three sections of Algebra I taught by two TFA teachers and one comparison teacher—held at the same time.

C. Recruitment of Districts, Schools, and Teachers

We conducted the study over a two-year period and recruited two cohorts of sample members—one cohort that participated in the 2009–2010 school year (cohort 1) and a second cohort that participated in the 2010–2011 school year (cohort 2). A separate sample of districts, schools, and teachers was recruited into the study for each of the two study years, but sample members could be included in both years. The final sample consisted of 11 states, 15 districts, 82 schools, 228 classroom matches, 287 teachers, and 517 classes (Table II.1, last column).

The sample consisted of two distinct sets of study participants. One set of participants, referred to as the TFA study sample, consisted of TFA and comparison teachers who taught matched classes and included the students who were assigned to those classes and the schools and districts in which those classes were located. The other set of participants, referred to as the Teaching Fellows study sample, consisted of Teaching Fellows and comparison teachers who taught matched classes, along with the corresponding students, schools, and districts. A comparison teacher could belong to both study samples if he or she was compared with both TFA and Teaching Fellows teachers.[9]

Study participants were recruited in similar ways for both the TFA and Teaching Fellows study samples. Next, we describe the recruitment of districts, schools, and teachers into the study. Appendix A provides details on the numbers of districts, schools, and potential classroom matches that were involved in each stage of recruitment.

[9] Fewer than three comparison teachers were included in both the TFA and Teaching Fellows samples; the exact number is not reported in accordance with National Center for Education Statistics statistical standards to protect respondent confidentiality (National Center for Education Statistics 2000).

Table II.1. Number of Districts, Schools, Classroom Matches, Teachers, and Classes in the Study

	Number of Sample Members in		
	TFA Study Sample	Teaching Fellows Study Sample	Both Samples Combined[a]
States	8	8	11
Districts	11	9	15
Schools	45	44	82
Classroom Matches[b]	111	118	228
With TC comparison teachers	58	92	149
With AC comparison teachers	53	30	83
Teachers	136	153	287
TFA or Teaching Fellows teachers	66	69	135
Comparison teachers	70	84	152
TC teachers	41	61	100
AC teachers	29	23	52
Classes	248	270	517
Taught by TFA or Teaching Fellows teachers	123	135	258
Taught by comparison teachers	125	135	259

[a]Counts of sample members in the TFA and Teaching Fellows study samples can sum to more than the total count in the combined sample because some comparison sample members belonged to both the TFA and Teaching Fellows study comparison samples.

[b]Counts of matches with TC teachers and those with AC teachers sum to more than the total number of matches because some matches contained both TC and AC teachers.

AC = alternative route to certification; TC = traditional route to certification; TFA = Teach For America.

1. Recruitment of Districts

We focused recruitment efforts on districts with large concentrations of secondary math teachers from TFA or a Teaching Fellows program. Using fall 2008 placement data from TFA and the Teaching Fellows programs, we identified districts with the largest numbers of secondary math teachers trained by these programs, and we contacted 42 of these districts prior to the first study year. Of those districts, 15 agreed to participate in the study by allowing the study team to conduct random assignment and data collection activities (Table II.1).

2. Recruitment of Schools

Within participating districts, we contacted schools in the spring prior to each study year to identify those in which the study could be implemented in the upcoming year. We placed priority on contacting schools in which TFA and Teaching Fellows programs had previously placed secondary math teachers because those schools had the greatest likelihood of having teachers eligible for the study in the upcoming study year. In each contacted school, we ascertained course schedules and teaching assignments to determine whether the school would have any eligible classroom matches in the following school year. Of the 792 schools that were initially contacted, the final sample of 82 schools consisted of those that contained 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 (Table II.1).

Because schools could be included in the study only if they had TFA teachers or Teaching Fellows, the distribution of participating schools across the 15 study districts was influenced by the number of TFA teachers and Teaching Fellows who taught in those districts. Some districts had substantially larger populations of teachers from these programs than other districts, contributing to an uneven distribution of study schools across districts. This was particularly true in the Teaching Fellows study sample; of the 44 schools with Teaching Fellows in the study, 22 belonged to a single district (Table II.2). As a result, this district contributed 54 percent (82 of 153) of the teachers in the Teaching Fellows study sample.

Table II.2. Distribution of Sample Sizes Across Districts in the Study

Type of Sample Size	Number of Districts	Distribution of Schools and Teachers Across Districts			
		Mean	Median	Minimum	Maximum
In Teach For America Study Sample:	11				
Number of study schools per district		4	3	1	12
Number of study teachers per district		12	8	2	41
In Teaching Fellows Study Sample:	9				
Number of study schools per district		5	3	1	22
Number of study teachers per district		17	8	2	82

As we discuss in Chapters IV and VII, the schools in the study were, on many dimensions, similar to all secondary schools with TFA and Teaching Fellows teachers. However, our recruitment approach led to an underrepresentation of certain types of schools. First, smaller schools were less likely to have multiple sections of the same math course—a requirement for classroom matches to be formed—and were thus underrepresented. Second, although we contacted charter schools to determine whether they could be included in the study, none had eligible matches. Finally, the schools in the study were disproportionately from the large, urban districts that the study targeted in its recruitment.

3. Classroom Matches and Teachers in the Final Study Sample

The final set of classroom matches in the study spanned several types of middle school and high school math courses (Table II.3). Courses were represented in different proportions in the TFA and Teaching Fellows study samples. In particular, middle school math courses constituted 75 percent of classroom matches with TFA teachers and 31 percent of classroom matches with Teaching Fellows.

Table II.3. Number of Classroom Matches and Study Classes, by Subject

Subject	Teach For America Study Sample			Teaching Fellows Study Sample		
	Classroom Matches	Classes with TFA Teachers	Classes with Comparison Teachers	Classroom Matches	Classes with Teaching Fellows	Classes with Comparison Teachers
6th-Grade Math	31	33	35	9	11	10
7th-Grade Math	33	33	34	14	18	14
8th-Grade Math	19	20	19	14	21	16
Algebra I or General High School Math	9	14	14	34	38	46
Geometry	14	18	18	23	23	23
Algebra II	5	5	5	24	24	26
Total	**111**	**123**	**125**	**118**	**135**	**135**

TFA = Teach For America.

Of the 287 teachers in the study, about half (135) were either TFA or Teaching Fellows teachers (Table II.1). The number of TFA teachers (66) was similar to the number of Teaching Fellows (69). The total number of teachers in the study (287) was not twice the number of classroom matches (228) because some matches had more than two teachers. Moreover, some teachers in the study taught classes in more than one match—in different periods during the school day and/or in both study years.

Our approach to recruiting may have led to an underrepresentation of TFA and Teaching Fellows teachers in their first year of teaching as well as more experienced TFA teachers. Because we placed priority on contacting schools in which TFA or Teaching Fellows teachers had been previously placed as indicated by the programs' placement data, we might not have been aware of schools that began hiring new teachers from these programs during the study's recruitment period. In addition, although TFA was aware of many schools with more experienced TFA teachers, it maintains less comprehensive information on teachers who have completed their two-year commitment to the program. Thus, more experienced TFA teachers may also have been underrepresented in the sample.

4. Differences between the TFA and Teaching Fellows Study Samples

We evaluated the impacts of TFA and Teaching Fellows teachers separately, in part, because of key differences in the study samples used for evaluating these two groups. Chapters IV and VII discuss the characteristics of the sample members in detail, but several differences between the TFA and Teaching Fellows study samples are evident in Tables II.1 through II.3. First, the TFA teachers and Teaching Fellows in our study taught largely in different districts. Of the 15 districts in the study, only 5 included both TFA and Teaching Fellows teachers. Second, even within a district, TFA and Teaching Fellows teachers taught in different schools. Seventy-five of the 82 study schools (91 percent) had either TFA or Teaching Fellows teachers in the study, but not both; as discussed earlier, TFA teachers were more likely than the Teaching Fellows to teach in middle schools (Table II.3). Third, the TFA and Teaching Fellows teachers were compared with different groups of comparison teachers. Whereas AC teachers represented 41 percent (29 of 70) of the comparison teachers with whom TFA teachers were compared, they represented

27 percent (23 of 84) of the comparison teachers with whom Teaching Fellows were compared (Table II.1). Differences in impact findings from the two study samples could be due to any of these differences in sample characteristics, as well as other unobserved differences stemming from the fact that students were not randomly assigned between TFA and Teaching Fellows teachers. For all of these reasons, study findings cannot be used to compare the effectiveness of TFA and Teaching Fellows teachers.

D. Selection and Assignment of Students

Before the start of each new school year, schools sent us lists of students whom they wanted placed into one of the classes in the identified classroom matches. We randomly assigned the students into the classes, specifying the teacher for each class. The schools then assigned the students to classes in accordance with the random assignment results. Students who wanted to enter one of the classes after this initial assignment but before the end of the first month of the school year were also randomly assigned to one of the classes. Schools could explicitly request a specific assignment for a given student, in which case the student was excluded from the study (which was rare, as described below). Appendix A provides further details on the random assignment process.

Based on the random assignment conducted in the study, a student was included in the *research sample*—that is, the beginning-of-year sample representing the students to whom study findings pertain—if he or she met two criteria. First, the student was randomly assigned to a study class by the end of the first month of the school year. Second, the student did not leave the school in which the study was being conducted prior to the first day of the school year. With these criteria, the research sample consisted of 5,790 students in the TFA study sample and 6,909 students in the Teaching Fellows study sample (Table II.4). As discussed in Section E, all research sample students with valid end-of-year math scores were included in the impact estimates; these students are referred to as the *analysis sample*.

Table II.4. Number of Students in the Study

	Teach For America Study Sample			Teaching Fellows Study Sample		
Subject	Assigned to TFA Teachers	Assigned to Comparison Teachers	Total	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Total
Number of Students in the Research Sample[a]	2,884	2,906	5,790	3,466	3,443	6,909
Number of Students in the Analysis Sample[b]	2,292	2,281	4,573	2,127	1,989	4,116

Note: 24 students from the research sample—of whom 20 were in the analysis sample—were assigned to a comparison teacher who was in both the Teach For America study sample and the Teaching Fellows study sample.

[a]Research sample includes all students who were randomly assigned to a study class by the end of the first month of school and did not leave the school prior to the start of the school year.

[b]Analysis sample includes all students in the research sample for whom the study obtained valid end-of-year math achievement scores.

TFA = Teach For America.

The vast majority of students (97.4 percent) who enrolled in a study class before or during the first month of school were randomly assigned and are included in the research sample. Rates were similar in the classes of TFA teachers and Teaching Fellows (97.2 percent) and the classes of comparison teachers (97.5 percent) (Table II.5). The remaining students either left the study school before the start of the school year or were excluded from random assignment at the school's request.

Table II.5. Changes in Composition of Study Classes During the School Year

	Average Number of Students per Class (unless otherwise indicated)		
	All Study Classes	Classes of TFA Teachers or Teaching Fellows	Classes of Comparison Teachers
Enrolled in Study Classes Before the End of the First Month of School[a]			
Number of students	25.2	25.3	25.2
Number of students who belong to the research sample	24.6	24.6	24.5
Percentage of students who belong to the research sample	97.4%	97.2%	97.5%
Listed on End-of-Year Class Rosters			
Number of students	23.3	23.3	23.2
Number of students who belong to the research sample and stayed in originally assigned class	19.0	19.1	18.9
Percentage of students who belong to the research sample and stayed in originally assigned class	81.5%	81.8%	81.3%
Number of Classes	**517**	**258**	**259**

Note: Calculations are unweighted. Tests of differences between TFA/Teaching Fellows and comparison classes were conducted using independent-sample *t*-tests with unequal variances. No differences were statistically significant at the 0.05 level.

[a]This category excludes students who were enrolled in study classes before the start of the school year but never showed up to school.

TFA = Teach For America.

There was some student movement into and out of the study classes after the random assignment period. Some research sample members transferred out of their originally assigned classes and some late-enrolling students were placed by schools into study classes after the first month 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, 81.5 percent of students in study classes were research sample members who had been originally assigned to those classes, with similar rates in the classes of TFA teachers and Teaching Fellows (81.8 percent) and the classes of comparison teachers (81.3 percent) (Table II.5). In Appendix A, we document the types of mobility experienced by research sample members.

E. Data Used in the Study

We collected data on students, teachers, schools, and the TFA and Teaching Fellows programs in the study (Table II.6). Appendix A provides details on the response rates for each major data source.

1. Data on Students

We attempted to collect data on math achievement and demographic characteristics for all students in the research sample for whom we received parental consent to collect this data.

Math achievement outcomes. We measured end-of-year math achievement based on different sources of data for middle school and high school students in the study. For students in grades 6 to 8, we obtained scores on state assessments from district administrative records. These assessments, required by federal law, were part of states' accountability systems, which gauge school performance in part based on students' test scores. The assessments were thus designed to measure the math skills that students were expected to learn at their grade level in each state. Because we requested district records, it was possible to obtain scores for any consenting middle school students in the study who had taken state assessments within a participating district, even if they had switched schools within the district during the school year.

Table II.6. Data Sources for the Evaluation

Domain	Data Source	Schedule of Data Collection	
		Cohort 1 (2009–2010)	Cohort 2 (2010–2011)
Student Math Achievement Outcomes			
Middle school	District administrative records	Summer/Fall 2010	Summer/Fall 2011
High school	NWEA assessments (study-administered)	Spring 2010[a]	Spring 2011[b]
Baseline Student Achievement and Characteristics	District administrative records	Summer/Fall 2010	Summer/Fall 2011
Student Mobility	Class rosters	Summer 2009, Fall 2009, Winter 2010, Spring 2010	Summer 2010, Fall 2010, Winter 2011, Spring 2011
Teachers' Route to Certification	Teacher background form	Summer/Fall 2009	Summer/Fall 2010
Teachers' Professional Background and Experiences	Teacher survey	Spring 2010	Spring 2011
Teachers' Math Content Knowledge	Praxis (study-administered or existing score from Educational Testing Service)	Fall 2009	Fall 2010
School Characteristics	Common Core of Data	2009–2010	2009–2010
TFA and Teaching Fellows Program Characteristics	Program administrator interviews	Spring 2010	n.a.

[a]Assessments for single-semester fall 2009 classes were conducted in winter 2009–2010.

[b]Assessments for single-semester fall 2010 classes were conducted in winter 2010–2011.

TFA = Teach For America; NWEA = Northwest Evaluation Association.

n.a. = not applicable.

For students in grades 9 to 12, we administered end-of-course math assessments developed by the Northwest Evaluation Association (NWEA). Based on the content of their math course, students took computer-adaptive assessments in general high school math, Algebra I, Geometry, or Algebra II. We attempted to administer assessments to all consenting high school students in the study who had not moved out of the school district by the time of testing, regardless of whether they had stayed in their original class or school. We chose not to use high school students' scores on state assessments for several reasons. First, high school state math

assessments were not consistently available for all states and grades in the study, as federal law required states to assess students in math in only one grade at the high school level. Second, we expected that some high school state assessments would measure minimum overall math competency required for graduation and, hence, would not be aligned with the content of particular courses.

Students in the research sample with valid outcome scores constituted the final sample used in the analysis. A student may not have had a valid outcome score for one of four reasons: (1) the student's parents did not give consent for data collection; (2) we were not able to administer an NWEA assessment to the student (if in high school); (3) the student's score on the NWEA assessment was invalid based on criteria defined by NWEA (Table A.12); or (4) the district did not have a state assessment score for the student (if in middle school). We obtained valid outcome scores for 68 percent of students in the research sample—in particular, for 70 percent of treatment group students and 67 percent of control group students (Figure A.2). As we discuss in Chapters IV and VII, treatment and control group students with outcome data were statistically similar on most baseline characteristics, suggesting that patterns of missing outcome data did not generate imbalance between these groups. Table A.17 provides further comparisons of the characteristics of students with and without outcome data.

Achievement and demographic characteristics at baseline. We collected information on baseline measures—measures whose values were determined prior to random assignment—of students' achievement and demographic characteristics. Baseline data for all students were based on district administrative records. These data served as covariates in the impact analyses, provided contextual information on the types of students included in the study, and indicated whether random assignment produced treatment and control groups that were similar at the beginning of the school year.

Baseline achievement was measured by students' prior scores on state assessments in math and reading. We used each student's math and reading score from the most recent prior grade at which end-of-grade state assessments were administered. Appendix B provides details on the construction of the baseline achievement measures. Because students' prior scores in both math and reading were predictive of their outcome math scores, controlling for baseline achievement in both subjects improved the precision of the impact estimates. Other characteristics available from district administrative records included age, gender, race and ethnicity, free and reduced-price lunch eligibility, special education status, and English language learner status.

Student mobility. To monitor compliance with random assignment and the mobility of students into and out of the classes in the study, we obtained class rosters from participating schools several times during the school year. Each roster provided an updated list of students who were enrolled in a specified class. We collected rosters four times: (1) immediately after random assignment; (2) at the end of the first month of school in the fall; (3) in the first week of classes after winter break; and (4) approximately eight weeks before the end of the school year.

2. Data on Teachers

Route to certification. Before each year of the study, we verified the certification route of all teachers whose classes could potentially be included in classroom matches by asking principals to complete a background form on each study teacher.

Professional background and experiences. In the spring of each 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, nonteaching and teaching work experience, preparation for teaching, support received during the school year, views toward teaching, and demographic characteristics.

Math content knowledge. Some researchers have argued that the effectiveness of math teachers is related to their knowledge of math concepts (Hill et al. 2005). To measure the math content knowledge of teachers in the study, we used teachers' scores on one of two assessments developed by the Educational Testing Service: (1) the Praxis II Middle School Mathematics assessment (for middle school math teachers) and (2) the Praxis II Mathematics Content Knowledge assessment (for high school math teachers). These assessments were designed to measure general and subject-specific teaching skills and knowledge appropriate for beginning math teachers. As of 2006, 37 states required new secondary math teachers to pass one of these tests as part of the licensure and certification process (Gitomer and Qi 2010).

The collection of Praxis II scores varied across states depending on whether the state required secondary math teachers to pass a Praxis II math assessment for certification. In states that required math teachers at a particular school level (middle or high) to pass the relevant Praxis II exam (Middle School Math or Math Content Knowledge, respectively), we asked teachers to release their existing scores to the study.[10] In states that did not require the relevant Praxis II exam, we administered the tests directly to teachers in the study.

3. Data on Schools

To describe the schools included in the study sample, we collected data on school characteristics, including grade span, school size, and general demographic information on the student body. We obtained these data from the 2009-2010 Common Core of Data. The Common Core of Data is a publicly available database covering all public schools in the United States.

4. Data on TFA and Teaching Fellows Programs

Through semi-structured telephone interviews with program officials, we collected detailed information about TFA and the Teaching Fellows programs to document how they selected, trained, placed, and supported program participants. The interviews solicited information on both national and local/regional program operations. Staff at the TFA national office and TNTP, the national organization that oversees the Teaching Fellows programs, identified the most appropriate respondents for the interviews. In some cases, national office staff gave us information about local/regional program operations, especially if policies and practices were standardized across different locations. In other cases, we spoke directly with local or regional program staff. We obtained information on all 10 regional TFA programs and all seven Teaching Fellows programs represented in the study, as well as national-level operations of both TFA and TNTP.

[10] One state in the study required middle school math teachers to pass the Math Content Knowledge assessment (rather than the Middle School Math assessment). In that case, we used scores on the Math Content Knowledge assessment for teachers at all grade levels.

F. Overview of the Analytic Approach

The study included two key types of analyses: (1) experimental analyses to estimate the impacts of TFA teachers and Teaching Fellows on student math achievement in the full sample and key subgroups, and (2) nonexperimental analyses to explore the association between teacher characteristics and teacher effectiveness. Appendix B provides more technical details of the analytic methods.

1. Experimental Analyses

The experimental analyses provided separate estimates for the impacts of TFA teachers and Teaching Fellows in the study relative to the comparison teachers in the same classroom matches. Random assignment was the key to the causal validity of these estimates because it ensured that students assigned to the TFA or Teachings Fellows teachers were no different, on average, than students assigned to comparison teachers in the same match at the time of random assignment. The analytic approaches we describe next were applied identically to estimating the impacts of the TFA teachers and Teaching Fellows in the study.

Outcome measure. Students' math assessment scores constituted the outcome measure for the analysis. The scales of the test scores differed between the state assessments and NWEA assessments and, among state assessments, differed across states and grade levels. To express test scores in a common unit, we converted each score into a *z*-score by subtracting the mean score of a reference population and dividing the difference by the standard deviation of scores in that reference population. For a student's score on a state assessment, the reference population was the full population of students in the same state, year, and grade who took the same assessment; for a student's score on an NWEA assessment in a given course, the reference population was the NWEA's nationwide norming sample for that assessment. Thus, impacts on *z*-scores in this analysis represented increments to math achievement expressed in standard deviations within a statewide or national student population.

Estimation method. Because teachers in the same classroom matches were assigned similar students at the beginning of the year, the simple average difference in end-of-year achievement between their students could provide causally valid estimates of their relative effectiveness. In our main analysis, we separately estimated the average impacts of TFA teachers and Teaching Fellows relative to comparison teachers on the basis of regression models that built upon the simple test score differences. The regression models pooled data across all classroom matches involving teachers from the specified group and the comparison teachers with whom they were compared. The model controlled for classroom match indicators so that comparisons were made only within the same match. The average impact estimated by each model was therefore similar to a weighted average of impacts from all relevant classroom matches, such that matches with more students received more weight. To enhance the precision of the estimates, the models controlled for students' baseline characteristics, including prior test scores in math and reading and several demographic characteristics.

Subgroup analyses. Impacts of TFA or Teaching Fellow teachers relative to comparison teachers could depend on the specific characteristics of the teachers who are compared with each other. Therefore, we applied similar analytic methods to estimate differences in effectiveness between TFA or Teaching Fellows teachers and comparison teachers in subgroups of matches defined by specific teacher characteristics. To define the subgroups, we chose a number of

characteristics that could potentially influence the magnitude or direction of the impacts. First, we estimated separately the impacts of TFA teachers and Teaching Fellows relative to comparison teachers from each of the two major types of certification routes—TC and less selective AC routes—as the two routes could attract very different types of comparison teachers and provide very different types of training. Second, given that a key expected difference between TFA or Teaching Fellows teachers and comparison teachers is the gap in their experience levels, we defined subgroups by experience to understand whether our impact findings would change if teachers with more similar levels of experience were compared with each other. Third, we estimated impacts within subgroups delineated by school level (middle or high)because math instruction at different levels might require different sets of skills.

Adjustments for spurious significant findings. Estimating multiple impacts increases the chance of falsely identifying an impact as significant when, in fact, there is no true impact. In this study, we estimated many impacts—for TFA and Teaching Fellows teachers separately and for various subgroups. Corrections can be made to reduce the chance of obtaining spurious significant findings, but adjusting for all impact estimates included in the study would have severely limited the study's power to detect any impacts at all. To balance these competing factors, we adjusted the significance tests (using the Benjamini and Hochberg [1995] method) only for the impact estimates we considered to be of primary policy relevance the impacts of TFA teachers or Teaching Fellows relative to all comparison teachers, TC teachers, and AC teachers.[11] We regarded the subgroups defined by the comparison teachers' route to certification (AC or TC) to be the most policy relevant of the subgroups we examined because of their potential to inform principals' hiring decisions. We did not adjust significance tests in any of the remaining subgroup analyses because we did not consider them to be of primary policy relevance and did not want to compromise the study's power to detect impacts for the groups of most interest. Hence, the significant findings from these remaining analyses are more likely to be spurious than the main significant findings. Further discussion of these adjustments is provided in Appendix B.

Accounting for students who switched to a different type of teacher. Our main estimates capture the relative effects of TFA or Teaching Fellows teachers and comparison teachers on all students who were randomly assigned to their classes, including students who left their original classes and switched to a different type of teacher during the school year. The advantage of these estimates is that they are based purely on the experimental design. In supplemental analyses, we also estimated impacts on only students who stayed with their originally assigned type of teacher.

2. Nonexperimental Analyses

We used nonexperimental methods to examine the extent to which teacher effectiveness was associated with teacher characteristics other than their routes to certification. These additional types of information could potentially be useful to hiring authorities in predicting applicants' likely effectiveness in the classroom. Accordingly, we focused on analyzing characteristics

[11] None of the impact findings changed as a result of these adjustments. In other words, no impact estimates that were statistically significant prior to the adjustments became insignificant after the adjustments.

observable at the time of hiring, including educational background, performance on math content knowledge tests, professional experience, and instructional training. Understanding the teacher characteristics associated with teacher effectiveness may also shed light on the reasons for any difference in effectiveness between teachers who enter teaching via different routes.

The objectives of this analysis differed from those of the experimental subgroup analysis described earlier. The experimental subgroup analysis was aimed at examining dimensions on which the *impacts* of TFA or Teaching Fellows teachers relative to comparison teachers could vary. Accordingly, for that analysis, we focused on a small set of characteristics that could lead to larger or smaller differences in effectiveness between TFA or Teaching Fellows teachers and comparison teachers. For the nonexperimental analysis, we sought to identify factors that could generally predict teacher effectiveness, regardless of the teachers' certification route, to generate hypotheses for future research. Therefore, as described below, this analysis focused on a broader set of characteristics.

The associations found by the nonexperimental analysis cannot provide rigorous evidence for the causal effects of teacher characteristics on student achievement. Any association between a given teacher characteristic and student achievement could be due to some unobserved factor rather than the one under study. Nonetheless, this analysis may suggest hypotheses that could be examined in future research.

We conducted two basic analyses in sequence: (1) estimating relationships between teacher characteristics and effectiveness based on the full sample of teachers in the study; and (2) assessing the extent to which those relationships could account for the actual, experimentally estimated differences in effectiveness between specific groups of teachers. We describe the methods for these analyses next.

Estimating relationships between teacher characteristics and effectiveness. We estimated regression models in which student math scores were the dependent variable and the characteristics of the students' math teachers were the key independent variables of interest. The models also controlled for classroom match indicators and all variables included in the experimental impact models. The regression coefficients on the variables of interest captured the association between teacher characteristics and student math achievement. Because the estimated associations were intended to apply generally to all teacher groups examined in the study, we pooled data from both the TFA and Teaching Fellows study samples to maximize the precision of the analyses.

Although student math achievement was the outcome variable, estimating relationships between teacher characteristics and student math achievement was equivalent to estimating relationships between teacher characteristics and teacher effectiveness. With controls for classroom match indicators, the regression models related differences in student achievement between classes in the same match to differences in the characteristics of their teachers. Crucially, students of matched classes were similar at the beginning of the school year due to random assignment, so achievement differences between matched classes could only have been the result of differences in teacher effectiveness. Consequently, teacher characteristics that were found to vary with student math achievement were those that were predictive of teacher effectiveness.

Characteristics examined. The characteristics in the analysis generally belonged to three broad categories: (1) measures of teachers' general academic ability; (2) measures of teachers' exposure to and knowledge of mathematics; and (3) measures of teachers' instructional training and experience. Table II.7 lists the main set of characteristics we examined and the specific measures we used for each characteristic.

Table II.7. Teacher Characteristics Examined in the Nonexperimental Analysis

Teacher Characteristic	Variables Measuring the Teacher Characteristic	Variable Structure
General Academic Ability		
College selectivity	Graduated from selective college or university[a]	Dichotomous
Exposure to and Knowledge of Math		
College-level math coursework	Number of college-level math courses taken is above median[b]	Dichotomous
Math-related nonteaching experience	Used college-level math in nonteaching job	Dichotomous
Math content knowledge	Score on Praxis II test in Math Content Knowledge (*z*-score)	Continuous
	Score on Praxis II test in Middle School Math (*z*-score)	Continuous
Instructional Training and Experience		
Math pedagogy instruction	Number of hours of math pedagogy instruction during training is above median[c]	Dichotomous
Student teaching in math	Number of days of student teaching in math during training is above median[d]	Dichotomous
Coursework during the school year	Hours of coursework during the school year (divided by 10)	Continuous
Teaching experience	Has more than one year of teaching experience	Dichotomous
	Number of additional years of teaching experience beyond two total years (until teacher has five total years of experience)	Continuous
	Number of additional years of teaching experience beyond five total years	Continuous

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

To enhance the precision of the analyses, we used a parsimonious set of variables to measure the characteristics of interest. Some characteristics, such as whether a teacher attended a selective college or university, were naturally captured by dichotomous variables. For characteristics based on survey questions whose possible responses were ordered categories—for example, categories denoting different ranges of student teaching experience—we constructed a dichotomous variable for whether a teacher's reported category was above the median reported

category in the sample.[12] Characteristics that could be expressed in well-defined increments, such as hours of coursework during the school year, were usually formulated as a single continuous variable. As an exception, the measure of teaching experience was formulated as three continuous variables such that the predicted gain in effectiveness per additional year of experience could differ at different levels of total experience (one to two years, three to five years, and more than five years). This formulation reflected previous evidence that the gains from additional experience decline with total experience (Rivkin et al. 2005; Hanushek et al. 2005).

Robustness of evidence. Our analytic approach assessed whether an observed association between a characteristic and teacher effectiveness was robust across model specifications. For each characteristic, we estimated its association with teacher effectiveness in two different ways: (1) without controls for any other teacher characteristics besides TFA/Teaching Fellow status and (2) with controls for all teacher-level variables in Table II.7. If the association between a specified characteristic and teacher effectiveness was statistically significant in both estimation approaches, the finding was regarded as representing more robust evidence than a finding of a statistically significant association in only one approach. In sensitivity analyses, we explored alternative formulations of the teacher characteristics in Table II.7 and considered models with additional characteristics (see Appendix G).

Explaining differences in effectiveness between specific groups of teachers. Findings on the relationships between teacher characteristics and effectiveness could potentially help account for differences in effectiveness between specific groups of teachers. Hence, in cases where teachers who entered teaching through either TFA or the Teaching Fellows programs differed in effectiveness from the full group of comparison teachers with whom they were compared, we assessed whether differences in teacher characteristics could account for—that is, predict—the groups' difference in effectiveness. Specifically, for each characteristic, we calculated the predicted difference in effectiveness, expressed in student *z*-score units, between TFA or Teaching Fellows teachers and comparison teachers based on only the specified characteristic. This predicted impact, which represented the portion of the actual impact accounted for by the specified characteristic, was equal to the product of two quantities: (1) the estimated association between the characteristic and teacher effectiveness and (2) the average difference in the characteristic between TFA or Teaching Fellows teachers and comparison teachers. Because the predicted impacts depended on nonexperimental estimates, they provided only suggestive—but not causally valid—explanations for the experimental findings.

[12] We constructed dichotomous variables rather than including an indicator for each category because, if we had included an indicator for each category, there would have been too few teachers in each category to detect differences between pairs of categories.

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

III. SELECTION, TRAINING, AND SUPPORT OF TFA PROGRAM PARTICIPANTS

Teachers who enter the profession through Teach For America (TFA) may differ from other teachers in their effectiveness in teaching mathematics for one or both of two broad reasons: (1) they may have different characteristics, knowledge, and experiences when they enter TFA, and (2) they may receive different preparation for teaching and support once they start teaching. This chapter describes how participants in TFA apply for the program, are selected for the program, are trained before they begin teaching ("preservice training"), receive instruction or training in local alternative certification (AC) programs, and are supported while teaching.[13] Appendix C presents some additional information on TFA.

A. Application and Selection

TFA has a national application and selection process. Individuals apply to TFA as a whole, not to a regional program, but they may specify ranked preferences for the regions in which they would like to teach (Appendix Figure C.1 lists the 43 TFA regions as of August 2011). At the time of our interviews, the application and selection process included three possible stages, depending how far an applicant advanced: (1) submission of an online application, (2) a telephone interview, and (3) an in-person assessment.

1. Application Form and Eligibility Rules

TFA estimates that the online application takes about two hours to complete. It collects information in four general areas:

1. **Personal data,** including contact information; gender and ethnicity (both optional); citizenship/residency status; disclosures regarding criminal charges and academic deficiencies or misconduct; and family background information (also optional). The family background information covers whether the applicant is a first-generation college student, his or her parents' income, and his or her parents' highest level of education completed.
2. **Academic and professional information,** including names of undergraduate and graduate schools attended, dates attended, areas of study, and degrees awarded or expected; cumulative undergraduate grade point average (GPA) and GPA by year; disclosure of failing grades and course withdrawals; leadership roles in extracurricular, paid work, or volunteer activities; longest amount of time spent pursuing a goal or refining a skill; and experience in low-income communities.
3. **Nonacademic activities during undergraduate and graduate school,** including names of organizations the applicant was involved with, whether the activity was extracurricular or for pay, organization size, title/position in the organization, primary

[13] The interview questions asked about "instruction or training." For simplicity, hereafter, we refer only to "instruction."

responsibilities, personal awards/recognitions received, number of people applicant led (if applicable), and length and timing of involvement with the organization.

4. **Postgraduate activities,** including names of organizations the applicant was involved with, title/position in the organization, primary responsibilities, most significant contribution, personal awards/recognitions received, and reasons for leaving (if applicable).

In addition to entering the requested information online, applicants must provide supplementary materials and information. All applicants must upload a resume and a "letter of intent" explaining why they want to join TFA, what they hope to accomplish in the program, and how they would determine their success in the program. Applicants who progress to the two stages beyond the application stage are asked to provide the names of two people who will submit online recommendations; one reference whom program staff could contact to discuss the applicant's candidacy; regional preferences; preferences for the subject and grade they would like to teach; information extracted from their postsecondary transcripts; official postsecondary transcripts; and proof of citizenship/residency status.

TFA also requires all applicants to meet three firm, objective criteria for eligibility: an applicant must (1) be a U.S. citizen or national or legal permanent resident; (2) have a cumulative undergraduate GPA of at least 2.5 on a 4.0 scale; and (3) receive a bachelor's degree from an accredited college/university before the start of TFA's training program, the "Summer Institute." Although the minimum GPA is 2.5, the average GPA of admitted applicants in recent years, according to national TFA staff, has been about 3.6.

Applicants who receive an offer of admission are notified about relevant hiring and certification eligibility requirements such as any coursework they will have to complete or any new-teacher examinations they will have to pass. The tests may measure basic academic skills in areas such as reading, writing, and mathematics or more advanced content knowledge in the teacher's subject area. Some states have their own tests, and others rely on various tests from *The Praxis Series*^{TM} developed and administered by the Educational Testing Service. Newly admitted corps members are encouraged to start working on these requirements as soon as possible.

2. Selection Process

The application and selection process is intended to produce a complex, "holistic view" of each applicant. It is also designed to produce multiple measures of each of seven core competencies that TFA officials view as broadly applicable to strong leaders in any field and, more important, believe to be common operating principles for highly successful corps members. The competencies are:

1. Demonstrated leadership and achievement in academic, professional, extracurricular, or volunteer arenas
2. Perseverance and sustained focus in the face of challenges
3. Critical thinking skills
4. Organizational ability
5. Respect for individuals' diverse experiences and the ability to work effectively with people from a variety of backgrounds

6. Interpersonal skills to motivate and lead others
7. An understanding of and a desire to work relentlessly in pursuit of the organization's vision

Most applicants who receive consideration beyond the application stage are asked to participate in a 25- to 45-minute telephone interview with a program representative. Interviewers follow a general set of guidelines, asking applicants about their accomplishments, how they approach problems, and issues such as educational inequity.

Applicants deemed promising based on the telephone interview are invited to an in-person assessment called the "final interview." (Applicants deemed very promising based on their application alone may bypass the telephone interview.) In 2010, final interviews were conducted in approximately 100 cities nationwide, with many cities hosting multiple sessions. Each session, which can last a full day, involves 4 to 14 applicants. At the time of our interviews, the final interview comprised four activities, in the following order:

- **Sample teaching lesson.** Candidates take turns presenting a five-minute sample teaching lesson. They "teach" the other candidates in attendance that day, with program staff observing. The lesson may target any grade level on any topic. Although all applicants know at the outset about the sample lesson requirement, they receive no detailed instructions for how to approach it until they are invited to the final interview, about 8 to 15 days before the event. At that time, they receive guidance on how to set up and introduce the lesson, and some tips on what to consider or include to help make their lesson successful.
- **Group discussion.** Applicants engage in a 15-minute group discussion among themselves. In advance of the final interview, applicants receive several articles to read on various education topics, such as the gap between minority and white student achievement or the nature of educational assessment, taken from a publication such as *Education Week* or *Educational Leadership*. On site, staff members introduce a subject or a hypothetical situation that teachers might encounter and that relates to the assigned reading. Staff members then simply observe the applicants as they discuss the issue.
- **Problem-solving activity.** Applicants are provided with two short articles similar to those described above; watch a video clip; and provide brief written answers to a series of questions about the issues addressed in the materials. They are also presented with a series of multiple-choice questions, both qualitative and quantitative. The activity, which takes place in an online format, lasts about one hour. (In 2011, as this program description was being prepared, the problem-solving activity was moved to before the telephone or final interview to help officials learn more about candidates earlier in the selection process.)
- **One-on-one interview.** Candidates participate in a one-on-one interview, lasting 25 to 45 minutes. The approach is generally similar to that used in the telephone interviews. Trained interviewers systematically collect certain kinds of information. The conversation may address information on the application, additional information obtained since submission of the application, or events that occurred earlier in the day.

At each stage of the process, a mathematical selection model heavily influences decisions about whether applicants will progress to the next stage. TFA developed the model using data about the characteristics of corps members deemed most effective in their classrooms, measured by the average annual achievement growth of their students. The model produces a score for each applicant that predicts the likelihood that he or she will be a highly effective corps member.

Virtually all the information available from applications, interviews, and related documents—with the exception of personal demographic data such as age, gender, and race/ethnicity—is entered into the model.[14] Qualitative information such as written answers to application questions and answers to interview questions are judged by selectors and quantified on various rubrics using 3-, 4-, or 5-point scales. The model includes more than 20 variables, or indicators, that collectively encompass the seven competencies listed above. Whenever possible, several measurements capture each of the seven competencies of interest. TFA officials do not publicly reveal details about the model, such as the exact number of indicators, what they measure, or how the weights are used in constructing an overall score. But officials are generally satisfied with the model's predictive abilities and its ability to select candidates who are very strong on several competencies. Moreover, as TFA collects more data on corps members' effectiveness, it reassesses and refines the model.

TFA runs the model three times to gauge which applicants should move to the next stage, each time adding more data. The first run uses data from only the application, resume, and letter of intent. Each applicant's score leads to an initial recommendation, based on different cutoff points, to move the applicant straight to the final interview, offer him or her a telephone interview, or terminate his or her consideration for admission. Staff members who act as selectors during a given admission cycle consider the model's recommendation in light of the evidence they reviewed regarding the applicant and may propose either supporting or changing the recommendation. A selection committee then considers both the model's recommendation and the selectors' recommendation and makes the final determination regarding each applicant. The model's second run adds data collected through the telephone interview or submitted after the original application. The numeric recommendation and selectors' opinions are again used to determine which applicants will be invited to the final interview. Finally, after all the data from the final interview are added to the model, TFA runs the model a third time and again uses both the numeric recommendation and selectors' opinions to determine which applicants will be offered admission to the program. TFA revises the model and the admissions cutoff point over time as it learns more about what competencies seem to matter most for teacher effectiveness in raising student test scores. But once it adopts a new model and sets a new cutoff point for a given admissions cycle, TFA does not lower the cutoff point to admit more candidates than would otherwise be selected—not even, for example, if the existing cutoff point would yield fewer secondary math teachers than desired.

Admission offers include a regional placement and teaching subject/grade assignment. In some years, some applicants may not gain admission immediately but are instead placed on a waiting list. For example, in 2010, TFA had a substantial increase in its applicant pool, but

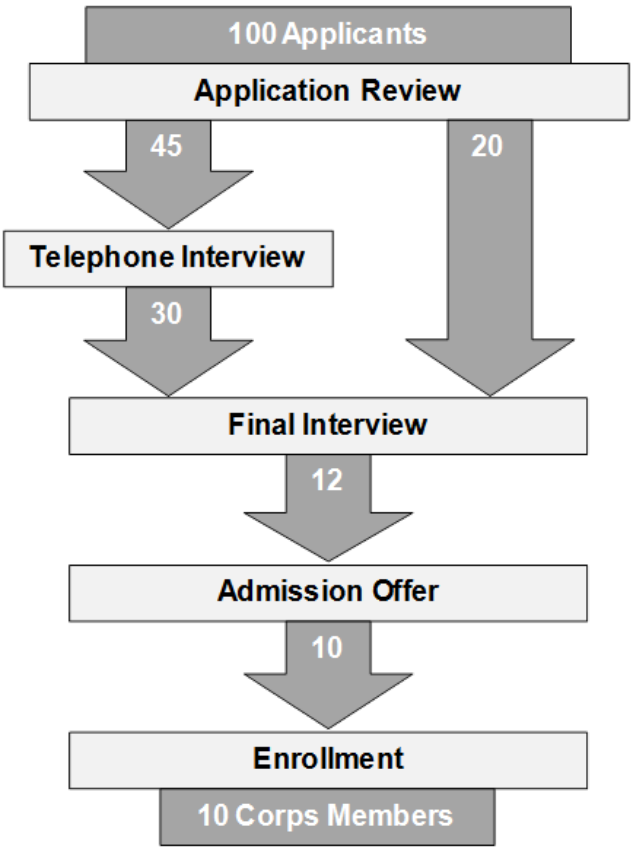
[14] TFA officials said that these factors have no bearing on admission decisions. TFA has internally important goals concerning the diversity of its corps members on these and other characteristics, but officials work toward those goals through targeted efforts to recruit a diverse applicant pool rather than adjusting admissions processes.

because of economic conditions and budget issues, some districts could not be sure how many openings they would have until later than usual.

3. Program Selectivity and Application Yields

For an overall perspective on program selectivity, we show in Figure III.1 the movement of applicants through the selection process in the three years preceding our interviews. According to national TFA officials, for every 100 applicants, on average, 65 advanced past the application stage, 50 were invited to the final interview (about 85 percent of those attended), 12 were offered admission, and 9.6 enrolled. So although 12 percent of applicants had an opportunity to join TFA, 80 percent of those offered the chance became corps members.

Figure III.1. TFA Selection Process and Applicant Flow



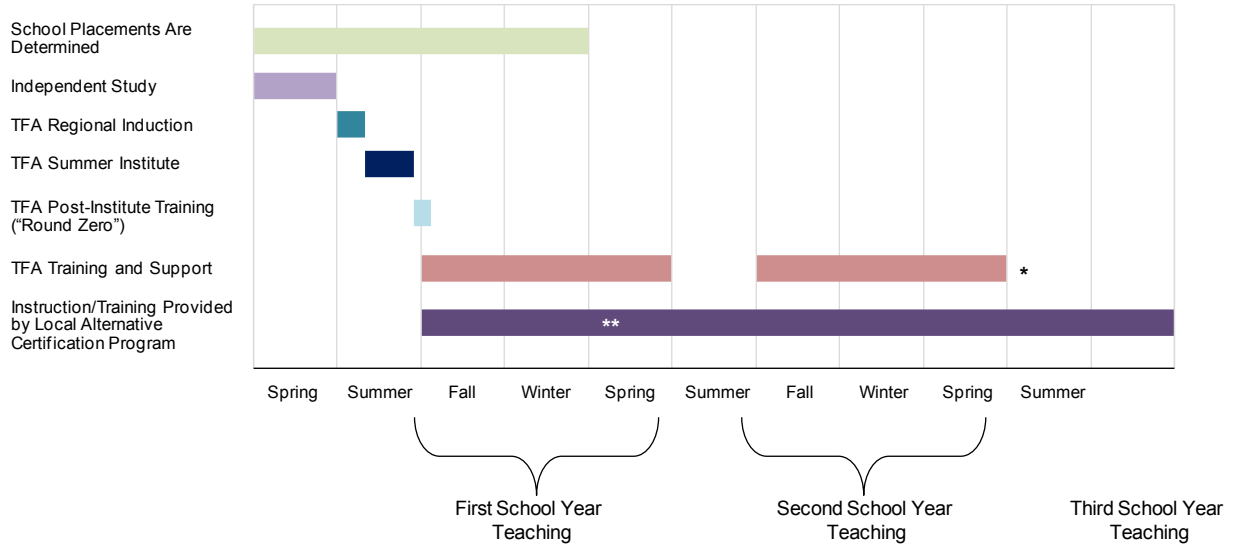
Source: National TFA officials.

B. Preservice Training for New Corps Members

The main preservice training for new TFA teachers is a full-time, residential summer institute located on university campuses around the country; it is preceded by two shorter training sessions and followed by another short training experience. All TFA training, both before and after corps members begin teaching, is guided by its Teaching As Leadership rubric, which lists six major principles of leadership: (1) set big goals, (2) invest students and their families/influencers in working hard to reach the big goals, (3) plan purposefully, (4) execute effectively, (5) continuously increase effectiveness, and (6) work relentlessly. The rubric also

lists 28 associated actions that teachers can take to promote student achievement (a list of the actions associated with each principle appears in Appendix Figure C.2). In Figure III.2, we outline the timing of placement and of all training and support activities before and after corps members begin teaching.

Figure III.2. Overview of Placement and Training of TFA Corps Members



Source: National TFA officials.

Note: The shaded areas of this figure indicate the seasons in which various activities generally take place; shading does not imply that any given activity takes up the full season. This chapter provides information on the hours associated with various activities.

*When their two-year commitment is complete, corps members become TFA alumni.

**Timing of coursework, where required, varies; it may start just before corps members begin teaching and may last into their third year of teaching; when coursework extends beyond the first year of teaching, it may or may not take place during summer.

1. Independent Study and Regional Induction

Soon after joining TFA, all new corps members must complete a few self-directed assignments to help them begin to prepare for their future teaching responsibilities and to learn more about TFA's overall approach and expectations. They read a set of six or seven curriculum texts that will be used at the summer institute, observe experienced teachers and watch video clips of classroom instruction, and reflect in writing on these materials and experiences. Altogether, TFA estimates that the independent study activities take 30 to 40 hours.

Early in the summer, new corps members attend an induction program in the region where they will teach in the fall. They participate in a series of meetings to learn more about TFA's mission and their roles as corps members and to gain familiarity with district- and school-specific curricula, policies, and programs. On average, regional induction may last two to four days.

2. Summer Institute

As of 2011, TFA held summer institutes in eight locations—Atlanta, Chicago, Houston, Los Angeles, the Mississippi Delta, New York, Philadelphia, and Phoenix—each hosting corps members from four to seven regional programs. Regardless of location, each institute follows the same model—it lasts five weeks, covers the same instruction guide, and involves the same activities. However, to improve the quality of instruction each year, a subset of institutes implements a few pilot programs. If the pilots prove successful, they are incorporated into the overall institute model. The five main types of activities during the summer institute include:

1. **Instructor-led coursework ("curriculum sessions").** Corps members spend part of virtually every weekday, most often in the evening, in small- or large-group curriculum sessions. Topics include lesson planning, content delivery, classroom management, student assessment, literacy, and effective interactions with diverse constituents. Some sessions are relevant for all corps members; other sessions are tailored for those who will teach certain subjects at certain grade levels, such as secondary math. TFA estimates the total amount of instruction at 109 hours.
2. **Practice teaching ("experiential learning").** Groups of three or four corps members team up in "collaboratives" to lead small-group or whole-class instruction in a real classroom in a local district's summer school program, under the direct supervision of the regular classroom teacher. Each participant, together with his or her teammates, spends about 20 hours tutoring small groups of students and about 20 hours leading instruction for the whole class. Participants may also spend an additional 4 to 8 hours observing each other, observing veteran teachers, or tutoring individual students.
3. **Independent work and reflection.** Corps members plan and rehearse lessons, watch video recordings of their own practice teaching, and review student work. These activities are estimated to take a total of about 71 hours.
4. **Discussions with advisors.** TFA instructors observe corps members in practice teaching several times each week and provide them with written and oral feedback to help them refine their skills. Formal one-on-one discussions total about 4 hours.
5. **Administrative activities.** Events such as opening and closing ceremonies and a campus tour take up about 19 hours.

Altogether, these activities add up to about 243 hours of effort for corps members during the summer institute, an average of about 49 hours per week.

At the time of our interviews, TFA did not widely offer any services to help candidates become qualified to teach secondary math. However, in partnership with a university, TFA had developed an online training course to help qualified secondary math teacher candidates brush up on their math knowledge and skills. The course was piloted in 2009 with a small number of corps members at one summer institute. TFA's perception that the pilot improved participants' preparedness and teaching led it to offer the online course to substantially more corps members who would be attending that same summer institute in 2011.

From 2008 to 2010, according to national TFA staff, an average of 98.5 percent of corps members who started the summer institute completed it and thus moved to post-institute activities.

3. Post-Institute Activities

After the summer institute, new corps members return to their placement regions for a final set of meetings and training activities, a period referred to as "Round Zero." While the summer institute focuses mainly on daily, general teaching skills, Round Zero focuses on longer term planning and strategies relevant to the subject, grade level, school, and district of corps members' teaching assignments. Corps members study these topics in part through required online coursework that takes about 25 hours to complete. Even though these activities may extend a month or two into the school year, we discuss them in this section on preservice training because they are foundational in nature and TFA does not classify them as part of the ongoing support provided to corps members after they start teaching (described below).

C. Placement and Retention of TFA Corps Members

Corps member placement is a two-stage process. First, at the point of admission, national program staff assign each corps member to a regional TFA program. Regional assignment decisions take into account the corps member's preferences, the number and types of teachers sought by each region, and whether the corps member meets a given region's coursework requirements for hiring or certification, such as any special requirements for secondary math and science teachers.

Second, new teachers are matched with appropriate vacancies in designated high-need schools in the regions to which they are assigned. The process is conducted within each region with the aid of a matching tool that combines data on the corps members and the vacancies in the region. In the 10 regional TFA programs in the study, the process varied in terms of how much effort new participants must exert in the job search and how much choice they have in where they interview or teach. Staff from all 10 regional programs reported that corps members interview with school principals. Interviews could start as early as mid-February, but more typically start in April or May, and most take place around the time of regional induction and the summer institute.

Corps members in different regions may receive different types of assistance during the matching process. In one region in the study, for example, new corps members receive a list of schools, particularly ones with a high percentage of students from low-income families, and are strongly encouraged to interview for openings at those schools. In other regions in the study, corps members are directed to one or more schools for interviews based on the program staff's views about which corps members would be a good fit for particular schools, requests from principals who have reviewed applicant materials and have requested to meet with certain candidates, or program and/or district staff's predeterminations about where corps members will be placed. In such predetermined cases, interviews are generally a courtesy rather than a means for reaching a hiring decision. Even when corps members interview in more than one school and placements are not decided by others in advance, corps members do not necessarily have free choice in deciding where they will teach. In one region in the study, for example, principals and district officials decide which corps members will teach at which school after interviews are completed.

In the three hiring cycles before our interviews, nearly all new teachers (98 percent, on average) and all new secondary math teachers in the 10 regional programs were hired before the

start of the school year, and all teachers not hired before the start of the school year were hired after the year started (Table III.1).

Table III.1. Estimated Percentage of New TFA Corps Members Who Received Teaching Placements, on Average, Across the 2007–2008, 2008–2009, and 2009–2010 School Years

	Hired Before Start of School Year	Hired After Start of School Year	**Total**
All Candidates	98%	2%	**100%**
Secondary Math Candidates	100%	0%	**100%**

Source: Interviews with regional TFA officials.

Regional staff typically try to place new corps members in schools where current or former TFA corps members are teaching, so that the new teachers can benefit from an added source of support. The programs in the study had typically succeeded in arranging such placements over the three school years preceding our interviews. On average, across fall 2007, 2008, and 2009, according to program representatives, an estimated 92 percent of all new teachers and 93 percent of new secondary math teachers were hired in schools where current or former corps members were teaching.

For the corps members nationwide who entered TFA as secondary math teachers in 2007, 2008, or 2009 (around the time of some sample members), TFA officials estimate that, on average, 94.3 percent completed their first year in the program, 91.2 percent started their second year, and 90.3 percent completed their two-year commitment in the program.

D. Instruction Provided to TFA Corps Members Through Local Teacher Certification Programs

In 9 of the study's 10 regional TFA programs, new secondary math teachers were required to enroll in a state-authorized alternative certification (AC) program. The 10th program had no such requirement, because state law allowed teachers to gain initial certification, good for three years, without enrolling in a certification program, as long as they had a relevant college major for the subject they would teach and passed both a general knowledge test and a test in their teaching subject area. All secondary math corps members in recent years reportedly had taken and passed the tests and did not enroll immediately in an AC program.

To gain a general sense of the instruction that TFA secondary math teachers in our study might have received from local AC programs, because teacher-level data collection was not feasible, we asked officials in each region to identify the one program that had enrolled the largest number, whether a majority or a plurality, of their secondary math teachers over the three most recent years before 2010. This was the time period, we believed, during which most study teachers would have enrolled in a local AC program. (The later addition of study teachers from TFA's summer 2010 entering cohort means that their AC program experiences may not be as well represented by the AC program information we collected.) In seven of the regions we examined, the identified AC programs reportedly served all the TFA secondary math teachers. In both of the other two regions, the identified program reportedly served 70 percent of secondary math TFA teachers in that region. Hereafter, we refer to these AC programs as "focal," as they were the focus of our questions.

The nine focal AC programs were operated by a variety of institutions; five were operated by a nearby university. Secondary math teachers from two regions attended a single AC program operated by TNTP. One program was operated by a local school district in collaboration with a university. In the final case, the regional TFA program was the state-authorized AC provider. Additional TFA regional programs also now serve as state-authorized certification providers.

1. Preservice Instruction

Two of the nine focal AC programs reportedly involved a preservice component in addition to the TFA's preservice and Round Zero activities. One involved one week of full-time, in-person instruction for a total 40 hours; the other involved 25 hours of independent study. Across all 10 regional TFA programs, the average preservice instruction requirement faced by new secondary math teachers was 6.5 hours.

2. Instruction During and After the First Year of Teaching

All nine focal AC programs involved instruction during the teachers' first year of teaching. The certification program with the most extensive requirement for participants' first year involved 112 hours of instruction; participants attended classes or other training activities 4 hours a week for 28 weeks. The certification program with the least extensive first-year requirement involved 54 hours of instruction; participants attended classes or other training activities 3 hours a week for 18 weeks. Across all 10 regional TFA programs, the average first-year instruction requirement faced by new secondary math teachers in their focal certification programs was 70 hours. Three of the programs reportedly involved instruction during the summer after corps members' first year of teaching, ranging from 40 to 60 hours (Table III.2).

Table III.2. Hours of Instruction Toward Certification Reportedly Required in Focal AC Programs for TFA Secondary Math Teachers

	Regions with This Instructional Component				
Instruction Provided	Number of Regions	Minimum	Maximum	Mean	Mean for All 10 Regions[b]
Before first year of teaching	2	25	40	32.5	6.5
During first year of teaching	9	54	112	77.8	70.0
During summer after first year of teaching	3	40	60	49.3	14.8
During second year of teaching	3	100	210	153.3	46.0
Total Hours, All Programs[b]	**10**	**0**	**315**	**137.3**	**137.3**

Source: Interviews with regional TFA officials.

Note: This table excludes training provided directly by the TFA summer institute and regional TFA programs.

[a]The focal AC program for each participating TFA regional program is the program that enrolled the largest number of its secondary math teachers from 2007 to 2009.

[b]Includes the one regional TFA program whose participants did not have to attend an AC program. For these calculations, the program received a value of 0 hours at each time period.

Finally, required instruction at three certification programs extended into new corps members' second year of teaching. The certification program with the most extensive requirement in this period involved about 7.5 hours of instruction a week for 28 weeks, for a

total of 210 hours. The certification program with the least extensive second-year requirement involved 5 hours of instruction a week for 20 weeks, for a total of 100 hours. Across all 10 TFA programs, the average second-year instruction requirement faced by new secondary math teachers in their focal certification programs was 46 hours.

3. Total Amount of Instruction in Certification Programs

The total hours of instruction that secondary math teachers in the 10 regions reportedly had to complete for certification, before and during their first two years of teaching, ranged from 0 to 315 hours, with a mean of about 137 hours. In Table III.2, we present average instruction hours required, overall and by time period.

E. Training and Support for New Corps Members After They Begin Teaching

Regional program staff continue to work with corps member after they begin teaching.

1. Instruction, Training, and Professional Development

All TFA regional programs provide some type of instruction, training, or professional development (hereafter "training") to their participants after they began teaching. In this paragraph, we exclude from our discussion the one program in the study that was a state-authorized AC provider. For this program it was difficult to distinguish between training provided as part of the state certification program and that provided as a regular part of TFA. The training in the other nine programs took place on weeknights and/or weekends during the school year. The programs provided an average of 42 hours of training, according to representatives' reports, with a range of 16 to 70 hours. Most of these programs (eight of nine) covered classroom management and lesson plans or unit plans (five of nine). Programs also sometimes covered goal setting, creating assessments or judging student work, and instructional practices or pedagogy. Subjects are not necessarily set in advance but sometimes are determined throughout the year according to participants' needs—gauged, for example, through classroom observations. Six of the nine programs reportedly provided some training specifically for secondary math teachers.

2. Other Support Services

All 10 TFA programs provided individualized support for new teachers, such as classroom observations and mentoring. Each program appointed staff members, known as "program directors," to work with participants after they started teaching. Each program director typically worked with about 30 new teachers. Where possible, new secondary math teachers were matched with program directors who were themselves former secondary math teachers. At the time of the interviews, six programs reported that 100 percent of their secondary math teachers were in such a pairing, and two programs reported that none of their secondary math teachers was in such a pairing; the mean of the percentages reported across the programs was 68 percent. Nine of the TFA programs reported that new teachers were supported by program directors for their full two years of program participation, whereas one program reported that the support lasted one year.

- **Classroom observations.** Program directors in all 10 regions reportedly conducted formal observations of new teachers in their classrooms. The average number of observations per teacher ranged from 1 to 16, with a mean of 6; six programs reported that they typically observed each new teacher four or five times over his or her two

years in the program. All program directors provided the teachers with feedback from the observations to help them with their development. In some cases, program directors made the assessment available to the principal (for example, if the teacher was struggling or the principal requested the information) or to other TFA staff (either routinely or if the teacher was struggling and the program director wanted additional advice on how to help the teacher).

- **One-on-one meetings.** Beyond providing feedback from classroom observations, program directors in 7 of the 10 regions reportedly also held other planned one-on-one meetings with new teachers over their two years in the program. The purpose of the meetings was to mentor the new teachers and provide information, advice, or encouragement. The average number of planned discussions with teachers across the programs ranged from 2 to 40, with a mean of 10.5, and the average length of discussions across programs ranged from 60 to 90 minutes, with a mean of 71 minutes.
- **Group meetings.** Program directors or other staff at 8 of the 10 regional TFA programs reportedly also scheduled meetings with groups of new teachers to provide various types of support or information. The topics covered in the sessions varied with assessments of corps members' needs. The average staff-to-teacher ratio in these meetings, across the six programs that provided an estimate, ranged from 1:5 to 1:10. The average number of meetings that TFA staff scheduled for new teachers over their two years in the program, across the five programs that provided an estimate, ranged from 2 to 20, with a mean of 10.8. The average length of the meetings, across the seven programs that provided an estimate, ranged from 90 to 150 minutes, with a mean of 99 minutes.

F. Financial Support for Corps Members

Corps members may receive financial support while in training. First, TFA covers the costs of room and board during the summer institute and regional orientation. Second, TFA offers needs-based no-interest loans and grants of $1,000 to $6,000 to cover transitional costs such as travel expenses to the summer institute and placement region, relocation expenses, and testing and certification fees. Third, TFA corps members may receive financial support from other sources. Most corps members, at the time of our program interviews (spring 2010), were eligible for an AmeriCorps education award of $5,350 at the end of each of their two years of teaching. (Participants who had received one AmeriCorps award before joining TFA were eligible for only one more award; those who had already received two awards were ineligible for additional awards.) AmeriCorps awards could be used to cover teacher training and certification costs as well as repayment of qualified student loans and future graduate school expenses. Fourth, at the time of the interviews, participants in four of the regional TFA programs could reportedly receive some form of financial assistance to help cover at least some of the current costs associated with the AC programs discussed above. The support ranged from $1,000 per participant in one region to $3,000 per participant, per year, in another region.

IV. TEACH FOR AMERICA STUDY TEACHERS: THEIR CHARACTERISTICS, THE SCHOOLS THEY TEACH IN, AND THE STUDENTS THEY TEACH

Teach For America (TFA) aims to place high quality teachers in schools that serve some of the most disadvantaged students in the United States. TFA's screening procedures suggest that teachers accepted into the program will differ from teachers who enter the profession through other routes. Its placement procedures suggest that the schools these teachers are placed in and the students they serve are also likely to differ from typical schools and students. To provide context for the estimates of the effectiveness of TFA teachers presented in Chapter V, this chapter describes the schools (Section A), teachers (Section B), and students (Section C) in the TFA study sample.

To summarize the findings from this chapter, we found that TFA teachers in the study were typically placed in urban schools that served economically disadvantaged students, largely from racial and ethnic minority groups. The non-TFA study teachers in these schools were primarily nonwhite women, typically with more than five years of teaching experience. In contrast, the TFA teachers in the study were overwhelmingly white, less likely to be female, and had significantly less teaching (and general work) experience. The TFA teachers were also far more likely to have attended a selective undergraduate university. While the TFA teachers were less likely to have majored in math, they had significantly greater math content knowledge test scores than the typical math teachers in the schools in which they taught.

A. Schools in the TFA Study Sample

Even though study schools were not randomly selected from the full set of secondary schools employing TFA teachers nationwide, the study schools were similar to secondary schools employing TFA teachers nationwide along many dimensions (Table IV.1). For instance, both sets of schools served predominantly students from racial and ethnic minority groups. Less than 10 percent of students at both the average study school and the average secondary school with TFA teachers nationwide were white, non-Hispanic, while 57 percent at both types of schools were black, non-Hispanic, and around 30 percent were Hispanic. Close to 80 percent of students at both types of schools were eligible for free or reduced-price lunch. More than 80 percent of both schools in the TFA study sample and schools with TFA teachers nationwide were eligible for Title I funds for schools with high percentages of low-income students. Consistent with TFA's mission to place its corps members in schools in low-income communities, schools in the study sample were on average more disadvantaged than the average secondary school nationwide. For instance, compared with students in the average secondary school nationwide, students in the average school in the TFA study sample were more likely to be eligible for free or reduced-price lunch and to be from racial and ethnic minorities.

There were also some differences between study schools and all TFA schools nationwide, many of which may have been due to our recruitment approach and study eligibility requirements. The average study school had significantly more students per grade than the average secondary school employing TFA teachers (240 versus 184 students per grade), consistent with the fact that schools with more students per grade were more likely to have eligible classroom matches. Although 23 percent of secondary schools with TFA placements nationwide were charter schools, there were no charter schools in the study sample. This reflects the fact that charter schools were typically smaller than average and therefore less likely to have

Table IV.1. Characteristics of Study Schools with TFA Teachers Compared with All Secondary Schools with TFA Teachers and All Secondary Schools Nationwide

	Study Schools with TFA Teachers	All Secondary Schools with TFA Teachers[a]		All Secondary Schools Nationwide	
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	2.8	3.8	0.150	3.5	0.091
Percentage Black, non-Hispanic	57.4	57.0	0.918	18.8**	0.000
Percentage Hispanic	31.3	32.1	0.852	20.2*	0.016
Percentage White, non-Hispanic	8.0	8.9	0.545	59.4**	0.000
Percentage other race/ethnicity	0.5	4.5**	0.000	9.8**	0.000
Student Socioeconomic Status					
Percentage eligible for free/reduced-price lunch	78.9	79.7	0.663	50.5**	0.000
Percentage Title I-eligible schools	82.2	88.5	0.267	60.0**	0.000
Enrollment and Staffing					
Average total enrollment	836.0	714.3	0.090	551.2**	0.000
Average enrollment per grade	240.2	183.6**	0.003	134.5**	0.000
Average student/teacher ratio	14.4	14.7	0.339	16.1**	0.000
Grade Level[b]					
Percentage middle schools	66.7	59.0	0.278	66.1	0.932
Percentage high schools	37.8	54.8*	0.019	45.3	0.301
School Type					
Percentage charter schools	0.0	22.5**	0.000	6.9**	0.000
Percentage magnet schools	26.7	6.2**	0.001	2.3**	0.000
School Location					
Percentage urban	80.0	70.1	0.098	25.5**	0.000
Percentage suburban	11.1	21.2*	0.038	37.8**	0.000
Percentage rural	8.9	8.5	0.922	36.4**	0.000
Census Bureau Region					
Percentage in Northeast	13.3	13.2	0.974	14.3	0.844
Percentage in Midwest	13.3	14.6	0.799	27.7**	0.005
Percentage in South	66.7	53.3	0.060	30.8**	0.000
Percentage in West	6.7	18.7**	0.003	26.9**	0.000
Sample Size	**45**	**471**		**63,148**	

Sources: TFA placement data; Common Core of Data, Public Elementary/Secondary School Universe Survey, 2009–2010 .

Note: Secondary schools are defined as any school that includes at least one grade above 5th grade.

[a]Estimates are based on schools in which new TFA teachers were placed in the 2009–2010 school year.

[b]Middle schools are defined as schools with at least one grade from 6 to 8. High schools are defined as those with at least one grade from 9 to 12. Thus the two categories are not mutually exclusive, as a given school could contain both middle and high school grades.

*Difference between this group and study schools with TFA teachers (first column) is statistically significant at the 0.05 level, two-tailed test.

**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.

eligible classroom matches. Study schools were also significantly more likely than the typical secondary school with TFA placements to be magnet schools (27 versus 6 percent) and significantly less likely to be high schools, defined as those serving at least one grade from 9 to 12 (39 versus 55 percent). Eighty percent of study schools were located in urban areas, as were 70 percent of the average secondary schools employing TFA teachers; this difference was not statistically significant at the 5 percent level. The average study school was less likely to be located in a suburban area (11 versus 21 percent). Study schools were also significantly less likely than the full set of TFA secondary schools to be located in the West (7 versus 19 percent).

B. TFA Teachers in the Study Sample Compared with Comparison Teachers

TFA selects teachers with characteristics and attitudes it believes are associated with effectiveness in the classroom, provides them with five weeks of training before they begin teaching, and provides additional training and support throughout the teachers' two-year commitment to teaching. Hence, TFA study teachers may differ from comparison teachers in the same classroom matches—the non-TFA teachers teaching the same math course in the same school—in their background characteristics, experience, preparation for teaching, training, support, or attitudes toward teaching. In fact, TFA teachers did differ from comparison teachers in many ways, as described further below.

Compared with comparison teachers, TFA teachers:

- Were younger and less likely to be members of racial or ethnic minorities
- Were more likely to have graduated from a selective college or university
- Had greater math content knowledge as measured by scores on the Praxis II math exams
- Were less likely to have majored in math and had taken fewer college-level math courses
- Had fewer years of teaching and nonteaching experience
- Completed similar amounts of math pedagogy instruction and student teaching in math during their preparation for teaching
- Were more likely to have taken coursework during the study school year and to have had a mentor during that year, which is consistent with the fact that TFA teachers were more likely to be in their first or second year of teaching
- Were less satisfied with some aspects of teaching related to school policy and staff.

The comparison teachers with whom TFA study teachers were compared included both teachers from traditional routes to certification (TC routes) and those from less selective alternative routes to certification (AC routes)—59 percent of comparison teachers were from TC routes, while 41 percent were from less selective AC routes. With a few exceptions discussed below, these patterns of differences were present both among the subset of TFA teachers who were compared with teachers from TC routes and the subset who were compared with teachers from AC routes (Appendix D).

1. Demographic Characteristics

TFA teachers differed from comparison teachers in age, gender, and race/ethnicity (Table IV.2).[15] Consistent with the fact that TFA primarily recruits recent college graduates and asks them to make only a two-year commitment to teaching (although some remain in teaching beyond this commitment), TFA study teachers were on average younger than comparison teachers (25 versus 38 years old). They were less likely to be female (61 versus 79 percent). TFA teachers were also less likely than comparison teachers to be members of racial or ethnic minorities. Eighty-nine percent of TFA study teachers were white, non-Hispanic, compared with 30 percent of comparison teachers.[16] Eight percent of TFA study teachers were black, non-Hispanic, compared with 57 percent of comparison teachers. Comparison teachers in the study were closer in age to the average secondary teacher nationwide than were TFA study teachers, while TFA study teachers looked more like the average secondary teacher in terms of both gender and racial and ethnic distribution.

2. Educational Background and Math Content Knowledge

TFA study teachers were more likely than comparison teachers to have graduated from a selective college or university (81 versus 23 percent) (Table IV.3).[17] However, they were also less likely to have completed a graduate degree (41 versus 70 percent). They were less likely than comparison teachers to have majored in math (8 versus 26 percent) or secondary math education (0 versus 16 percent), but more likely to have majored in some other math-related subject (statistics, engineering, computer science, finance, economics, physics, or astrophysics) (27 versus 12 percent). There was no significant difference in the percentage of TFA and comparison teachers who majored in any math-related subject (including math, secondary math education, and the other math-related subjects listed above).

Patterns of differences in educational background were similar for the subset of TFA teachers who were compared with comparison teachers from TC routes and the subset who were compared with comparison teachers from AC routes, with one exception. Teachers from TC routes were significantly more likely to have majored or minored in education than the TFA teachers with whom they were compared, while none of the teachers from AC routes or the TFA teachers with whom they were compared had an education major or minor (Appendix Table D.2).

[15] In this and all teacher-level analyses presented in this chapter, each teacher is counted only once, regardless of whether he or she taught multiple classes in the study. For teachers in both cohorts, the value of the variable as reported in cohort 1 is used if cohort 1 and cohort 2 reports differ. We also conducted analyses that weight each teacher according to the number of study classes taught, and findings were similar.

[16] For comparison, among all TFA secondary math teachers who began teaching in fall 2010, 69 percent were white, non-Hispanic, according to data provided by TFA (not shown in table). We are not able to test the significance of the difference between this estimate and that for the study teachers.

[17] College selectivity was defined based on *Barron's Profiles of American Colleges 2003*. Highly selective colleges were defined as those ranked by Barron's as most competitive, and selective colleges were defined as those ranked by Barron's as very competitive, highly competitive, or most competitive.

Table IV.2. Demographic Characteristics of TFA and Comparison Teachers in the Study and All Secondary Teachers Nationwide (percentages unless otherwise indicated)

	Secondary Teachers Nationwide	TFA Teachers	Comparison Teachers	Difference Between TFA and Comparison Teachers	*p*-Value
Age (Average Years)	42.7††^^	24.5	37.9	-13.4**	0.000
Female	59.3^^	60.9	79.4	-18.4*	0.023
Race/Ethnicity[a]					
Asian, Non-Hispanic	1.3†^	9.4	11.1	-1.7	0.749
Black, Non-Hispanic	6.9^^	7.8	57.1	-49.3**	0.000
Hispanic	6.8	4.7	12.7	-8.0	0.111
White, Non-Hispanic	83.5^^	89.1	30.2	58.9**	0.000
Sample Size	**1,099,770**	**64**	**63**		

Source: Data for secondary teachers nationwide from the Schools and Staffing Survey Teacher Questionnaire, 2007–2008; data for study teachers from the Survey of Secondary Math Teachers.

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

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

†Difference between secondary teachers nationwide and TFA teachers is statistically significant at the 0.05 level, two-tailed test.
††Difference between secondary teachers nationwide and TFA teachers is statistically significant at the 0.01 level, two-tailed test.

^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
^^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Consistent with the fact that they were less likely to have majored in math, TFA teachers had also completed fewer college-level math courses than comparison teachers (5 versus 8 courses) (Table IV.4).[18] TFA teachers had taken significantly fewer courses than comparison teachers in several topic areas, including advanced algebra, analysis, advanced geometry/topology, probability and statistics, and discreet mathematics; however, there was no significant difference in the number of calculus courses taken. TFA teachers took fewer college-level math courses on average than comparison teachers at both the middle and high school levels, although this difference was more pronounced at the high school level.

[18] Teachers were asked to indicate the number of college or graduate-level math courses they had taken at a college or university, or high school courses for which they had received college credit, across various categories of courses. They were not expected to include high school courses for which they did not receive college credit.

Table IV.3. 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 selective college or university[a]	81.3	22.7	58.5**	0.000
Major[b]				
Math	7.8	25.6	-17.8*	0.020
Secondary math education	0.0	16.3	-16.3**	0.005
Other math-related subject[c]	26.6	11.6	14.9*	0.047
Other subject	90.6	67.4	23.2**	0.005
Math, secondary math education, or other math-related subject[c]	34.4	48.8	-14.5	0.141
Major or Minor[b]				
Math	12.5	30.2	-17.7*	0.033
Secondary math education	0.0	20.9	-20.9**	0.001
Other math-related subject[c]	28.1	14.0	14.2	0.072
Other subject	93.8	72.1	21.7**	0.005
Graduate Degree				
Any graduate degree	40.6	69.8	-29.2**	0.001
Subject[b]				
Math	0.0	4.8	-4.8	0.081
Secondary math education	18.8	15.9	2.9	0.671
Other subject	21.9	54.0	-32.1**	0.000
Sample Size	**64**	**63**		

Source: Survey of Secondary Math Teachers.

[a]Selective colleges are those ranked by *Barron's Profiles of American Colleges 2003* as being very competitive, highly competitive, or most competitive.

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

[c]Includes statistics, engineering, computer science, finance, economics, and physics/astrophysics.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Despite the fact that the TFA teachers were less likely to have majored in math or a related subject than comparison teachers, the TFA teachers displayed greater math content knowledge, as measured by the Praxis II Mathematics Content Knowledge Test (taken primarily by the high school teachers in the sample, along with a few middle school teachers in states that allowed or required middle school teachers to take this test) or the Praxis II Middle School Mathematics Test (taken by the remaining middle school teachers in the sample) (Table IV.5).[19] TFA teachers who took the Mathematics Content Knowledge Test outperformed comparison teachers by 22 points (or 0.93 standard deviations); those who took the Middle School Mathematics Test also outperformed comparison teachers by 22 points (or 1.19 standard deviations). Similar differences persisted among both teacher matches for whom the Praxis was "high stakes" (those in districts that required the exam for certification) and those for whom the Praxis was "low stakes" (not required for certification, and taken only at the request of the study team).

[19] Response rates on the Praxis II differed for TFA and comparison teachers—91 percent of the TFA teachers in the sample had Praxis II scores, compared with only 73 percent of comparison teachers (Appendix Table A.18). In the main analysis, we omitted teachers with missing scores. As a sensitivity test, in Appendix Table E.1 we examined differences in Praxis II scores when we imputed values that would be otherwise missing. Results were similar under the two approaches, although differences between the two groups were somewhat more pronounced, suggesting that teachers with missing Praxis data might have been those who would have scored lower had they taken the test.

Table IV.4. Average Number of College-Level Math Courses Taken by TFA and Comparison Teachers in the Study

	TFA Teachers	Comparison Teachers	Difference	*p*-Value
All Teachers				
Total Number of College-Level Math Courses	4.9	8.3	-3.4**	0.001
Calculus	1.7	2.1	-0.4	0.179
Advanced Algebra	0.5	1.3	-0.8**	0.000
Analysis	0.3	0.8	-0.5**	0.008
Advanced Geometry/Topology	0.1	0.6	-0.5**	0.000
Probability and Statistics	1.2	1.5	-0.3	0.146
Discrete Mathematics	0.2	1.0	-0.8**	0.000
Applied Mathematics	0.8	1.0	-0.1	0.669
Sample Size	**64**	**63**		
Middle School Teachers				
Total Number of College-Level Math Courses	4.2	6.7	-2.5**	0.006
Sample Size	**47**	**50**		
High School Teachers				
Total Number of College-Level Math Courses	6.6	14.4	-7.7*	0.011
Sample Size	**17**	**13**		

Source: Survey of Secondary Math Teachers.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Although it would be useful to compare average scores of study teachers to average scores of the full set of teachers who took the Praxis II mathematics tests nationwide, published data on national average scores are available only for test takers who passed the test. For comparison to these statistics, we also examined the average scores of the teachers in our sample who passed the test. On the Mathematics Content Knowledge Test, TFA teachers in the study who passed the test scored an average of 165 points, compared with an average of 155 for all test-takers nationwide who passed. Similarly, on the Middle School Mathematics Test, TFA teachers in the study who passed the test scored an average of 182 points, compared with an average of 169 for all test-takers nationwide who passed.

3. Teaching and Nonteaching Work Experience

Previous research has shown that teacher effectiveness generally increases with years of teaching experience (Hanushek and Rivkin 2006; Kane et al. 2008). Because TFA only asks its teachers to make a two-year commitment to teaching (although they may choose to remain longer), critics of the program have claimed that TFA teachers are less effective than other teachers because they lack experience (Heilig and Jez 2010). TFA teachers in the study did have less teaching experience than comparison teachers (Table IV.6). The TFA teachers had been teaching an average of two years as of the end of the study school year, compared with an average of 10 years among comparison teachers. Eighty-three percent of the TFA teachers were in their first or second year of teaching, compared with 10 percent of comparison teachers. Seventy percent of the comparison teachers had been teaching more than five years, while none of the TFA teachers had been teaching this long.

Table IV.5. Praxis II Scores of TFA and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Test-Takers Nationwide	TFA Teachers	Comparison Teachers	TFA-Comparison Difference	*p*-Value
Overall					
Mathematics: Content Knowledge Test					
Average score	n.a.	162.0	140.1	21.9*	0.010
Average score of those who scored above the state's passing threshold[a]	154.5†	164.6	152.7	11.9	0.074
Sample size	5,785	15	11		
Middle School Mathematics Test					
Average score	n.a.	179.8	158.3	21.6**	0.000
Average score of those who scored above the state's passing threshold[a]	169.0††	181.7	166.8	14.9**	0.000
Sample size	9,532	45	40		
High-Stakes Test-Takers[b]					
Mathematics: Content Knowledge Test					
Average score	n.a.	159.9	140.1	19.8*	0.018
Sample size		14	11		
Middle School Mathematics Test					
Average score	n.a.	179.3	163.0	16.3**	0.002
Sample size		20	16		
Low-Stakes Test-Takers[c]					
Middle School Mathematics Test					
Average score	n.a.	180.3	155.1	25.2**	0.000
Sample size		25	24		

Source: Data on all test-takers nationwide are from Gitomer and Qi (2010), and include test-takers who did and did not eventually enter the teaching profession. Data for study teachers are from study-administered assessment or were provided by the Educational Testing Service.

[a]The Educational Testing Service does not report the average score of all test-takers, but only the average score of those who passed the test in their states. To compare the study sample with the national statistics from the Educational Testing Service, we computed the average scores of sample members who scored above their states' passing thresholds using the Praxis passing thresholds from 2011. In states in which this test was not required, the score was compared with the average passing threshold across all states that did require this test.

[b]High-stakes test-takers are those who are in districts that require secondary math teachers to pass the relevant Praxis II assessment for certification.

[c]Low-stakes test-takers are those who are in districts that do not require secondary math teachers to pass the relevant Praxis II assessment for certification. Because fewer than three low-stakes test-takers in the TFA study sample took the Mathematics: Content Knowledge test, we do not present estimates for this group, to protect respondent confidentiality.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

†Difference between secondary teachers nationwide and TFA teachers is statistically significant at the 0.05 level, two-tailed test.

††Difference between secondary teachers nationwide and TFA teachers is statistically significant at the 0.01 level, two-tailed test.

^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.05 level, two-tailed test.

^^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

n.a. = not available.

TFA teachers in the sample also had less nonteaching work experience than the comparison teachers (Table IV.6). Seventeen percent of the TFA teachers had held a nonteaching job after college, compared with 54 percent of comparison teachers. Nine percent of TFA teachers had used college-level math in a nonteaching job, compared with 37 percent of comparison teachers.

Table IV.6. Nonteaching and Teaching Work Experience of TFA and Comparison Teachers in the Study (percentages unless otherwise indicated)

	TFA Teachers	Comparison Teachers	Difference	*p*-Value
Nonteaching Work Experience				
Had a nonteaching job after college	17.2	54.0	-36.8**	0.000
Years of experience in nonteaching job (average)[a]	0.4	2.7	-2.3**	0.000
Used college-level math in nonteaching job[a]	9.4	36.5	-27.1**	0.000
Teaching Experience (end of study year)				
Years of teaching experience (average)	1.9	10.1	-8.3**	0.000
1 or 2 years of teaching experience	82.8	9.5	73.3**	0.000
3 to 5 years of teaching experience	17.2	20.6	-3.4	0.623
More than 5 years of teaching experience	0.0	69.8	-69.8**	0.000
Years of experience teaching math (average)	1.8	7.8	-6.0**	0.000
1 or 2 years of math teaching experience	85.9	9.5	76.4**	0.000
3 to 5 years of math teaching experience	14.1	28.6	-14.5*	0.047
More than 5 years of math teaching experience	0.0	61.9	-61.9**	0.000
Sample Size	**64**	**63**		

Source: Survey of Secondary Math Teachers.

[a]Calculations are based on all teachers, regardless of whether they had a nonteaching job after college.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

4. Teacher Training, Professional Development, and Support

Teacher training and professional development. TFA and comparison teachers reported completing similar amounts of math pedagogy instruction and student teaching in math as part of their teacher training programs (Table IV.7).[20] Differences in average hours of math pedagogy instruction and average days of student teaching in math between TFA and comparison teachers were not statistically significant. However, the average days of student teaching in math mask some differences between the two groups—the TFA teachers were more likely than comparison teachers to have completed at least some student teaching in math and were more likely to have completed from 11 to 20 days. The TFA teachers also reported receiving less professional development in math during the study school year than comparison teachers (6 versus 11 hours).

Teachers from TC routes typically receive different teacher training than teachers from AC routes (Constantine et al. 2009), and so it is not surprising that there were some differences in teacher training and support in the subset of TFA teachers who were compared with teachers from TC routes relative to the subset who were compared with teachers from AC routes (Appendix Table D.6). For instance, TFA teachers completed significantly less student teaching

[20] Because questions about teacher training required more experienced teachers (predominantly comparison teachers) to recall experiences that occurred several years in the past, these data may be less reliable than data pertaining to the study year.

Table IV.7. Training and Professional Development of TFA and Comparison Teachers in the Study (percentages unless otherwise indicated)

	TFA Teachers	Comparison Teachers	Difference	*p*-Value
Average Hours of Math Pedagogy Instruction as Part of Teacher Training[a]	35.2	37.1	-1.9	0.755
Days of Student Teaching in Math as Part of Teacher Training[b]				
No days	21.9	47.6	-25.7**	0.002
1 to 10	6.3	4.8	1.5	0.716
11 to 20	42.2	15.9	26.3**	0.001
More than 20	29.7	31.7	-2.1	0.803
Average Days of Student Teaching in Math	18.1	24.7	-6.6	0.141
Hours per Day Spent in Student Teaching in Math as Part of Teacher Training (average)[c]	1.3	3.3	-2.0**	0.000
Average Hours Spent in Math Pedagogy Professional Development During School Year[d]	6.1	11.0	-4.9**	0.000
Sample Size	**64**	**63**		

Source: Survey of Secondary Math Teachers.

[a]Teachers were asked, "As part of your training to become a math teacher, did you receive any instruction in math pedagogy or strategies to teach math?" If so, "In total, how many hours of instruction in math pedagogy or strategies to teach math 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 math 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 in which you went to an elementary or secondary school and taught one or more math lessons to a whole classroom of students?" If so, "On approximately how many days, in total, did you teach at least one full math lesson (at least one class period) to a whole classroom of students during your teacher education/preparation program?" Possible responses were none, 1 to 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 in math, we created a continuous variable equal to zero for teachers who did not do any student teaching in math, 80 for those who did more than 80 days, and the midpoint of the range for all other categories.

[c]Calculated only among those teachers who said they did some student teaching in math.

[d]Teachers were asked, "During this school year, did you attend any professional development classes, workshops, or seminars *provided by the school district* in math pedagogy or strategies to teach math?" If so, "In total, how many hours did you spend attending these professional development classes, workshops, or seminars in math pedagogy or strategies to teach math?" 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 between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

in math, on average, than comparison teachers from TC routes (20 versus 38 days), while the difference between TFA teachers and comparison teachers from AC routes (17 versus 8 days) was not statistically significant at the 5 percent level.

Coursework during study year. Most states require teachers from AC routes, including TFA, to take coursework in their first few years of teaching to fulfill state certification requirements (Feistritzer 2005). Consistent with the fact that they were more likely to be in their first or second year of teaching and thus likely still fulfilling coursework requirements for certification, TFA teachers were more likely than comparison teachers to have taken coursework during the study year (Table IV.8). Fifty percent of TFA teachers reported taking coursework related to their teaching job during the study year, compared with 21 percent of comparison

Table IV.8. 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
All Teachers				
Took Coursework Related to Teaching Job	50.0	20.6	29.4**	0.000
Total Hours Spent During School Year on Coursework (average)	89.4	49.9	39.5	0.095
Hours Spent in Class During School Year on Coursework (average)	41.8	22.8	19.1	0.089
Hours Spent Out of Class During School Year on Coursework (average)	47.5	27.1	20.4	0.137
Sample Size	**64**	**63**		

Source: Survey of Secondary Math Teachers.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

n.a = not applicable.

teachers. TFA teachers spent an average of 89 hours on coursework during the study year (roughly evenly divided between time in and out of class), compared with 50 hours among comparison teachers; this difference was not statistically significant at the 5 percent level.

Mentoring. Under federal guidelines, novice teachers from AC programs must participate in "a program of intensive supervision that consists of structured guidance and regular ongoing support for teachers or in a teacher mentoring program" to be considered "highly qualified" (U.S. Department of Education 2005). In keeping with the fact that they were more likely to be in their first or second year of teaching than comparison teachers, TFA teachers were more likely than comparison teachers to report having had a mentor during the study school year (Table IV.9). Sixty-seven percent of the TFA teachers reported having had a mentor, compared with 29 percent of comparison teachers. TFA teachers were more likely than comparison teachers to have had a mentor assigned by their school as well as to have had a mentor assigned by their program. Among those who had mentors, TFA teachers and comparison teachers spent similar amounts of time both observing other teachers and being observed by their mentors. On average, TFA teachers spent 87 more minutes in formal meetings with their mentors than comparison teachers, a statistically significant difference; there was no statistically significant difference between the two groups in the amount of time spent in informal meeting with mentors.

Table IV.9. 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 Mentor During School Year	67.2	28.6	38.6**	0.000
Mentor Assigned by School	62.5	27.0	35.5**	0.000
Mentor Assigned by Program	31.3	0.0	31.3**	0.000
Average Time Spent Observing Other Teachers (minutes)	92.6	92.7	-0.1	0.997
Average Time Spent Being Observed by Mentors (minutes)	16.8	15.1	1.7	0.842
Average Time Spent in Formal Meetings with Mentors (minutes)	147.9	61.4	86.5**	0.007
Average Time Spent in Informal Meetings with Mentors (minutes)	102.2	73.5	28.7	0.495
Number of Times Received Written Feedback on Teaching Performance	2.2	1.6	0.6	0.447
Felt that Mentoring Was Very Helpful	48.4	23.8	24.6**	0.004
Sample Size	**64**	**63**		

Source: Survey of Secondary Math Teachers.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

5. Views Toward Teaching

Because teachers' levels of satisfaction with their jobs have the potential to influence how long they stay in teaching, we measured the job satisfaction of both TFA and comparison teachers. TFA teachers were less satisfied with some aspects of their job than comparison teachers, particularly those aspects related to school policies and staff (Table IV.10). Teachers were given a list of 10 aspects of their teaching job and asked whether they were "very dissatisfied," "somewhat dissatisfied," "somewhat satisfied," or "very satisfied" with this aspect of their job. TFA and comparison teachers reported statistically similar levels of satisfaction with aspects of their jobs related to their own classrooms (autonomy and control over their own classrooms, students' motivation to learn, students' discipline and behavior, and availability of resources for the classroom). However, TFA teachers reported significantly lower levels of satisfaction for aspects of their jobs related to school policies and staff, including their influence over school policy, opportunities for professional development, procedures for performance evaluation, and the professional caliber of their colleagues.

Table IV.10. Job Satisfaction of TFA and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Percentage of Teachers Who Were Somewhat or Very Satisfied with this Aspect of Job[a]			
Aspect of Job	TFA Teachers	Comparison Teachers	Difference	*p*-Value
Autonomy and Control Over Own Classroom	84.4	85.7	-1.3	0.834
Students' Motivation to Learn	39.1	38.7	0.4	0.968
Students' Discipline and Behavior	27.0	34.9	-7.9	0.339
Availability of Resources for Classroom	65.1	66.1	-1.0	0.903
Recognition/Support from Administration	59.4	67.7	-8.4	0.333
Influence Over School Policies and Practices	37.5	62.9	-25.4**	0.004
Opportunities for Professional Development	54.7	82.5	-27.9**	0.001
Principal's Leadership and Vision	56.3	68.3	-12.0	0.165
Procedures for Performance Evaluation	50.0	73.0	-23.0**	0.007
Professional Caliber of Colleagues	57.8	76.2	-18.4*	0.028
Sample Size	**64**	**63**		

Source: Survey of Secondary Math Teachers.

[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 between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

C. Students in the TFA Study Sample

Among students in the analysis sample (those for whom we have outcome test score data), those assigned to the TFA teachers (the treatment group) and those assigned to the comparison teachers (the control group) were statistically similar in terms of their baseline characteristics (Table IV.11). This suggests that random assignment was properly implemented and that student attrition after random assignment (due to lack of outcome test score data) did not lead to differences in observable baseline characteristics between the two groups. Of the 23 differences in baseline characteristics examined, three were statistically significant at the 5 percent level, and these differences were small in magnitude. These same differences were present for the full sample of students who were randomly assigned (including those with missing outcome data), as shown in Appendix Table A.5—this indicates that the differences occurred at the time of random assignment and were not due to differential attrition between the treatment and control groups after random assignment.

Consistent with TFA's goal of serving disadvantaged students, treatment and control group students in the analysis sample scored, on average, about half a standard deviation below average for their states in both reading and math in the years prior to the evaluation, as indicated by average *z*-scores for both groups of around -0.50. Both treatment and control group students were 13.4 years old on average. The majority of students (around 62 percent) in both groups were black, non-Hispanic, while 28 percent in both groups were Hispanic, 7 percent were white, non-Hispanic, and 2 percent were Asian. Thus a higher percentage of the study sample were members of racial and ethnic minority groups relative to public secondary school students nationwide,

Table IV.11. Average Baseline Characteristics of Students in the Analysis Sample Who Were Assigned to TFA Teachers or Comparison Teachers (percentages unless otherwise indicated)

Characteristic	Secondary School Students Nationwide	Assigned to TFA Teachers	Assigned to Comparison Teachers	Difference Between TFA and Comparison	*p*-Value
Baseline Math Achievement					
Baseline math score (average *z*-score)	0.00	-0.51	-0.50	-0.01	0.531
Score below 25th percentile of state	25.0	39.6	39.4	0.2	0.741
Score below 50th percentile of state	50.0	73.1	72.7	0.3	0.641
Baseline Reading Achievement					
Baseline reading score (average *z*-score)	0.00	-0.51	-0.51	0.00	0.734
Score below 25th percentile of state	25.0	40.6	39.8	0.7	0.393
Score below 50th percentile of state	50.0	71.9	73.0	-1.1	0.163
Age (average years)	n.a.	13.44	13.39	0.05**	0.002
Female	50.2	48.6	50.0	-1.5	0.102
Race and Ethnicity					
Asian, non-Hispanic	4.7	2.1	1.8	0.2	0.391
Black, non-Hispanic	15.9	62.1	62.5	-0.4	0.656
Hispanic	18.2	28.3	27.7	0.5	0.471
White, non-Hispanic	60.1	6.5	6.6	-0.1	0.802
Other race/ethnicity	1.2	1.1	1.4	-0.3	0.319
Eligible for Free/Reduced-Price Lunch	35.0	89.9	90.5	-0.7	0.469
Limited English Proficient	6.5	8.0	8.4	-0.4	0.475
Individualized Education Plan	12.5	6.4	6.0	0.4	0.358
Grade Level in Study Year					
6th grade	7.4	27.2	27.2	0.0	.[a]
7th grade	7.5	28.6	28.7	-0.1	0.164
8th grade	7.6	19.5	19.5	0.1	0.164
9th grade	8.6	14.6	14.5	0.2	0.555
10th grade	7.9	7.3	7.8	-0.6*	0.012
11th grade	7.2	2.6	2.4	0.2	0.237
12th grade	6.6	0.2	0.0	0.2*	0.044
Number of Students		**2,292**	**2,281**		
Number of Classroom Matches		**111**	**111**		
Number of Teachers		**66**	**70**		
Number of Schools		**45**	**45**		

Source: Nationwide statistics on demographic characteristics from the 2007–2008 Schools and Staffing Survey, available at http://nces.ed.gov/pubs2009/2009321/. Data on nationwide grade distribution from the 2006–2007 "Digest of Education Statistics: Enrollment in Public Elementary and Secondary Schools, by Grade," available at http://nces.ed.gov/programs/digest/d08/tables/ dt08_037.asp. Data on study students from 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 and clustering at the teacher level.

[a] *p*-value is not defined because there is no variation in this variable within classroom matches.

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

TFA = Teach For America.

among whom 16 percent were black, 18 percent were Hispanic, and 60 percent were white, non-Hispanic. Around 90 percent of students in both groups were eligible for free or reduced-price lunch, higher than the national average of 35 percent. In both groups, around 8 percent of students had limited English proficiency (similar to the national average of 7 percent) and around 6 percent had an Individualized Education Plan (IEP) for a special education program or services. The percentage of students with an IEP in the TFA sample is lower than the national average of 13 percent, perhaps because students with more severe disabilities or restrictive IEPs were exempted from random assignment and excluded from the study sample.

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

V. TEACH FOR AMERICA: IMPACTS ON SECONDARY STUDENT MATH ACHIEVEMENT

A central goal of school administrators, policymakers, and parents is to select teachers who are most effective in increasing student achievement. In this chapter, we examine the impact of Teach For America (TFA) secondary math teachers on student achievement in math, relative to the comparison teachers who taught the same courses in the same schools. To estimate impacts, we compared the end-of-year math achievement of students assigned to TFA teachers and those assigned to comparison teachers. Random assignment ensured that these two groups of students had similar achievement levels and other characteristics at the start of the school year, so that differences in achievement between the two groups at the end of the school year could be reliably attributed to the effectiveness of the teachers. TFA and comparison teachers could differ in effectiveness as a result of two types of factors: (1) the characteristics they had before entering their preparation programs and (2) the training and support they received from their programs. The impact findings presented here are based on TFA and comparison teachers whose classes were fully comparable and whose schools and districts agreed to be in the study; the findings may not necessarily apply to TFA teachers in other types of schools, subjects, or demographic environments.

For all impact analyses, the outcome of interest was a student's end-of-year math score. We expressed the outcome as a *z*-score based on means and standard deviations of scores for the full student population in the same state and grade (in middle school grades) or the national norming sample of the student's assessment (in high school grades). Therefore, differences in effectiveness between TFA and comparison teachers, expressed in student *z*-score units, can be regarded as effect sizes measured in standard deviation units within a statewide or national reference population. We estimated impacts within the full set of classroom matches in which TFA teachers were included (Section A) as well as within subgroups of matches defined by characteristics of the TFA or comparison teachers (Sections B and C).

A. Impacts of TFA Teachers Relative to All Comparison Teachers

1. Average Impact

On average, TFA teachers in the study were more effective than comparison teachers. Students assigned to TFA teachers scored 0.07 standard deviations higher on end-of-year math assessments than students assigned to comparison teachers (Table V.1). Although the average end-of-year math score in both groups was below the mean in the statewide or national reference populations (indicated by negative average *z*-scores), TFA teachers raised the average achievement of their students to a level that was closer to the population mean.

The magnitude of the difference in effectiveness between TFA and comparison teachers can be interpreted in several ways. First, the effect size can be expressed as a change in percentiles of achievement within the reference populations used in the study. If assigned to a comparison teacher, the average student in the study would have had a *z*-score of -0.60, equivalent to the 27th percentile of achievement in his or her reference population based on a normal distribution for test scores. If assigned to a TFA teacher, this student would, instead, have had a *z*-score of -0.52—equivalent to the 30th percentile. Thus, the average student in the study would gain three percentile points from being assigned to a TFA teacher rather than a comparison teacher. A limitation of this interpretation, however, is that a given effect size equates to a different

percentile gain at different points in the achievement distribution. Specifically, the effect size of 0.07 translates to a 3 percentile point change for a student who, like the average student in the study, would have scored at the 27th percentile if assigned to a comparison teacher. However, the same effect size would imply a smaller percentile point change for lower-performing students and a larger percentile point change for students closer to the 50th percentile.

Table V.1. Difference in Effectiveness Between TFA Teachers and All Comparison Teachers

	Mean Outcome		Difference in Outcomes Between Groups	
Outcome Variable	Assigned to TFA Teachers	Assigned to Comparison Teachers	Estimate	*p*-Value
Student Math Achievement (*z*-score)	-0.52	-0.60	0.07**	0.000
Sample Sizes				
Number of Students	**2,292**	**2,281**		
Number of Classroom Matches	**111**	**111**		
Number of Teachers	**66**	**70**		
Number of Schools	**45**	**45**		

Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: The difference between groups is adjusted for classroom match fixed effects and all covariates in Appendix Table B.1. The mean outcome for students assigned to TFA teachers is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups.

*Estimate is statistically significant at the 0.05 level based on a two-tailed test after adjusting for multiple hypothesis testing.

**Estimate is statistically significant at the 0.01 level based on a two-tailed test after adjusting for multiple hypothesis testing.

TFA = Teach For America.

Alternatively, the effect size can be compared with educationally relevant benchmarks. An illustrative benchmark is the average one-year gain in achievement exhibited by students on nationally normed assessments in grades 6 through 11, as tabulated by Hill et al. (2008). On the basis of this benchmark, the difference in achievement between students of TFA teachers and students of comparison teachers amounted to 26 percent of an average year of learning by students nationwide—that is, 2.6 months of learning in a 10-month school year.

The finding that TFA teachers were more effective than comparison teachers is robust to multiple sensitivity analyses. We estimated models that (1) excluded covariates, (2) used alternative methods of handling missing data and invalid test scores, (3) used alternative approaches to weighting classroom matches, (4) dropped classroom matches with high rates of student mobility, (5) dropped matches with supplemental math classes, (6) used alternative ways of standardizing the outcome variable, and (7) accounted for students who switched to a different type of teacher than their originally assigned teacher (see Appendix Table F.1). Students of TFA teachers were estimated to outperform students of comparison teachers by a statistically significant degree in all models.

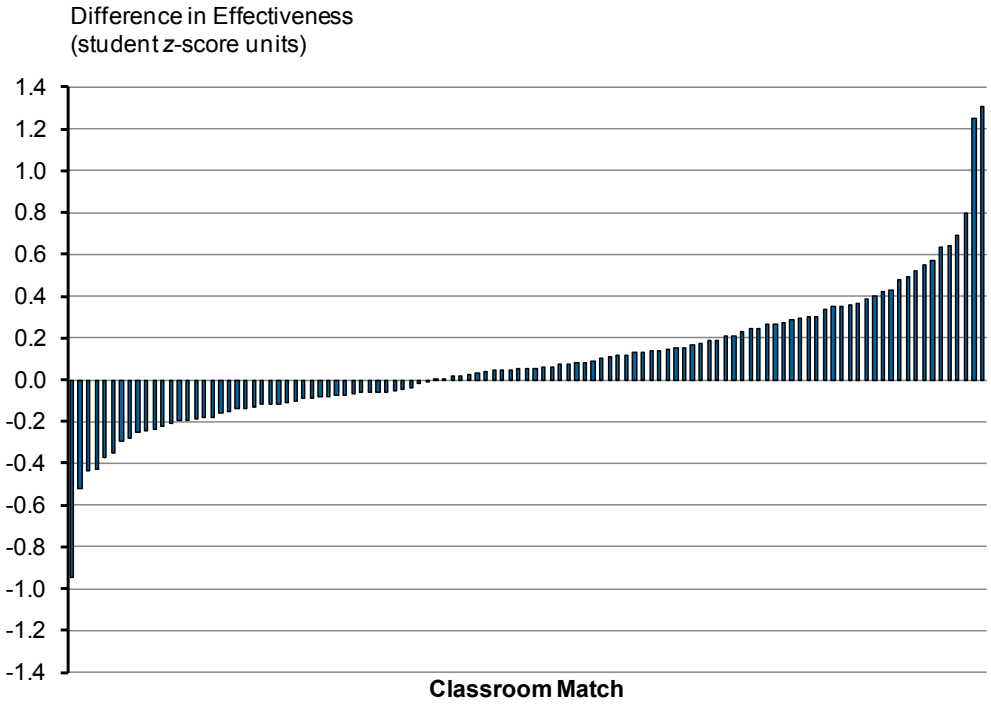
In supplemental analyses, we examined impacts on measures of student absences, a channel through which teachers could affect student math achievement. We found no statistically significant differences in the rate of absences during the study year between students of TFA teachers and students of comparison teachers (Appendix Table F.2).

2. Variation in Impacts

Although TFA teachers had a positive average impact on student math achievement relative to comparison teachers, impacts from individual classroom matches varied in both sign and magnitude (Figure V.1). Notably, not all TFA teachers were more effective than their counterparts; without regard to statistical significance, the estimated difference in effectiveness between TFA and comparison teachers was positive in 60 percent of classroom matches (67 out of 111) and negative in the remaining 40 percent. Because each match-specific estimate was based on a small number of students, random statistical error contributed to the variation in impact estimates across classroom matches. Nevertheless, on the basis of an *F*-test, we found that the observed variation in impact estimates across classroom matches exceeded the variation that would be expected from pure statistical chance.

Therefore, although TFA teachers were more effective than comparison teachers on average, our findings suggest that an individual TFA teacher could be more or less effective than an individual comparison teacher. In other words, even after accounting for teachers' route to certification (TFA or comparison), teachers varied in effectiveness. Chapter IX explores additional factors that might account for this variation in effectiveness.

Figure V.1. Differences in Effectiveness Between TFA Teachers and Comparison Teachers from Individual Classroom Matches



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Notes: Differences in effectiveness from individual classroom matches are estimated from a regression of students' end-of-year math scores on classroom match indicators, an indicator for being assigned to a TFA teacher fully interacted with all classroom match indicators, and all covariates in Appendix Table B.1. There are 111 classroom matches represented in the figure. An *F*-test of the null hypothesis that differences in effectiveness between TFA teachers and comparison teachers are equal across classroom matches has a *p*-value of less than 0.001.

TFA = Teach For America.

Given that the difference in effectiveness between TFA and comparison teachers varied across matches, we examined whether differences were more or less pronounced when specific types of TFA and comparison teachers were compared. The subgroup analyses that address this question are discussed in the remainder of this chapter; detailed results are presented in Appendix Table F.3.

B. Impacts of TFA Teachers Relative to Comparison Teachers from Traditional and Alternative Routes to Certification

The comparison teachers who were compared with TFA teachers consisted of a mix of teachers from two distinct routes to certification: traditional routes to certification (TC routes) and less selective alternative routes to certification (AC routes). As we discussed in Chapter IV, on a number of dimensions of academic background and training, patterns of differences between TFA and comparison teachers varied according to whether the comparison teachers entered teaching through TC or AC routes. Therefore, the average impact of TFA teachers might mask differences in how TFA teachers performed relative to each type of comparison teacher. In fact, one criticism of AC routes, such as TFA, is that teachers from those routes who have entered the classroom without completing all certification requirements have less pedagogical knowledge—and are therefore less effective—than counterparts from TC routes (Darling-Hammond 1990, 2000; Darling-Hammond et al. 2005). To examine this possibility, we compared the effectiveness of TFA and comparison teachers within two separate subgroups of classroom matches: (1) those with TC comparison teachers and (2) those with AC comparison teachers.

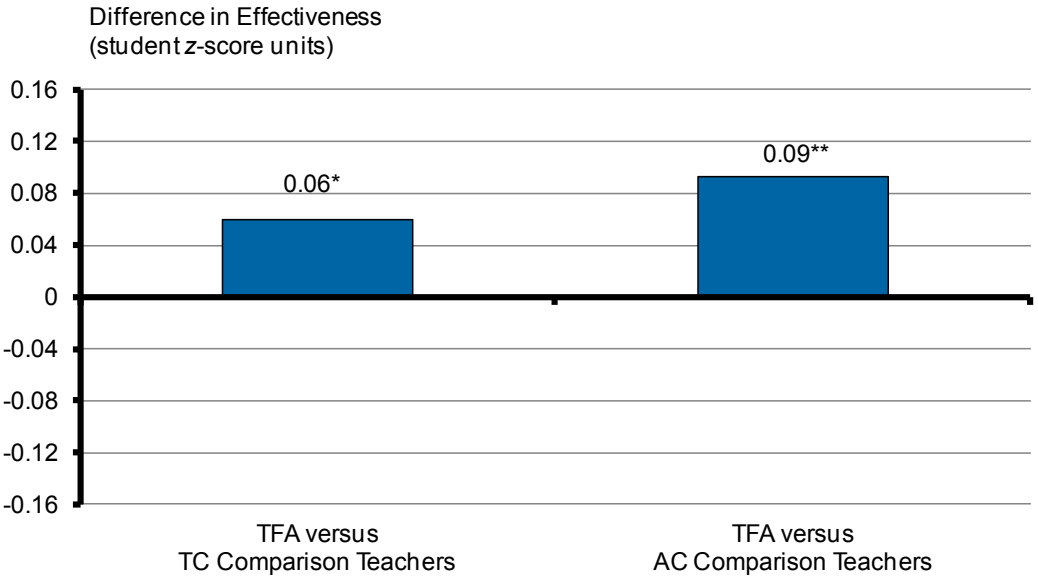
On average, TFA teachers were more effective than comparison teachers from both TC and less selective AC routes. In classroom matches in which TFA teachers were compared with TC comparison teachers, students of TFA teachers outperformed those of TC comparison teachers by 0.06 standard deviations (Figure V.2). Likewise, students of TFA teachers scored 0.09 standard deviations higher than students of AC comparison teachers in the same classroom match. These findings suggest that a strategy of hiring TFA teacher to teach the classes in the study would lead to higher student math achievement than would a strategy of hiring either TC or AC comparison teachers.

C. Impacts Within Other Teacher Subgroups

1. Subgroups Defined by Teaching Experience

Although the findings presented thus far shed light on the effectiveness of TFA teachers who are less experienced, on average, than their counterparts from other routes, school administrators might also want to know how student achievement at their school would differ in the short term as a result of hiring a new TFA teacher rather than some other new, inexperienced teacher. We gauged the short-term effects of choosing a new TFA teacher over a new non-TFA teacher by comparing TFA and comparison teachers who had recently entered teaching. Specifically, we classified all teachers into two experience categories: (1) novice teachers, defined as those in their first three years of teaching, and (2) experienced teachers, defined as

Figure V.2. Differences in Effectiveness Between TFA Teachers and Comparison Teachers from Traditional and Alternative Routes to Certification



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.1

*Estimate is statistically significant at the 0.05 level based on a two-tailed test after adjusting for multiple hypothesis testing.

**Estimate is statistically significant at the 0.01 level based on a two-tailed test after adjusting for multiple hypothesis testing.

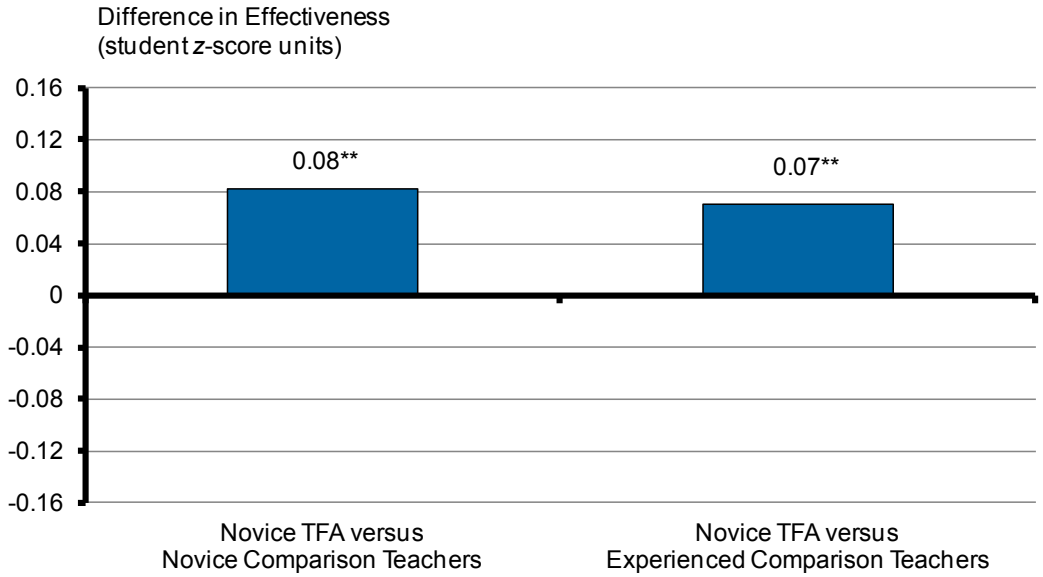
AC = alternative route to certification; TC = traditional route to certification; TFA = Teach For America.

those in their fourth year of teaching or beyond. We examined impacts within the subgroup of classroom matches in which novice TFA teachers were compared with novice comparison teachers. We also examined impacts within other experience subgroups to determine whether the positive impact of TFA teachers was limited to particular types of comparisons or was observed more broadly. Because very few (4 of 66) TFA teachers were experienced, the remaining comparisons that included TFA teachers in the study were primarily those in which novice TFA teachers were compared with experienced comparison teachers.

Novice TFA teachers were more effective than novice comparison teachers in the study. Average math achievement was higher by 0.08 standard deviations for students of novice TFA teachers compared with students of novice comparison teachers (Figure V.3). The evidence implies that a net gain in student achievement from hiring a new TFA teacher rather than a new non-TFA teacher would be realized even in the short run.

Novice TFA teachers also raised student math achievement by 0.07 standard deviations relative to experienced comparison teachers despite having 11 fewer years of experience, on average (Figure V.3). Taken together, our findings indicate that novice TFA teachers were more effective than comparison teachers irrespective of whether the comparison teachers were novice or experienced.

Figure V.3. Differences in Effectiveness Between Novice TFA Teachers and Comparison Teachers, by Experience Level of Comparison Teacher



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.1. Novice teachers are defined as teachers in their first three years of teaching. Experienced teachers are defined as teachers in their fourth or more year of teaching.

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

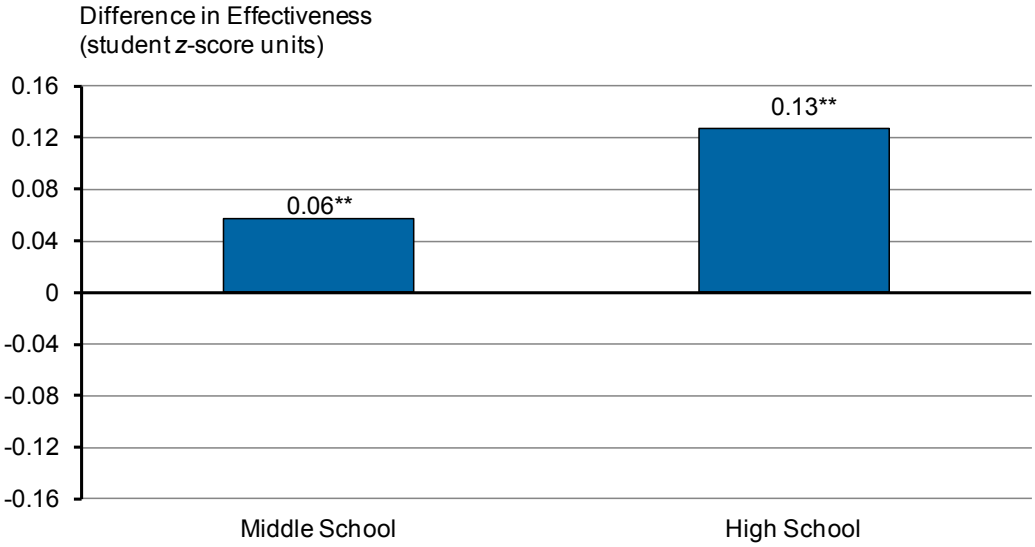
TFA = Teach For America.

2. Subgroups Defined by Middle and High School

We examined differences in effectiveness between TFA and comparison teachers separately for middle school and high school teachers due to a number of differences between the two sets of grade levels. First, the state assessments taken by middle school students differed from the study-administered Northwest Evaluation Association (NWEA) assessments taken by high school students. For example, the state assessments were expected to be aligned more closely with state curriculum standards, but the NWEA assessments were expected to be less prone to floor and ceiling effects due to their computer-adaptive structure. Second, high school courses covered more advanced mathematics than did middle school courses. This difference could influence impact findings if the effective teaching of math at different levels required different sets of instructional skills. As Chapter IV shows, TFA teachers at both the middle and high school levels were more likely than comparison teachers to have graduated from a selective college or university, had higher average scores on the Praxis II, and were less likely to have majored in math. They also had taken fewer college-level math courses than comparison teachers at both the middle and high school levels, although this difference was more pronounced among high school teachers.

At both the middle school and high school levels, TFA teachers had positive impacts relative to comparison teachers (Figure V.4). Students of TFA teachers outscored those of comparison teachers by 0.06 standard deviations in middle school and 0.13 standard deviations in high school.

Figure V.4. Differences in Effectiveness Between TFA Teachers and Comparison Teachers Within Middle Schools and High Schools



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.1.

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

TFA = Teach For America.

D. Conclusions

TFA aims to place high-quality teachers in schools that serve some of the most disadvantaged students in the United States. As we discuss in Chapter III, the program focuses its recruitment efforts primarily on recent college graduates and asks them to make a two-year commitment to teaching. TFA is highly selective, with a multi-stage application process, and accepts only about 12 percent of applicants. The program attempts to identify applicants it believes will be most successful in the classroom based on both staff judgment and predictions of a mathematical selection model. It provides its newly selected teachers with an intense five-week summer training program as well as mentoring and support throughout the teachers' two-year commitment to the program.

Consistent with TFA's mission to place teachers in high-need schools, TFA teachers in the study were typically placed in urban schools that served economically disadvantaged students, largely from racial and ethnic minority groups. As we discuss in Chapter IV, the non-TFA study teachers in these schools were primarily nonwhite women, typically with more than five years of teaching experience. In contrast, the TFA teachers in the study were overwhelmingly white, less

likely to be female, and had significantly less teaching (and general work) experience. The TFA teachers were also far more likely to have attended a selective undergraduate university. Although the TFA teachers took fewer college-level math courses, they had significantly greater math content knowledge test scores than the typical math teachers in the schools in which they taught.

The experimental analysis indicates that TFA teachers were more effective than comparison teachers. On average, students assigned to TFA teachers scored 0.07 standard deviations higher on end-of-year math assessments than students assigned to comparison teachers. The impact translated into an additional 2.6 months of learning for the average student nationwide. Findings were similar when we compared specific subgroups of TFA and comparison teachers:

- **TFA teachers were more effective than comparison teachers from both TC and less selective AC routes.** Students of TFA teachers outperformed students of teachers from TC routes by 0.06 standard deviations and outperformed students of teachers from less selective AC routes by 0.09 standard deviations.
- **Novice TFA teachers were more effective than both novice and experienced comparison teachers.** Students of novice TFA teachers—those in their first three years of teaching—outscored students of novice comparison teachers by 0.08 standard deviations. Students of novice TFA teachers also outperformed students of more experienced comparison teachers (those with more than three years of experience) by 0.07 standard deviations. This latter finding is particularly important given the fact that TFA requires its teachers to make only a two-year commitment to teaching.
- **TFA teachers were more effective than comparison teachers in both middle and high schools.** Students of TFA teachers outscored those of comparison teachers by 0.06 standard deviations in middle school and 0.13 standard deviations in high school.

VI. SELECTION, TRAINING, AND SUPPORT OF TEACHING FELLOWS PROGRAM PARTICIPANTS

Like teachers who enter the profession through Teach For America (TFA), teachers who enter teaching through Teaching Fellows programs may differ from other teachers in their effectiveness in teaching mathematics. This may be because of their characteristics, knowledge, and experiences when they enter the program, and/or because of the training they receive from the program and the support they receive after they start teaching. This chapter presents details on how Teaching Fellows programs select, train, and support their participants. Appendix C presents some additional information on the programs and summarizes the features of both the Teaching Fellows programs and TFA.

A. Application and Selection

The Teaching Fellows application and selection process has two stages: (1) submission of an online application, and (2) an in-person assessment. Individuals apply to a specific Teaching Fellows program based on where they want to teach (Appendix Figure C.3 lists the 19 Teaching Fellows programs as of August 2011). Each program has its own application form and may have some different eligibility requirements, but all the programs follow a similar selection process, according to officials from TNTP, the national organization that oversees the Teaching Fellows programs.

1. Application Form and Eligibility Rules

The online application form collects information on the colleges or universities applicants have attended, the highest degree they have earned, their fields of study, their overall undergraduate grade point average (GPA), any languages they speak in addition to English, and whether they are U.S. citizens. Applicants must also provide short written answers (200 to 400 words) to three questions on such topics as (1) why they want to teach in high-need schools and what experience they have in such settings, (2) what skills or abilities they have that will help them succeed in the classroom, and (3) what steps they have taken to overcome challenges they have faced. Responses to application form questions about personal demographic information—sex, age, race/ethnicity—are optional. Applicants can specify subject area preferences (for example, secondary math); however, those who reach the in-person assessment stage are informed what type of position they are being considered for, based on program officials' reviews of their applications, transcripts, and districts' needs.

Applicants also must submit material at different points in the application and selection process. They must submit a resume with their application. Those who advance beyond the application stage must submit unofficial undergraduate and graduate school transcripts; if they reach the point of accepting a teaching position, they must submit official transcripts. At the hiring stage, candidates must provide proof of citizenship. Applications and related documents from all Teaching Fellows programs are reviewed online by a centrally overseen set of "pre-screeners" who determine which candidates should be invited to the in-person assessment stage.

To be eligible for admission, all applicants to all Teaching Fellows programs must meet three basic, objective eligibility criteria: (1) hold a bachelor's degree, (2) not have been certified

to teach, and (3) be able to pass a background investigation. In addition, some eligibility criteria vary by state or program, or by teacher subject area or level:

- **Undergraduate institution.** In some states, the bachelor's degree must be from a U.S. institution.
- **GPA.** Requirements for cumulative undergraduate GPA vary by program. Of the seven Teaching Fellows programs in this study still operating in 2011, two listed on their website an official GPA requirement of 3.0, and five listed a requirement of 2.75. Four of the programs, however, allow some exceptions to these standards, giving applicants a chance to explain why their GPA was below the standard or to demonstrate academic achievement in other ways. One of the programs with a 3.0 GPA requirement allowed such exceptions for applicants with a GPA of at least 2.75, and one of the programs with a 2.75 GPA requirement allowed exceptions for applicants with a GPA of at least 2.5.
- **Coursework.** Requirements may vary by state for teachers of certain subjects and grade levels. For example, some states require secondary math teacher candidates to have majored in math, to have earned a requisite GPA in math courses, to have taken a requisite number of college-level math courses, or to agree to take certain math courses while working toward initial certification.
- **New-teacher tests.** All candidates must meet applicable state policies for tests required of all new teachers to be deemed highly qualified to begin teaching. Teaching Fellows programs require candidates to meet the state-specified passing scores on all applicable examinations; they do not set higher thresholds. Candidates typically must have passed all applicable tests before the start of the school year, although some programs require candidates to pass tests before the start of the summer institute.

State- and program-specific eligibility requirements are made clear to applicants at the start of the application process.

2. Selection Process

The application and selection process used in every Teaching Fellows program is designed to highlight the degree to which each applicant embodies seven core competencies:

1. Critical thinking
2. Achieving ambitious goals
3. Personal responsibility and accountability
4. Commitment to students in high-need schools
5. Dedication to continuous improvement
6. Communication skills
7. Respectful and positive in all situations

In general, when TNTP pre-screeners decide which candidates for each local program should be invited to the second and final stage of the application and selection process, known as

the "Interview Event," they consider two factors: (1) the completeness and quality of the application, including the quality of responses to open-ended questions; and (2) the applicant's experience, including academic and other accomplishments, and exposure to high-need schools and communities. Pre-screeners place applicants into one of five tiers, ranging from "concerns" to "strong." In general, candidates must be in the upper tiers to be invited to the Interview Event, but exceptions may be made in response to district needs for certain types of teachers.

Each Teaching Fellows program conducts its own local Interview Event. The event, which lasts about six hours, consists of four activities:

1. **Sample teaching lesson.** Candidates take turns presenting a five-minute sample teaching lesson to one another as program staff observe and evaluate. Applicants may target any subject and grade level for the lesson. The amount of time candidates have to prepare their sample lesson depends on how close to the Interview Event they receive their invitation; some applicants may have just four or five days' notice, while others may know about the requirement weeks in advance.
2. **Writing exercise.** Applicants have 20 minutes to describe in writing how they would handle a realistic, challenging classroom scenario.
3. **Group discussion.** Groups of about 10 applicants spend 17 minutes discussing two questions posed by program staff, such as how they would respond to a particular, challenging classroom scenario. Participants are also instructed to draw on articles sent to them before the Interview Event. The articles, often taken from major newspapers, relate to current or recent education policy issues in low-income urban schools. Staff observe the discussion but do not participate in it. When the discussion time concludes, participants have three minutes to respond in writing to a summary question posed by the session organizer.
4. **One-on-one interview.** Applicants participate in a one-on-one interview, lasting about 25 minutes. Following an interview guide developed by TNTP, staff may either use a standard set of questions or choose from a bank of questions, across three areas. They ask at least one question about preceding activities in the Interview Event and often a question about the applicant's resume or answers to the written application questions. They ask several "scenario questions" about how the applicant would respond to a hypothetical classroom situation and several "targeted questions" to elicit more evidence of the applicant's skills, accomplishments, and dedication and to round out a "holistic picture" of the applicant. Interviewers are encouraged to ask several follow-up questions to elicit complete, specific answers from applicants, and to gather robust, concrete evidence needed to assess applicants' candidacy.

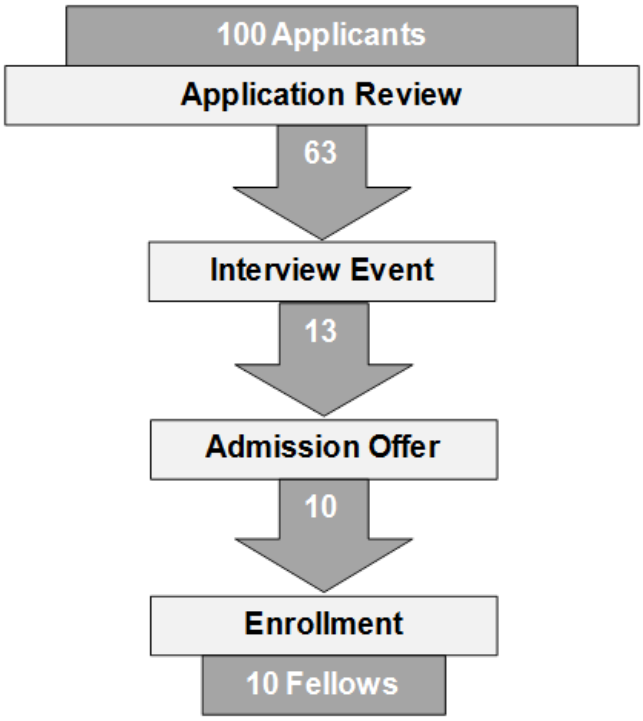
After the Interview Event, local selectors (a different group than the pre-screeners) consider all available information and complete a final evaluation form on each candidate. The form includes a rating on each of the seven competencies listed above. The basic ratings are "exemplary," "fully acceptable," and "not fully acceptable"; a plus or minus sign may be attached to each rating when candidates appear to be slightly above or below the basic ratings. To be recommended for admission, applicants must earn "exemplary" on at least one of the first four competencies and no ratings of "not fully acceptable" on any competency.

In the three years preceding our interviews, according to TNTP officials, the application and selection process for Teaching Fellows programs changed in three important ways as officials worked to refine and improve the process. First, they developed the five-tier system for screening applications. Second, they added "commitment to students in high-need schools" as a ranked competency. Third, they introduced the option to use pluses and minuses around each of the three basic ratings (exemplary, fully acceptable, not fully acceptable) applied to the seven core competencies. These changes, intended to help identify candidates who would be effective Teaching Fellows, provided additional information to help program officials make strategic decisions about who should receive an offer to join the program, and reportedly had the incidental effect of making it somewhat harder for applicants to gain admission.

3. Program Selectivity and Application Yields

For an overall perspective on program selectivity, Figure VI.1 shows the flow of applicants through the selection process in the three years preceding our interviews. According to TNTP officials, for every 100 applicants, on average, 63.3 advanced past the application stage to be invited to the Interview Event (about 55 percent of those attended), 12.6 were offered admission, and 9.7 enrolled. About 13 percent of applicants were given an offer to join a Teaching Fellows program, and, ultimately, 76.7 percent of those offered the opportunity chose to enroll in a Teaching Fellows program.

Figure VI.1. Teaching Fellows Selection Process and Applicant Flow



Source: TNTP officials.

TNTP officials noted that math teacher candidates were accepted at slightly lower rates than the overall rates shown above. One reason, they said, was that such candidates were less likely to be experienced in working in low-income communities or with underserved youth. They were also less likely to enroll when offered admission; officials thought that the refusals might

indicate that candidates were considering other teacher programs, including TFA, or nonteaching jobs.

For another view on program selectivity, we asked TNTP officials how Teaching Fellows programs would respond if a district's projected need for secondary math teachers were higher than usual or higher than expected, such that the district would likely face difficulty in meeting that need, given its applicant pool and typical process for making admission decisions. The officials indicated that programs may adjust their degree of selectivity but would never make admission offers to candidates not deemed "fully acceptable." Officials also emphasized that rather than adjust selectivity, they would prefer to have programs encourage more candidates to apply, even extending the application deadline, if necessary, to attract potentially more qualified candidates. In addition, the hypothetical situation presented here may not apply to individual Teaching Fellows programs because placement goals are typically broad—for example, "50 teachers in the areas of math and science"—rather than stating a specific target for the number of secondary math teachers.

B. Preservice Training for New Teaching Fellows

For new participants in Teaching Fellows programs, the main preservice training experience is a summer institute, but it is preceded by self-guided study and an orientation session. A few programs also provide a supplementary math immersion program, typically before the institute. Figure VI.2 shows an overview of the timing of placement and all training and support activities before and after Teaching Fellows begin teaching.

1. Independent Study and Orientation

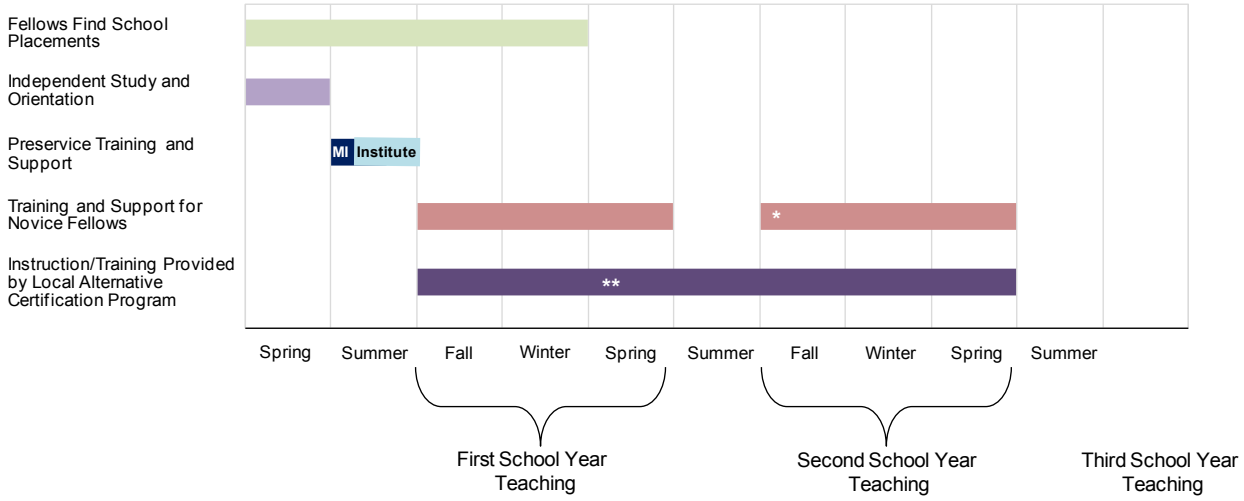
Teaching Fellows first complete assignments described in an Independent Study Guide, including a good deal of reading, classroom observations (either a full day or a series of shorter sessions), and written responses to questions. Altogether, independent study may take about 25 hours. Next, Teaching Fellows participate in a 4-hour orientation session, either in person or via webinar, that provides information on district hiring and on future training and support and includes an overview of the summer institute.

2. Summer Institute

Each Teaching Fellows program hosts its own full-time summer institute. (Programs serving several regions or districts may hold several institutes.) The summer institutes in nearly all Teaching Fellows programs reportedly last between five and six weeks, and Teaching Fellows attend every weekday. One program in the study runs a seven-week institute, and Teaching Fellows attend four days a week. The two primary activities Teaching Fellows engage in during the institute are fieldwork in public summer school classrooms and courses conducted in school buildings or a university. Institutes are customized for the districts they serve, but all follow the same training curriculum.

Training for Teaching Fellows is organized into what TNTP calls the Teaching for Student Achievement Framework. For general education teachers, the framework encompasses 10 competencies across two domains, as shown in Table VI.1.

Figure VI.2. Overview of Placement and Training of Teaching Fellows



Source: TNTP officials.

Note: The shaded areas of this figure are meant to show that various activities generally do, or may, take place during a particular season; shading does not imply that any given activity takes up the full season. The text of this chapter provides information on the hours associated with various activities.

MI = Math Immersion Program; required of some Teaching Fellows in some programs, to help them meet requirements to teach secondary math courses.

*Although not typical, a modest amount of support may be provided, on an as-needed basis, to Teaching Fellows in their second year of teaching.

**Timing of required coursework varies; it may start as early as the summer before Teaching Fellows begin teaching and may last into their third year of teaching. When coursework extends beyond the first year of teaching, it may or may not take place during summer(s).

One major component of the institute is formal instruction. Teaching Fellows engage in coursework nearly every day of the institute, an average of 27 days in all. On days when fieldwork is scheduled in the morning (described below), coursework takes place in the late afternoon. Formal instruction is led by Teaching Fellows program staff and district staff, many of whom are Teaching Fellows or TFA alumni. To help new general education teachers develop the 10 competencies listed in Table VI.1, the curriculum is delivered in six "framework modules," each spanning approximately three to five sessions covering theory and pedagogy:

- Module 1—How can I use the Teaching for Student Achievement Framework to help me close the achievement gap in a "high-need" school?
- Module 2—How do I prevent and handle disruptions to maximize the time I spend teaching?
- Module 3—How can I effectively teach a class with students whose readiness varies greatly?
- Module 4—How do I plan and deliver rigorous and engaging instruction?
- Module 5—What does good literacy instruction look like?
- Module 6—How can I establish credibility in the classroom and community?

Table VI.1. Domains and Corresponding Competencies of TNTP's Teaching for Student Achievement Framework

Instructional Design and Delivery	Classroom Management and Culture
• Set rigorous academic goals • Use appropriate assessments • Create standards-based lessons and units • Apply differentiated instruction • Use high-impact teaching strategies	• Effect change as a newcomer • Use diversity to promote achievement • Address misbehavior effectively • Develop rules, procedures, and consequences • Create a "no excuses" classroom culture

Source: TNTP officials.

In addition to the standard Teaching Fellows curriculum delivered though the modules above, district officials (such as central administrators or curriculum and professional development staff) or school site leaders may schedule instructional workshops to cover district- or school-specific topics. Overall, counting both the framework modules and the additional district-specific sessions, Teaching Fellows reportedly spend an average of 2.5 hours per day receiving instruction, for an estimated average of 67.5 hours over 27 days.

Another major component of the institute, and an important way for teachers to develop the 10 competencies, is fieldwork, referred to as "practice teaching," in a district's summer school program. Program staff strive to maximize the time that Teaching Fellows spend in practice teaching. The programs expect all Teaching Fellows to engage in practice teaching for a minimum of 15 days. The scheduling of fieldwork—whether it is all on consecutive days and how soon it starts after the beginning of the institute—varies by program and depends on the schedule of local summer school sessions. Program staff would like Teaching Fellows to spend 5 hours a day on-site, but this is not possible if summer school operates for only 4 hours a day. National staff estimate that Teaching Fellows spend an average of 4 to 4.5 hours a day on-site on fieldwork days, or an estimated average of 64 hours in fieldwork over 15 days. Teaching Fellows are assigned in pairs to work with a regular summer school teacher (the "cooperating teacher"), and secondary math candidates typically are assigned to secondary math classrooms. Teaching Fellows may spend their first few days in the classroom mainly observing the teacher and students but are expected to assume increasing responsibility for leading instruction and other routine activities. They first work with small groups during sessions led by the cooperating teacher and move toward leading instruction for the whole class in back-to-back lessons as the fieldwork period progresses.

The final component of the summer institute is assessment and intervention. Through formative and summative assessments, program staff evaluate a Teaching Fellow's performance and provide feedback to help with his or her development. For example, staff observe each Teaching Fellow in practice teaching sessions three times and provide feedback after each session.

During the institute, Teaching Fellows are reportedly involved in training for 8.5 to 9 hours per day, including the activities described above, the commute to/from fieldwork sites, and optional activities (such as visiting institute staff during their office hours and participating in other meetings). Even when instructional sessions and other scheduled activities conclude for the day, Teaching Fellows often remain busy into the evening, completing course assignments or lesson planning for practice teaching sessions.

Over the three years preceding the spring 2010 interviews, according to TNTP officials, 94 percent of those who enrolled in a Teaching Fellows program, on average, completed the summer institute. Most (72 percent) of those who did not complete the institute left for personal reasons; others failed to meet eligibility requirements or exited from the program due to performance. Math teacher candidates were reportedly no more or less likely than other participants to complete the summer institute.

3. Post-Institute Activities

Three Teaching Fellows programs in the study offered "math immersion" programs during the summer. These programs involve an intensive review of mathematical concepts to help individuals otherwise ineligible to teach secondary math become qualified to do so and to achieve more success in the classroom. Table VI.2 lists key characteristics of the math immersion programs. The amount of instruction provided in the programs ranged from 30 to 80 hours, with a mean of 62 hours. Respondents reported that each of the three math immersion programs had just one curriculum for all participants and that the curriculum did not include math pedagogy. They also reported that the math immersion programs were open only to Teaching Fellows for whom participation was mandatory; they were not an option for other Teaching Fellows.

Table VI.2. Characteristics of Math Immersion Programs Offered by Three Teaching Fellows Programs in the Study

	Program 1	Program 2	Program 3
Candidates required to attend	Anyone who did not have 24 college credits in math and 6 related credits	Anyone who did not have 30 college credits in math or who did not major in math	Anyone who did not major in math as an undergraduate or graduate student
Estimated percentage of secondary math teacher candidates who participated in the program, past 3 years	61%	80%	80%
Amount of instruction provided	75 hours (10 days, 7.5 hours per day)	80 hours (10 days, 8 hours per day)	30 hours (10 days, 3 hours per day)
College-level math subjects covered in program	Algebra, geometry, calculus, trigonometry, linear algebra	Algebra, geometry, calculus, trigonometry, linear algebra	Algebra, calculus

Source: Interviews with Teaching Fellows officials.

C. Placement and Retention of Teaching Fellows

After they accept an offer of admission to a Teaching Fellows program, Teaching Fellows must take an active role in finding a suitable, appropriate teaching position; they are not placed in schools by program or district staff. Candidates receive a list of local schools that, in the district's assessment, have the greatest need for Teaching Fellows. Candidates are encouraged to apply to those schools, but they may also consider schools not on the list, if necessary, to find a suitable vacancy in the subject area and grade range in which they wish to teach. The timing of the process reportedly varies from program to program and year to year. Most hiring takes place during the summer institute; however, Teaching Fellows may receive job offers before the

summer institute or not be hired until after the institute (in some cases, not until after the school year begins). Teaching Fellows programs do not place a priority on securing placements for their new Teaching Fellows in schools that already have teachers from the program.

In the three hiring cycles before our interviews, an average of 92 percent of all new candidates in the eight Teaching Fellows programs received a job offer before the start of the school year, and an additional 3 percent were hired after the school year started (Table VI.3). Hiring rates were 3 percentage points higher for secondary math candidates. TNTP officials noted that the averages were pulled down somewhat by the most recent year (placements for fall 2009), which saw lower placement rates than in any previous year. According to TNTP officials, many districts had fewer openings than expected because of layoffs and hiring freezes brought about by districts' budget problems. They said that, in earlier years, the rates were always close to 100 percent.

Table VI.3. Estimated Percentage of New Teaching Fellows Who Received Teaching Placements, on Average, Across the 2007–2008, 2008–2009, and 2009–2010 School Years

	Hired Before Start of School Year	Hired After Start of School Year	Total
All Candidates	92%	3%	95%
Secondary Math Candidates	95%	3%	98%

Source: Interview with TNTP officials.

In recent years, according to TNTP officials, an average of 75 percent of Teaching Fellows who received a teaching placement remained in teaching for more than two school years. Math teachers were reportedly less likely than others to complete two years of teaching (specific figures were not available).

D. Instruction Provided to Teaching Fellows Through Local Teacher Certification Programs

All eight Teaching Fellows programs required new secondary math teachers to enroll in a state-authorized program that provided an alternative route to certification (AC program). To gain a general sense of the instruction that secondary math Teaching Fellows in our study might have received from local AC programs, we asked officials from each participating Teaching Fellows program to identify the one program that had enrolled the largest number, whether a majority or plurality, of their new secondary math teachers over the three most recent years before 2010. This was the time period, we believed, during which most study teachers would have enrolled in a local AC program. (The later addition of study teachers from Teaching Fellows programs' summer 2010 entering cohort means that their AC program experiences may not be as well represented by the AC program information we collected.) In three of the Teaching Fellows programs we examined, the identified AC program reportedly served all the secondary math Teaching Fellows. In the remaining five Teaching Fellows programs, the identified AC program reportedly served 50 to 85 percent of secondary math Teaching Fellows. Hereafter, we refer to these as the "focal" AC programs, as they were the focus of our questions.

A variety of institutions operated the eight focal AC programs. Five were operated by a nearby college or university, one by the local school district, and two by TNTP. As of 2011, TNTP operated six such programs, serving the District of Columbia, Louisiana, Maryland, Oakland (California), Rhode Island, and Texas.

1. Preservice Instruction

Secondary math teachers (and other teachers) in three of the programs in the study were required to participate in some instruction for their AC program in the summer before they started teaching (in addition to the summer institute). All three programs were run by local colleges. One program met seven hours a day for four days, separate from the Teaching Fellows summer institute. The other two programs required participants to take their first course toward certification concurrent with the summer institute. One of these programs' classes met for two hours, three nights a week, for six weeks. The other program's summer classes met for 2.5 hours, one night a week, for eight weeks, with an additional 8-hour session on a Saturday.

2. Instruction During and After the First Year of Teaching

In all eight Teaching Fellows programs in the study, secondary math teachers had to attend classes or other training activities toward their initial certification during their first year of teaching. The focal AC program with the most extensive first-year requirement involved an estimated 140 hours of instruction. The least extensive first-year requirement was 45 hours, with participants attending 3 hours a week for 15 weeks. Across the eight focal certification programs, the average amount of instruction required of first-year secondary math teachers was about 86 hours.

Just one of the focal AC programs featured an instruction component during the summer after teachers' first year of teaching. Participants reportedly met for 2.5 hours a week for four weeks, for a total of 10 hours.

In four Teaching Fellows programs in the study, secondary math teachers had to continue taking classes or other training activities toward their initial certification during their second year of teaching. The AC program with the most extensive second-year coursework requirement involved 4 hours a week for 30 weeks, for a total of 120 hours of instruction. The least extensive of the four programs involved 3 hours a week for 15 weeks, for a total of 45 hours. Across all eight programs, the average second-year instruction requirement for new secondary math teachers was about 38 hours.

3. Total Amount of Instruction in Certification Programs

The total hours of instruction reportedly required by the certification programs that were the focus of our interviews—before and after secondary math Teaching Fellows started teaching—ranged from about 61 to 250 hours, with a mean of about 137 hours. In Table VI.4, we present the average instruction hours required, overall and by time period.

Table VI.4. Hours of Instruction Toward Certification Reportedly Required in Focal AC Programs for Secondary Math Teaching Fellows

Instruction Provided	Programs with This Instructional Component				Mean for All 8 Programs
	Number of Programs	Minimum	Maximum	Mean	
Before first year of teaching	3	28	36	30.7	11.5
During first year of teaching	8	45	140	86.2	86.2
During summer after first year of teaching	1			10	1.3
During second year of teaching	4	45	120	76.5	38.3
Total Hours, All Programs	**8**	**60.8**	**250**	**137.2**	**137.2**

Source: Interviews with Teaching Fellows programs and TNTP officials.

Notes: This table excludes training provided directly by Teaching Fellows programs.

The focal AC program for each participating Teaching Fellows program is the one that enrolled the largest number of its secondary math teachers from 2007 to 2009.

E. Training and Support for New Teaching Fellows After They Begin Teaching

Program staff continued to work with Teaching Fellows in a variety of ways after they began teaching.

1. Instruction, Training, and Professional Development

Teaching Fellows programs reportedly deliver 10 hours of professional development to new participants in group sessions that take place on weeknights and weekends. Topics include classroom management, how to use data to inform instruction, and how to tailor instruction for different students. Staff also poll the participants or develop ideas for other professional development topics based on classroom observations (described below). TNTP staff described the material covered in the sessions as more practical and less theoretical than the instruction provided in the college-based AC programs that some participants attend while teaching. None of the professional development is uniquely designed for secondary math teachers.

2. Other Support Services

In the standard model for Teaching Fellows programs, each new teacher is assigned to a staff member, known as a "training and resource manager" or a "training and instruction manager" (hereafter, "training manager"), who interacts one-on-one with the teachers in several ways during their first year or two of teaching to monitor and foster their development.

- **Classroom observations.** The training manager conducts at least two formal classroom observations of each new teacher, with the first taking place within the first six weeks of the school year. (When a program is too large for the training manager to conduct all the observations, additional staff, called "instructional specialists" or "field visitors," are hired to conduct observations.) Some Teaching Fellows programs reportedly conduct considerably more classroom observations of new Teaching Fellows. The observer discusses the observation assessments mainly with the teacher

but will speak to a teacher's principal if he or she thinks the teacher might benefit from additional district- or school-based supports.

- **One-on-one meetings.** Training managers schedule two meetings during the school year with each of their teachers to talk one-on-one, acting more or less as a mentor. The meetings last an average of 20 minutes. (New teachers in most Teaching Fellows programs also have a formal mentor assigned to them by their district or school, according to information collected through our teacher surveys [Chapter VII]).
- **Informal discussions.** Training managers may call or email their teachers occasionally for informal check-in discussions (for example, to determine whether the teachers need any information or resources that the training manager could provide). The average frequency and length of these casual, impromptu interactions is not systematically tracked.

Although the assignments of new Teaching Fellows to program staff typically remain in place only for a Teaching Fellow's first year of teaching, a staff member might maintain contact with his or her teachers during the second year (for example, sending them newsletters or inviting them to optional professional development events). More generally, program staff provide both formal and informal support to any Teaching Fellows they believe need assistance.

F. Financial Support for Teaching Fellows

At the time of our interviews, participants in six of the eight Teaching Fellows programs reportedly could receive direct or indirect financial assistance to help cover training expenses. The percentage of eligible participants who received available financial assistance ranged from 73 to 100 percent; the mean percentage receiving financial assistance was about 93 percent. Support ranged from a district covering the full cost of the AC program, estimated at $3,500, to a district covering about $14,000 of the cost of required coursework toward a master's degree, which reportedly cost in total about $20,600.

VII. TEACHING FELLOWS STUDY TEACHERS: THEIR CHARACTERISTICS, THE SCHOOLS THEY TEACH IN, AND THE STUDENTS THEY TEACH

The Teaching Fellows programs aim to recruit and train teachers to teach in low-income school districts, with recruitment efforts that focus on both recent college graduates and mid-career professionals. Given the particular application and hiring processes the Teaching Fellows programs use, Teaching Fellows may differ from their counterparts who entered the teaching profession through other routes. Because the programs aim to place their Teaching Fellows in high-poverty schools, the schools they work in and the students they serve are also likely to differ from typical schools and students nationwide. This chapter provides descriptive information on the schools (Section A), teachers (Section B), and students (Section C) in the Teaching Fellows study sample.

To summarize the findings from this chapter, we found that Teaching Fellows in the study were typically placed in urban schools that served economically disadvantaged students, largely from racial and ethnic minority groups. The Teaching Fellows in the study were typically white, with an average of four years of teaching experience and five years of nonteaching work experience, and included roughly equal numbers of men and women. The other secondary math teachers in these same schools were primarily white or black women, with considerably more teaching experience (13 years) and three years nonteaching work experience, on average. Although the Teaching Fellows teachers were less likely to have majored in math than the teachers from other routes to certification, they were far more likely to have attended a selective undergraduate university and had significantly higher math content knowledge test scores.

A. Schools in the Teaching Fellows Study Sample

Even though schools in the Teaching Fellows sample were not randomly selected from the full set of secondary schools with Teaching Fellows placements nationwide, the study schools were similar to secondary schools employing Teaching Fellows nationwide along many dimensions (Table VII.1). For example, most students at both sets of schools were members of racial and ethnic minority groups, although the study schools had more racial and ethnic minorities than did the full set of Teaching Fellows schools. On average, 6 percent of students in the study schools in the Teaching Fellows sample were white, non-Hispanic, compared with 17 percent in Teaching Fellows secondary schools nationwide. About 46 percent of students at both study schools and all secondary schools with Teaching Fellows were black, non-Hispanic. Forty percent of students at the study schools and 35 percent of students in all secondary schools with Teaching Fellows placements nationwide were Hispanic. Teaching Fellows study schools had a large proportion of students eligible to receive free or reduced-price lunch—80 percent of study school students and 74 percent of all Teaching Fellows secondary school students. Eighty-two percent of the schools in the Teaching Fellows study sample and 85 percent of secondary schools with Teaching Fellows nationwide were eligible for Title I funds for schools with high percentages of low-income students. The grade distribution of study schools was also similar to the grade distribution of secondary schools with Teaching Fellows nationwide—about 48 percent of both sets of schools were middle schools (including at least one grade from 6 to 8), and 60 to 70 percent were high schools (including at least one grade from 9 to 12).

Consistent with the goal of the Teaching Fellows programs to serve low-income communities, schools in the study sample were, on average, more disadvantaged than the typical

secondary school nationwide. For example, compared with students in the average secondary school nationwide, students in the average school in the Teaching Fellows study sample were more likely to be eligible for free or reduced-price lunch and to be from racial and ethnic minorities.

Table VII.1. Characteristics of Study Schools with Teaching Fellows Compared with All Secondary Schools with Teaching Fellows and All Secondary Schools Nationwide

	Study Schools with Teaching Fellows	All Secondary Schools with Teaching Fellows[a]		All Secondary Schools Nationwide	
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	6.6	4.2	0.227	3.5	0.099
Percentage Black, non-Hispanic	46.5	45.6	0.877	18.8**	0.000
Percentage Hispanic	40.2	35.0	0.333	20.2**	0.000
Percentage White, non-Hispanic	6.3	16.6**	0.000	59.4**	0.000
Percentage other race/ethnicity	0.5	2.7	0.059	9.8**	0.000
Student Socioeconomic Status					
Percentage eligible for free/ reduced-price lunch	80.4	74.4	0.132	50.5**	0.000
Percentage Title I-eligible schools	81.8	86.6	0.476	60.0**	0.000
Enrollment and Staffing					
Average total enrollment	1,211.0	989.3	0.288	551.2**	0.001
Average enrollment per grade	298.0	248.3	0.346	134.5**	0.001
Average student/teacher ratio	16.3	15.8	0.464	16.1	0.768
Grade Level[b]					
Percentage middle schools	47.7	47.1	0.939	66.1*	0.015
Percentage high schools	70.5	61.3	0.258	45.3**	0.000
School Type					
Percentage magnet schools	6.8	5.9	0.832	2.3	0.237
Census Bureau Region					
Percentage in Northeast	50.0	19.3**	0.000	14.3**	0.000
Percentage in Midwest	11.4	19.3	0.187	27.7**	0.001
Percentage in South	27.3	45.4*	0.025	30.8	0.601
Percentage in West	11.4	16.0	0.392	26.9**	0.001
Sample Size	**44**	**119**		**63,148**	

Sources: TNTP placement data; Common Core of Data, Public Elementary/Secondary School Universe Survey, 2009–2010.

Note: Secondary schools are defined as any schools that include at least one grade above 5th grade.

[a]Estimates are based on schools in which new teaching fellows were placed in the 2009–2010 school year.

[b]Middle schools are defined as schools with at least one grade from 6 to 8. High schools are defined as those with at least one grade from 9 to 12. Thus, the two categories are not mutually exclusive, as a given school could contain both middle and high school grades.

*Difference between this group and study schools with Teaching Fellows (first column) is statistically significant at the 0.05 level, two-tailed test.
**Difference between this group and study schools with Teaching Fellows (first column) is statistically significant at the 0.01 level, two-tailed test.

Study schools and all Teaching Fellows secondary schools nationwide also differed on some dimensions, many of which may have been due to our recruitment strategies and study eligibility requirements. Nine percent of secondary schools employing Teaching Fellows nationwide were charter schools, compared with fewer than 5 percent of the study schools, a statistically significant difference (estimates not shown in table and exact percentage for study schools not reported to protect respondent confidentiality in accordance with National Center For Education Statistics statistical standards [National Center for Education Statistics 2000]). Charter schools typically were smaller than average and therefore less likely to have eligible classroom matches. Study schools also were significantly more likely to be in urban areas—more than 95 percent of study schools were in urban areas, versus 78 percent of Teaching Fellows secondary schools nationwide (also not shown in table and exact percentage for study schools not reported to protect respondent confidentiality), consistent with the fact that we concentrated on recruiting for the study districts with the greatest numbers of Teaching Fellows placements. Study schools were also more likely than the full set of Teaching Fellows secondary schools to be in the Northeast (50 versus 19 percent) and less likely to be in the South (27 versus 45 percent).

B. Teaching Fellows in the Study Sample Compared with Comparison Teachers

The Teaching Fellows programs aim to place effective teachers in low-income schools by selecting people with the qualities and attitudes they believe are predictive of successful teaching, providing these people with an intensive five- to seven-week training, and then supporting them through their first year of teaching. For these reasons, Teaching Fellows study teachers are likely to differ from the comparison teachers with whom they were compared—the non-Teaching Fellows teaching the same math course in the same school. In fact, we found that Teaching Fellows did differ from comparison teachers in important ways, with a few similarities, as described in the box below.

Compared with comparison teachers, Teaching Fellows:

- Were younger and less likely to be members of racial and ethnic minority groups
- Were more likely to have graduated from a selective college or university
- Had greater math content knowledge as measured by scores on the Praxis II math exams
- Were less likely to have majored in math or secondary math education
- Had fewer years of teaching experience, although they had more nonteaching work experience
- Completed similar amounts of math pedagogy instruction and math pedagogy professional development, but spent less time student teaching in math during their teacher training
- Were about as likely to have taken coursework during the study school year and to have had a mentor during that year
- Were less satisfied in general with some aspects of teaching, including school administration.

The comparison teachers with whom Teaching Fellows were compared included teachers from traditional routes to certification (TC routes) and those from less selective alternative routes to certification (AC routes)—73 percent of comparison teachers were from TC routes, while 27 percent were from less selective AC routes. With a few exceptions discussed below, patterns of differences between Teaching Fellows and comparison teachers were present both among the subset of Teaching Fellows who were compared with teachers from TC routes and the subset who were compared with teachers from AC routes (Appendix D).

1. Demographic Characteristics

Teaching Fellows and comparison teachers differed along several demographic characteristics (Table VII.2).[21] Teaching Fellows study teachers were younger than comparison teachers, on average (33 versus 41 years old), and were less likely to be members of racial and ethnic minority groups. Seventy-one percent of Teaching Fellows in the study were white, non-Hispanic, compared with 43 percent of comparison teachers. Seventeen percent of the Teaching Fellows in the study were black, non-Hispanic, compared with 36 percent of comparison teachers. Comparison teachers in the study were closer in age and gender distribution to the average secondary teacher nationwide than were Teaching Fellows in the study, whereas Teaching Fellows in the study more closely resembled the average secondary teacher in terms of racial and ethnic distribution.

2. Educational Background and Content Knowledge

Teaching Fellows in the study were more likely than comparison teachers to have graduated from a selective college or university (72 versus 34 percent, as well as from a highly selective college or university (21 versus 7 percent) (Table VII.3).[22] They were about as likely as comparison teachers to have completed a graduate degree (83 versus 80 percent; difference not significant at the 5 percent level). They were less likely to have majored in math (25 versus 43 percent), secondary math education (0 versus 13 percent), or any math-related subject, including math, secondary math education, statistics, engineering, computer science, finance, economics, physics, or astrophysics (56 versus 75 percent).

Despite the fact that they were less likely to have majored in math, Teaching Fellows completed approximately the same number of college-level math courses as comparison teachers (Table VII.4). Both groups of teachers took about the same number of each type of college-level math course, including calculus, advanced algebra, analysis, advanced geometry/topology, probability and statistics, discrete math, and applied math. Differences in average number of college-level math courses were similar at both the middle and high school levels.

[21] In this and all teacher-level analyses presented in this chapter, each teacher is counted only once, regardless of whether he or she taught multiple classes in the study. For teachers in both cohorts, the value of the variable as reported in cohort 1 is used if cohort 1 and cohort 2 reports differ. We also conducted analyses that weight each teacher according to the number of study classes taught, and findings were similar.

[22] College selectivity was defined based on *Barron's Profiles of American Colleges 2003*. Highly selective colleges were defined as those ranked by Barron's as most competitive, and selective colleges were defined as those ranked by Barron's as very competitive, highly competitive, or most competitive.

Table VII.2. Demographic Characteristics of Teaching Fellows and Comparison Teachers in the Study and All Secondary Teachers Nationwide (percentages unless otherwise indicated)

	Secondary Teachers Nationwide	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Age (Average Years)	42.7††^^	33.3	41.0	-7.8**	0.000
Female	59.3	53.8	57.3	-3.5	0.681
Race/Ethnicity[a]					
Asian, Non-Hispanic	1.3†^^	9.1	18.7	-9.6	0.099
Black, Non-Hispanic	6.9†^^	16.7	36.0	-19.3**	0.009
Hispanic	6.8	9.4	13.3	-4.0	0.464
White, Non-Hispanic	83.5†^^	71.2	42.7	28.5**	0.001
Sample Size	**1,099,770**	**66**	**75**		

Sources: Data for secondary teachers nationwide from the Schools and Staffing Survey Teacher Questionnaire, 2007–2008; data for study teachers from the Survey of Secondary Math Teachers.

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

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

†Difference between secondary teachers nationwide and Teaching Fellows is statistically significant at the 0.05 level, two-tailed test.
††Difference between secondary teachers nationwide and Teaching Fellows is statistically significant at the 0.01 level, two-tailed test.

^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
^^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

Moreover, although Teaching Fellows were less likely to have majored in math and reported taking roughly the same number of college-level math courses as comparison teachers, the Teaching Fellows demonstrated higher mathematics content knowledge (Table VII.5). This was measured by the Praxis II Mathematics Content Knowledge Test (taken primarily by the high school teachers in the sample, along with a few middle school teachers in states that allowed or required middle school teachers to take this test) or the Praxis II Middle School Mathematics Test (taken by the remaining middle school teachers in the sample).[23] On the Mathematics Content Knowledge Test, Teaching Fellows performed an average of 19 points higher than comparison teachers (or 0.80 standard deviations). Similarly, those teachers who took the Middle School Mathematics Test outperformed comparison teachers by 17 points (or 0.92 standard deviations). These patterns persisted both among teacher matches for which the Praxis was "high

[23] Response rates on the Praxis II differed for Teaching Fellows and comparison teachers—94 percent of the Teaching Fellows teachers in the sample had Praxis II scores, compared with only 81 percent of comparison teachers (Appendix Table A.18). In the main analysis, we omitted teachers with missing scores. As a sensitivity test, in Appendix Table E.2 we examined differences in Praxis II scores when we imputed values that would be otherwise missing. Results were similar under the two approaches, although differences between the two groups were even more pronounced, suggesting that teachers with missing Praxis data might have been those who would have scored lower had they taken the test.

stakes" (those in districts that required the exam for certification) and those for whom the Praxis was "low stakes" (not required for certification, and taken only at the request of the study team).

Table VII.3. Educational Background of Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Bachelor's Degree				
From a highly selective college or university[a]	21.1	7.1	13.9*	0.033
From a selective college or university[a]	71.9	33.9	38.0**	0.000
Major[b]				
Math	24.6	42.9	-18.3*	0.040
Secondary math education	0.0	12.5	-12.5**	0.006
Other math-related subject[c]	33.3	23.2	10.1	0.236
Other subject	66.7	35.7	31.0**	0.001
Math, secondary math education, or other math-related subject[c]	56.1	75.0	-18.9*	0.035
Major or minor[b]				
Math	29.8	51.8	-22.0*	0.017
Secondary math education	0.0	25.0	-25.0**	0.000
Other math-related subject[c]	38.6	30.4	8.2	0.361
Other subject	78.9	53.6	25.4**	0.004
Graduate Degree				
Any graduate degree	83.3	80.0	3.3	0.612
Subject[b]				
Math or math-related subject[c]	12.1	22.7	-10.5	0.098
Secondary math education	57.6	40.0	17.6*	0.037
In other education subject	19.7	24.0	-4.3	0.540
Other subject	12.1	9.3	2.8	0.598
Sample Size	**66**	**75**		

Source: Survey of Secondary Math Teachers.

[a]Highly selective colleges are those ranked by *Barron's Profiles of American Colleges 2003* as being most competitive. Selective colleges are those that are ranked by *Barron's* as being very competitive, highly competitive, or most competitive.

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

[c]Math-related subjects include statistics, engineering, computer science, finance, economics, and physics/astrophysics.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

Table VII.4. Average Number of College-Level Math Courses Taken by Teaching Fellows and Comparison Teachers in the Study

	Teaching Fellows	Comparison Teachers	Difference	p-Value
All Teachers				
Total Number of College-Level Math Courses	9.8	11.0	-1.2	0.320
Calculus	2.6	2.9	-0.4	0.192
Advanced Algebra	1.3	1.6	-0.3	0.190
Analysis	0.7	0.9	-0.2	0.295
Advanced Geometry/Topology	0.6	0.8	-0.2	0.155
Probability and Statistics	1.7	1.7	0.0	0.983
Discrete Mathematics	1.3	1.2	0.1	0.739
Applied Mathematics	1.5	1.8	-0.2	0.616
Sample Size	**66**	**75**		
Middle School Teachers				
Total Number of College-Level Math Courses	8.2	9.3	-1.1	0.648
Sample Size	**26**	**22**		
High School Teachers				
Total Number of College-Level Math Courses	10.8	11.7	-0.9	0.495
Sample Size	**40**	**53**		

Source: Survey of Secondary Math Teachers.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

While it would be useful to compare average scores of study teachers to average scores of the full set of teachers who took the Praxis II mathematics tests nationwide, published data on national average scores are available only for test takers who passed the test. For comparison to these statistics, we also examined the average scores of the teachers in our sample who passed the test. On the Mathematics Content Knowledge Test, Teaching Fellows in the study who passed the test scored an average of 164 points, compared with an average of 155 for all test-takers nationwide who passed. Similarly, on the Middle School Mathematics Test, Teaching Fellows in the study who passed the test scored an average of 187 points, compared with an average of 169 for all test-takers nationwide who passed.

Table VII.5. Praxis II Scores of Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Test-Takers Nationwide	Teaching Fellows	Comparison Teachers	Teaching Fellows-Comparison Difference	*p*-Value
Overall					
Mathematics: Content Knowledge Test					
Average score	n.a.	157.8	138.5	19.3**	0.000
Average score of those who scored above the state's passing threshold[a]	154.5††	163.8	154.1	9.7*	0.036
Sample size	5,785	39	47		
Middle School Mathematics Test					
Average score	n.a.	186.7	170.1	16.5**	0.001
Average score of those who scored above the state's passing threshold[a]	169.0††	186.7	172.6	14.0**	0.002
Sample size	9,532	26	21		
High-Stakes Test-Takers[b]					
Mathematics: Content Knowledge Test					
Average score	n.a.	154.1	134.3	19.8*	0.040
Sample size		14	10		
Low-Stakes Test-Takers[c]					
Mathematics: Content Knowledge Test					
Average score	n.a.	159.8	139.6	20.2**	0.001
Sample size		25	37		
Middle School Mathematics Test					
Average score	n.a.	186.7	170.1	16.5**	0.001
Sample size		26	21		

Sources: Data on all test-takers nationwide are from Gitomer and Qi (2010), and include test-takers who did and did not eventually enter the teaching profession. Data for study teachers are from study-administered assessment or were provided by the Educational Testing Service.

[a]Educational Testing Service does not report the average score of all test-takers, but only the average score of those who passed the test in their states. To compare the study sample with the national statistics from the Educational Testing Service, we computed the average scores of sample members who scored above their states' passing thresholds using the Praxis passing thresholds from 2011. In states that did not require this test, the score was compared with the average passing threshold across all states that did require this test.

[b]High-stakes test-takers are those who are in districts that require secondary math teachers to pass the relevant Praxis II assessment for certification. Because no high-stakes test-takers in the Teaching Fellows study sample took the Middle School Mathematics test, we do not present estimates for this group.

[c]Low-stakes test-takers are those who are in districts that do not require secondary math teachers to pass the relevant Praxis II assessment for certification.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

†Difference between secondary teachers nationwide and Teaching Fellows is statistically significant at the 0.05 level, two-tailed test.
††Difference between secondary teachers nationwide and Teaching Fellows is statistically significant at the 0.01 level, two-tailed test.

^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
^^Difference between secondary teachers nationwide and comparison teachers is statistically significant at the 0.01 level, two-tailed test.
n.a. = not available.

3. Teaching and Nonteaching Work Experience

Prior studies have found that, as teachers gain years of experience in the classroom, they become more effective teachers (Hanushek and Rivkin 2006; Kane et al. 2008). The Teaching Fellows in the study had, on average, 4 years of teaching experience at the end of the study school year, compared with an average of 13 years among comparison teachers (Table VII.6). Although the Teaching Fellows in the study did have less teaching experience on average than comparison teachers, most Teaching Fellows were not novice teachers, consistent with the program's expectation that Teaching Fellows will remain in the teaching field for an extended period. In fact, the majority of Teaching Fellows in the study (56 percent) had between three and five years of teaching experience, compared with 25 percent of the comparison teachers. Twenty-three percent of the Teaching Fellows in the study were in their first or second year of teaching, compared with 7 percent of comparison teachers. Sixty-eight percent of the comparison teachers had been teaching for more than 5 years, compared with 21 percent of Teaching Fellows.

Table VII.6. Nonteaching and Teaching Work Experience of Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Nonteaching Work Experience				
Had a nonteaching job after college	65.2	42.7	22.5**	0.007
Years of experience in nonteaching job (average)[a]	5.0	3.0	2.0	0.080
Used college-level math in nonteaching job[a]	40.9	28.0	12.9	0.110
Teaching Experience (End of Study Year)				
Years of teaching experience (average)	4.0	13.0	-9.0**	0.000
1 or 2 years of teaching experience	22.7	6.7	16.1**	0.008
3 to 5 years of teaching experience	56.1	25.3	30.7**	0.000
More than 5 years of teaching experience	21.2	68.0	-46.8**	0.000
Years of experience teaching math (average)	3.6	10.8	-7.2**	0.000
Fewer than 3 years of math teaching experience	24.2	8.0	16.2**	0.010
3 to 5 years of math teaching experience	60.6	29.3	31.3**	0.000
More than 5 years of math teaching experience	15.2	62.7	-47.5**	0.000
Sample Size	**66**	**75**		

Source: Survey of Secondary Math Teachers.

[a]Calculations are based on all teachers, regardless of whether they had a nonteaching job after college.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.

**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

Consistent with the fact that the Teaching Fellows programs recruit experienced professionals to transition into the teaching field, the Teaching Fellows in the study generally had more nonteaching work experience than comparison teachers. Sixty-five percent of the Teaching Fellows held a nonteaching job after college, compared with 43 percent of comparison teachers (Table VII.6). This difference was driven entirely by the subset of Teaching Fellows who were compared with comparison teachers from TC routes—68 percent of these Teaching Fellows had held a nonteaching job, compared with 33 percent of TC comparison teachers, while there was no significant difference between comparison teachers from AC routes and the

Teaching Fellows with whom they were compared (Appendix Table D.14). Forty-one percent of Teaching Fellows had used college-level math in a nonteaching job, as had 28 percent of comparison teachers; this difference was not significant at the 5 percent level.

4. Teacher Training, Professional Development, and Support

Teacher training and professional development. Teaching Fellows reported having received similar amounts of math pedagogy instruction as comparison teachers, but Teaching Fellows reported having fewer days of math student teaching as part of their teacher training (Table VII.7). [24] Both Teaching Fellows study members and comparison teachers completed approximately 48 hours of math pedagogy instruction, on average. However, Teaching Fellows were less likely than comparison teachers to have student taught a math class for more than 20 days during their teacher training (12 versus 61 percent). On average, Teaching Fellows reported that they taught at least one math class to a full classroom of students for 11 days during their teacher training programs, whereas comparison teachers taught a full math class during 38 days on average.

As discussed in Chapter VI, three of the eight Teaching Fellows programs in the study required Teaching Fellows who did not take enough math courses in college to complete a math immersion course in the summer before they began teaching. To determine the number of Teaching Fellows in the study who took math immersion, as well as whether any study teachers from other routes to certification completed a similar course, the teacher survey included the following question: "Some districts require prospective teachers without degrees in math to take an intensive math immersion program in order to become certified to teach secondary math. These programs typically cover the necessary content knowledge to teach secondary math. Before becoming a secondary math teacher, did you participate in this type of program?" Fifty percent of Teaching Fellows in the study reported completing a math immersion course, compared with 4 percent of comparison teachers. Across the three districts in the sample that offered math immersion, 43 percent of Teaching Fellows in the sample took math immersion in the first district, 68 percent took it in the second district, and all took it in the third.

Some differences between Teaching Fellows and their comparison teachers depended on whether the comparison teachers were teachers from AC or TC routes (Appendix Table D.15). For example, Teaching Fellows completed less student teaching in math, on average, than comparison teachers from TC routes (10 versus 47 days), while Teaching Fellows and the AC teachers with whom they were compared completed similar amounts of student teaching. Similarly, Teaching Fellows received more professional development in math during the study year than comparison teachers from TC routes (9 versus 4 hours), while there was no significant difference in the amount of math professional development received by Teaching Fellows and comparison teachers from AC routes.

[24] Because questions about teacher training required more experienced teachers (predominantly comparison teachers) to recall experiences that occurred several years in the past, these data may be less reliable than data pertaining to the study year.

Table VII.7. Training and Professional Development of Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Average Hours of Math Pedagogy Instruction as Part of Teacher Training[a]	48.7	47.5	1.2	0.843
Participated in Intensive "Math Immersion" Program as Part of Teacher Training	50.0	4.0	46.0**	0.000
Days of Student Teaching in Math as Part of Teacher Training[b]				
No days	10.6	22.7	-12.1	0.053
1 to 10	53.0	6.7	46.4**	0.000
11 to 20	24.2	9.3	14.9*	0.019
More than 20	12.1	61.3	-49.2**	0.000
Average Days of Student Teaching in Math	10.8	37.8	-27.0**	0.000
Hours per Day Spent in Student Teaching in Math as Part of Teacher Training (Average)[c]	2.1	3.1	-1.0**	0.000
Average Hours Spent in Math Pedagogy Professional Development During School Year[d]	7.8	5.7	2.2	0.078
Sample Size	**66**	**75**		

Source: Survey of Secondary Math Teachers.

[a]Teachers were asked, "As part of your training to become a math teacher, did you receive any instruction in math pedagogy or strategies to teach math?" If so, "In total, how many hours of instruction in math pedagogy or strategies to teach math 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 math 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 in which you went to an elementary or secondary school and taught one or more math lessons to a whole classroom of students?" If so, "On approximately how many days, in total, did you teach at least one full math lesson (at least one class period) to a whole classroom of students during your teacher education/preparation program?" Possible responses were none, 1 to 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 in math, we created a continuous variable equal to zero for teachers who did not do any student teaching in math, 80 for those who did more than 80 days, and the midpoint of the range for all other categories.

[c]Calculated only among those teachers who said they did some student teaching in math.

[d]Teachers were asked, "During this school year, did you attend any professional development classes, workshops, or seminars *provided by the school district* in math pedagogy or strategies to teach math?" If so, "In total, how many hours did you spend attending these professional development classes, workshops, or seminars in math pedagogy or strategies to teach math?" 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 between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

Coursework during study year. Many states require teachers from alternative routes, including Teaching Fellows, to take coursework in their first few years of teaching to meet certification requirements (Feistritzer 2005). Consistent with the fact that the majority of both Teaching Fellows and comparison teachers in the sample had more than two years of experience, there were no significant differences in the percentage who took coursework during the study year (Table VII.8). There was also no statistically significant difference between Teaching Fellows and comparison teachers in the average number of hours spent doing coursework.

Table VII.8. Coursework Taken During the School Year by Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
All Teachers				
Took Coursework Related to Teaching Job	28.8	22.7	6.1	0.411
Total hours spent during school year on coursework (average)	59.9	36.6	23.3	0.240
Hours spent in class during school year on coursework (average)	27.6	15.9	11.7	0.174
Hours spent out of class during school year on coursework (average)	32.3	20.7	11.6	0.330
Sample Size	**66**	**75**		
Middle School Teachers				
Took Coursework Related to Teaching Job	34.6	9.1	25.5*	0.030
Sample Size	**26**	**22**		
High School Teachers				
Took Coursework Related to Teaching Job	25.0	28.3	-3.3	0.724
Sample Size	**40**	**53**		

Source: Survey of Secondary Math Teachers.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

Mentoring. Teaching Fellows and comparison teachers received similar amounts of mentoring during the study year (Table VII.9). Less than 25 percent of both groups reported having a mentor during the school year, consistent with the fact that less than 5 percent of both groups were in their first years of teaching, when mentoring tends to be most common.

5. Views Toward Teaching

Because teachers' levels of satisfaction with their jobs have the potential to influence how long they stay in teaching, we measured the job satisfaction of both Teaching Fellows and comparison teachers. Teaching Fellows were less satisfied with some aspects of their job than comparison teachers, particularly those related to school policies and staff (Table VII.10). Teachers asked whether they were "very dissatisfied," "somewhat dissatisfied," "somewhat satisfied," or "very satisfied" with 10 different aspects of their job. Teaching Fellows and comparison teachers reported statistically similar levels of satisfaction with aspects of their jobs related to their own classrooms (autonomy and control over their own classrooms, students' motivation to learn, students' discipline and behavior, and availability of resources for the classroom). However, Teaching Fellows reported significantly lower levels of satisfaction for three out of six aspects of their jobs related to school policies and staff, including their influence over school policy, recognition and support from the administration, and the principal's leadership and vision.

Table VII.9. Mentoring Received During the School Year by Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Had a Mentor During School Year	22.7	17.3	5.4	0.430
Average Time Spent Observing Other Teachers (minutes)	70.9	43.5	27.4	0.461
Average Time Spent Being Observed by Mentors (minutes)	4.0	6.5	-2.5	0.656
Average Time Spent in Formal Meetings with Mentors (minutes)	71.2	27.0	44.1	0.145
Average Time Spent in Informal Meetings with Mentors (minutes)	21.3	23.4	-2.1	0.897
Number of Times Received Written Feedback on Teaching Performance	1.0	0.2	0.8	0.067
Felt that Mentoring Was Very Helpful	18.2	14.7	3.5	0.578
Sample Size	**66**	**75**		

Source: Survey of Secondary Math Teachers.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

C. Students in the Teaching Fellows Study Sample

Among students in the impact analysis sample (those for whom we have outcome test score data), those assigned to the Teaching Fellows (the treatment group) and those assigned to the comparison teachers (the control group) were similar in most of their baseline characteristics (Table VII.11). Of the 23 baseline characteristics examined, there were statistically significant differences in five, but these differences were all around 2 percent or less. This suggests that random assignment was properly implemented and that student attrition after random assignment (due to lack of outcome test score data) did not lead to differences in observable baseline characteristics between the two groups. As shown in Appendix Table A.6, treatment and control group students in the full research sample (including both those with and without outcome test score data) were also balanced in terms of baseline characteristics.

Consistent with the Teaching Fellows programs' goal of serving disadvantaged students, both treatment and control group students in the analysis sample scored, on average, about 0.10 standard deviations below average for their states in math and about 0.05 standard deviations below average for their states in reading in the years before the evaluation, as indicated by average z-scores for both groups. Both treatment and control group students were about 14 years old, on average. Approximately half of the students in both groups were black, non-Hispanic, while around 35 percent in both groups were Hispanic, 5 percent were white, non-Hispanic, and 8 percent were Asian. Thus, a higher percentage of the study sample were members of racial and ethnic minority groups relative to public secondary school students nationwide, among whom

Table VII.10. Job Satisfaction of Teaching Fellows and Comparison Teachers in the Study (percentages unless otherwise indicated)

	Percentage of Teachers Who Were Somewhat Satisfied or Very Satisfied with this Aspect of Job[a]			
Aspect of Job	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Autonomy and Control Over Own Classroom	95.5	89.3	6.1	0.169
Students' Motivation to Learn	36.4	45.3	-9.0	0.282
Students' Discipline and Behavior	39.4	48.0	-8.6	0.307
Availability of Resources for Classroom	66.7	69.3	-2.7	0.737
Recognition/Support from Administration	51.5	71.6	-20.1*	0.015
Influence Over School Policies and Practices	38.5	58.7	-20.2*	0.017
Opportunities for Professional Development	59.1	65.8	-6.7	0.422
Principal's Leadership and Vision	57.6	76.0	-18.4*	0.021
Procedures for Performance Evaluation	56.1	64.0	-7.9	0.341
Professional Caliber of Colleagues	74.2	84.0	-9.8	0.159
Sample Size	**66**	**75**		

Source: Survey of Secondary Math Teachers.

[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 between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

16 percent were black, 18 percent were Hispanic, and 60 percent were white, non-Hispanic. Around 75 percent of students in both groups were eligible for free or reduced-price lunch, higher than the national average of 35 percent. In both groups, around 7 percent of students had limited English proficiency (similar to the national average of 7 percent), and around 6 percent had an Individualized Education Plan (IEP) for a special education program or services. The percentage of students with an IEP in the TFA sample is lower than the national average of 13 percent, perhaps because students with more severe disabilities or restrictive IEPs were exempt from random assignment and excluded from the study sample.

Table VII.11. Average Baseline Characteristics of Students in the Analysis Sample Who Were Assigned to Teaching Fellows or Comparison Teachers (percentages unless otherwise indicated)

Characteristic	Secondary School Students Nationwide	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Difference Between Teaching Fellows and Comparison	*p*-Value
Baseline Math Achievement					
Baseline math score (average z-score)	0.00	-0.11	-0.10	-0.01	0.636
Score below 25th percentile of state	25.0	26.0	25.3	0.7	0.453
Score below 50th percentile of state	50.0	56.4	57.1	-0.7	0.493
Baseline Reading Achievement					
Baseline reading score (average z-score)	0.00	-0.06	-0.05	-0.02	0.418
Score below 25th percentile of state	25.0	22.8	22.3	0.5	0.619
Score below 50th percentile of state	50.0	57.1	57.6	-0.5	0.570
Age (average years)		14.31	14.27	0.04**	0.005
Female	50.2	53.1	54.1	-1.0	0.326
Race and Ethnicity					
Asian, non-Hispanic	4.7	8.9	8.1	0.8	0.242
Black, non-Hispanic	15.9	50.4	48.8	1.6*	0.047
Hispanic	18.2	34.9	36.8	-1.9*	0.038
White, non-Hispanic	60.1	5.2	5.5	-0.3	0.580
Other race/ethnicity	1.2	0.6	0.8	-0.3	0.142
Eligible for Free/Reduced-Price Lunch	35.0	73.7	75.9	-2.1*	0.017
Limited English Proficient	6.5	7.7	7.2	0.5	0.332
Individualized Education Plan	12.5	6.7	6.1	0.6	0.233
Grade Level in Study Year					
6th grade	7.4	8.1	8.1	0.0	[a]
7th grade	7.5	14.8	14.8	0.0	[a]
8th grade	7.6	17.6	17.6	0.0	[a]
9th grade	8.6	26.5	27.1	-0.6	0.140
10th grade	7.9	17.9	18.1	-0.1	0.766
11th grade	7.2	14.1	13.1	1.0*	0.040
12th grade	6.6	1.1	1.3	-0.2	0.541
Number of Students		**2,127**	**1,989**		
Number of Classroom Matches		**118**	**118**		
Number of Teachers		**69**	**84**		
Number of Schools		**44**	**44**		

Source: Nationwide statistics on demographic characteristics from the 2007–2008 Schools and Staffing Survey, available at http://nces.ed.gov/pubs2009/2009321/. Data on nationwide grade distribution from the 2006–2007 "Digest of Education Statistics: Enrollment in Public Elementary and Secondary Schools, by Grade," available at http://nces.ed.gov/programs/digest/d08/tables/ dt08_037.asp. Data on study students from 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 Teaching Fellows indicator and classroom match indicators, accounting for sample weights and clustering at the teacher level.

[a] *p*-value is not defined because there is no variation in this variable within classroom matches.

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

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

VIII. TEACHING FELLOWS: IMPACTS ON SECONDARY STUDENT MATH ACHIEVEMENT

In this chapter, we examine the impact of secondary math teachers from Teaching Fellows programs on student achievement in math, relative to the comparison teachers who taught the same courses in the same schools. Our analytic approach is identical to that used in estimating the impacts of Teach For America (TFA) teachers. To estimate the impacts of Teaching Fellows, we compared the end-of-year math achievement of students assigned to Teaching Fellows and those assigned to comparison teachers. Random assignment ensured that these two groups of students had similar achievement levels and other characteristics at the start of the school year, so that differences in achievement between the groups at the end of the school year could be reliably attributed to the effectiveness of the teachers. Teaching Fellows and comparison teachers could differ in effectiveness as a result of two types of factors: (1) the characteristics they had before entering their preparation programs and (2) the training and support they received from their programs. The impact findings presented here are based on Teaching Fellows and comparison teachers whose classes were fully comparable and whose schools and districts agreed to be in the study; the findings may not necessarily apply to Teaching Fellows in other types of schools, subjects, or demographic environments.

As we did for the impacts of TFA teachers in Chapter V, we expressed the impacts of Teaching Fellows in terms of standard deviations of student achievement within a statewide or national reference population. We estimated impacts within both the full set of classroom matches in which Teaching Fellows were included (Section A) as well as within subgroups of matches defined by characteristics of the Teaching Fellows or comparison teachers (Sections B and C).

A. Impacts of Teaching Fellows Relative to All Comparison Teachers

1. Average Impact

On average, Teaching Fellows in the study were no more or less effective than comparison teachers. Students assigned to Teaching Fellows and those assigned to comparison teachers both scored, on average, 0.39 standard deviations below the mean achievement in their statewide or national reference populations (Table VIII.1).

The finding that Teaching Fellows were similar in effectiveness to comparison teachers is robust to a variety of sensitivity analyses. We estimated models that (1) excluded covariates, (2) used alternative methods of handling missing data and invalid test scores, (3) used alternative approaches to weighting classroom matches, (4) dropped classroom matches with high rates of student mobility, (5) dropped matches with supplemental math classes, (6) used alternative ways of standardizing the outcome variable, and (7) accounted for students who switched to a different type of teacher than their originally assigned teacher (Appendix Table F.4). In all models, there was no statistically significant difference in math achievement between students of Teaching Fellows and students of comparison teachers.

These findings suggest that student math achievement within the classes in the study would, on average, be no better or worse if the schools in the study were to hire Teaching Fellows rather than teachers from other certification routes to teach those classes. One important consideration

Table VIII.1. Difference in Effectiveness Between Teaching Fellows and All Comparison Teachers

Outcome Variable	Mean Outcome		Difference in Outcomes Between Groups	
	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Estimate	*p*-Value
Student Math Achievement (*z*-score)	-0.39	-0.39	0.00	0.956
Sample Sizes				
Number of Students	**2,127**	**1,989**		
Number of Classroom Matches	**118**	**118**		
Number of Teachers	**69**	**84**		
Number of Schools	**44**	**44**		

Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: The difference between groups is adjusted for classroom match fixed effects and all covariates in Appendix Table B.2. The mean outcome for students assigned to Teaching Fellows is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups.

*Estimate is statistically significant at the 0.05 level based on a two-tailed test after adjusting for multiple hypothesis testing.
**Estimate is statistically significant at the 0.01 level based on a two-tailed test after adjusting for multiple hypothesis testing.

in interpreting these results, however, is that the Teaching Fellows in the study had, on average, nine fewer years of teaching experience than the comparison teachers. This experience differential could be due to a higher turnover rate among Teaching Fellows relative to comparison teachers. On the other hand, the experience gap could be due to the fact that many of the Teaching Fellows programs were still relatively new at the time of this study, and thus it was not possible for the Teaching Fellows in the sample to have accumulated as much experience as their comparison group counterparts. The findings from Table VIII.1 do not necessarily shed light on how effective Teaching Fellows would be if they were to have experience levels closer to those of their counterparts from other routes, a question we explore further below.

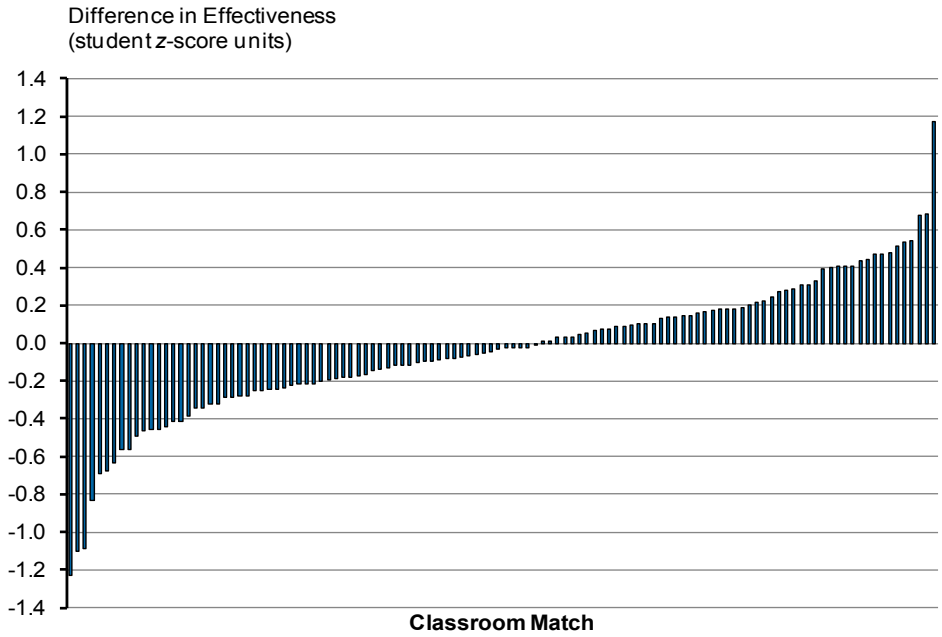
In supplemental analyses, we examined impacts on measures of student absences, a potential channel through which teachers could affect student math achievement. Students assigned to Teaching Fellows were less likely to be chronically absent—defined as being absent more than 25 percent of the time—from their math class during the study year than students assigned to comparison teachers (Appendix Table F.5). The difference in the rate of absences between these groups did not translate into differences in math achievement.

2. Variation in Impacts

Although there was no average difference in effectiveness between Teaching Fellows and comparison teachers, impacts of Teaching Fellows relative to comparison teachers from individual classroom matches varied in both sign and magnitude (Figure VIII.1). Across the 118 matches, the estimated difference in effectiveness between Teaching Fellows and comparison teachers, without regard to statistical significance, was positive in 46 percent of matches but negative in a similar proportion (54 percent) of matches. As with the TFA matches examined in Chapter V, each match-specific estimate for the impact of Teaching Fellows was

based on a small number of students and, thus, had random statistical error that contributed to the variation in impact estimates across classroom matches. Nevertheless, on the basis of an *F*-test, we found that the observed variation in impact estimates across classroom matches exceeded the variation that would be expected from pure statistical chance.

Figure VIII.1. Differences in Effectiveness Between Teaching Fellows and Comparison Teachers from Individual Classroom Matches



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Notes: Differences in effectiveness from individual classroom matches are estimated from a regression of students' end-of-year math scores on classroom match indicators, an indicator for being assigned to a Teaching Fellow fully interacted with all classroom match indicators, and all covariates in Appendix Table B.2. There are 118 classroom matches represented in the figure. An *F*-test of the null hypothesis that differences in effectiveness between Teaching Fellows and comparison teachers are equal across classroom matches has a *p*-value of less than 0.001.

Therefore, although Teaching Fellows and comparison teachers were similar in effectiveness on average, our findings suggest that an individual Teaching Fellow could be more or less effective than an individual comparison teacher. In other words, even after accounting for teachers' route to certification (Teaching Fellow or comparison), teachers varied in effectiveness. Chapter IX explores additional factors that might account for this variation in effectiveness.

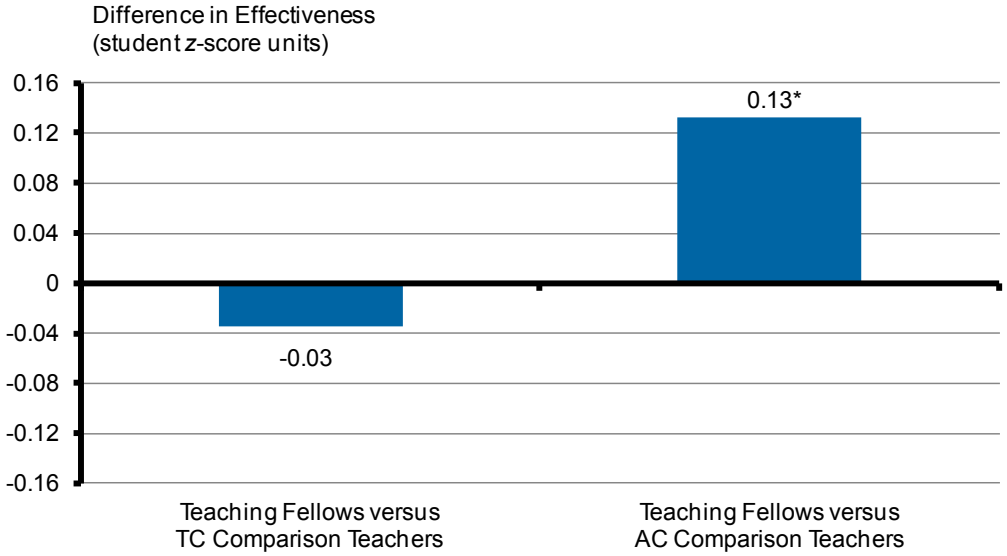
Given that the difference in effectiveness between Teaching Fellows and comparison teachers varied across matches, we examined whether differences were more or less pronounced when specific types of Teaching Fellows and comparison teachers were compared. The subgroup analyses that address this question are discussed in the remainder of this chapter; detailed results are presented in Appendix Table F.6.

B. Impacts of Teaching Fellows Relative to Comparison Teachers from Traditional and Alternative Routes to Certification

We compared the effectiveness of Teaching Fellows with that of comparison teachers in two separate subgroups of matches: (1) those whose comparison teachers were certified through traditional routes to certification (TC routes) and (2) those whose comparison teachers were certified through less selective alternative routes to certification (AC routes). As an AC route, Teaching Fellows programs are subject to the same criticisms mentioned in Chapter V regarding TFA—specifically, that they provide insufficient preparation for teaching relative to TC programs. Therefore, we explored whether the average similarity in effectiveness between Teaching Fellows and comparison teachers masked differences in how Teaching Fellows performed relative to TC and AC teachers separately.

While we found no statistically significant difference in achievement between students of Teaching Fellows and students of comparison teachers from TC routes, students of Teaching Fellows outperformed students of comparison teachers from less selective AC routes by 0.13 standard deviations (Figure VIII.2). This finding has two distinct implications. First, for study schools whose main choice is between Teaching Fellows and TC teachers, math achievement in the classes examined by the study would be no different if they filled vacancies with one type of teacher rather than another. Second, for study schools whose main choice is between Teaching Fellows and AC teachers, relying on Teaching Fellows instead of the AC teachers to teach the secondary math classes in the study would, on average, lead to higher student math achievement.

Figure VIII.2. Differences in Effectiveness Between Teaching Fellows and Comparison Teachers from Traditional and Alternative Routes to Certification



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.2.

*Estimate is statistically significant at the 0.05 level based on a two-tailed test after adjusting for multiple hypothesis testing.
**Estimate is statistically significant at the 0.01 level based on a two-tailed test after adjusting for multiple hypothesis testing.
AC = alternative route to certification; TC = traditional route to certification.

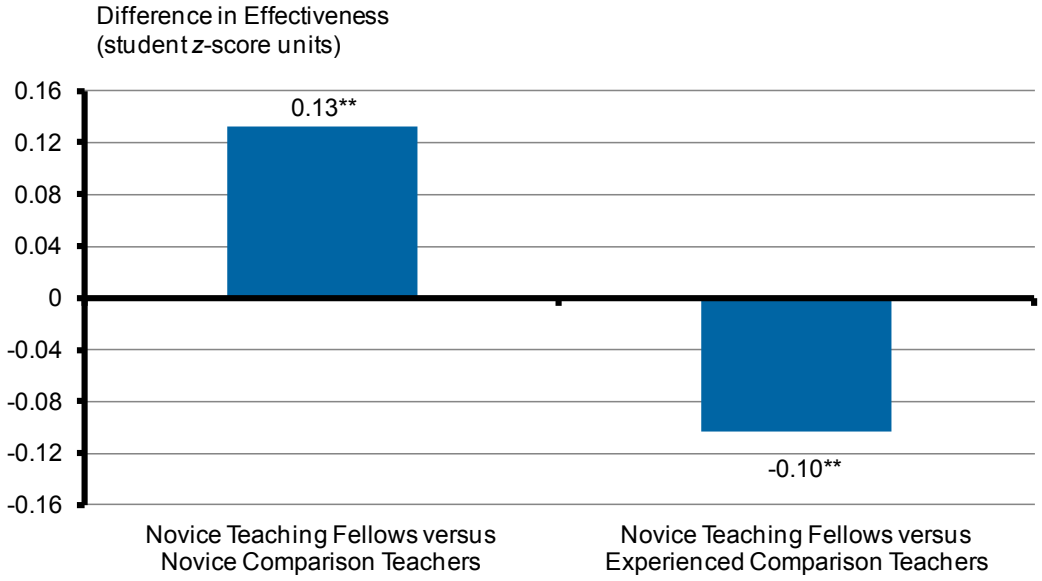
C. Impacts Within Other Teacher Subgroups

1. Subgroups Defined by Specific Levels of Teaching Experience

School administrators might want to know how student achievement in their school would differ as a result of hiring a new Teaching Fellow rather than some other new, inexperienced teacher. We gauged the effects of choosing a new Teaching Fellow over a new teacher from another certification route by comparing Teaching Fellows and comparison teachers who had recently entered teaching. As in Chapter V, we classified all teachers into two experience categories: (1) novice teachers, defined as those in their first three years of teaching, and (2) experienced teachers, defined as those in their fourth year of teaching or beyond. We then examined impacts within the subgroup of classroom matches in which novice Teaching Fellows were compared with novice comparison teachers. We also examined impacts within other experience subgroups to determine how each type of comparison contributed to the overall similarity in effectiveness between Teaching Fellows and comparison teachers.

Among teachers in the study, novice Teaching Fellows demonstrated greater effectiveness than novice comparison teachers. Students of novice Teaching Fellows scored 0.13 standard deviations higher than students of novice comparison teachers (Figure VIII.3).

Figure VIII.3. Differences in Effectiveness Between Novice Teaching Fellows and Comparison Teachers, by Experience Level of Comparison Teacher



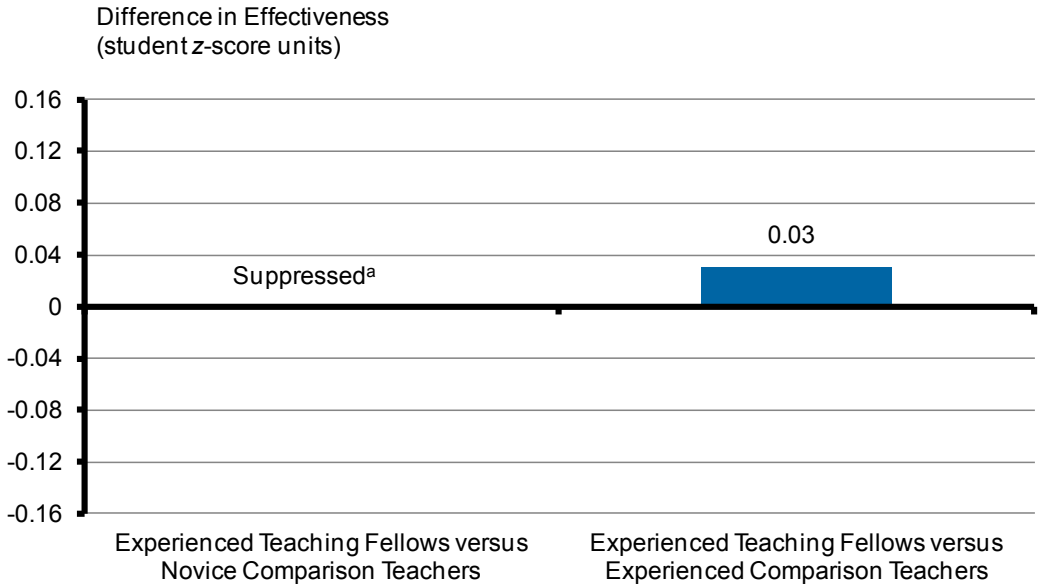
Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.2. Novice teachers are defined as teachers in their first three years of teaching. Experienced teachers are defined as teachers in their fourth or more year of teaching.

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

Novice Teaching Fellows were less effective than experienced comparison teachers. Novice Teaching Fellows lowered student math achievement by 0.10 standard deviations relative to experienced comparison teachers (Figure VIII.3)—a difference that could reflect both the imbalance in experience and other differences in teacher characteristics. As experienced Teaching Fellows were rarely matched (in only 11 classroom matches) with novice comparison teachers, we did not estimate differences in the effectiveness of those teacher groups. In the 65 classroom matches in which experienced Teaching Fellows were compared with experienced comparison teachers, the two groups did not statistically differ in effectiveness (Figure VIII.4).

Figure VIII.4. Differences in Effectiveness Between Experienced Teaching Fellows and Comparison Teachers, by Experience Level of Comparison Teacher



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.2. Novice teachers are defined as teachers in their first three years of teaching. Experienced teachers are defined as teachers in their fourth or more year of teaching.

[a]Due to small sample sizes, this estimate did not meet the minimum level of precision required to present a subgroup estimate in this report.

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

2. Subgroup of Teaching Fellows and Comparison Teachers with Similar Levels of Teaching Experience

As discussed earlier, the average difference in effectiveness between all Teaching Fellows and all comparison teachers includes any difference reflecting the gap in teaching experience between these groups. However, it is possible that the experience gap could narrow over time as the program matures and earlier cohorts of Teaching Fellows have more time to gain experience. We explored this scenario by considering the impacts that Teaching Fellows would have if their experience levels were similar to those of their counterparts in the same schools.

To assess the relative effectiveness of Teaching Fellows and comparison teachers in the absence of an experience gap, we estimated impacts based only on Teaching Fellows in the study who were compared with comparison teachers with similar levels of teaching experience (an absolute difference in experience of no more than two years). Within this subgroup of 46 teachers, the average experience levels of Teaching Fellows (3.6 years) and comparison teachers (3.9 years) were similar.

We found no statistically significant difference in effectiveness between Teaching Fellows and comparison teachers who had similar levels of teaching experience (Table VIII.2). Therefore, our main conclusion from the average impacts presented earlier—that Teaching Fellows and comparison teachers in the study were indistinguishable in effectiveness—also held true for comparisons in which the two groups had equivalent amounts of teaching experience.

Table VIII.2. Difference in Effectiveness Between Teaching Fellows and Comparison Teachers with Similar Levels of Teaching Experience

	Mean Outcome		Difference in Outcomes Between Groups	
Outcome Variable	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Estimate	*p*-Value
Student Math Achievement (*z*-score)	-0.17	-0.20	0.03	0.399
Sample Sizes				
Number of Students	**661**	**622**		
Number of Classroom Matches	**33**	**33**		
Number of Teachers	**22**	**24**		
Number of Schools	**17**	**17**		

Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

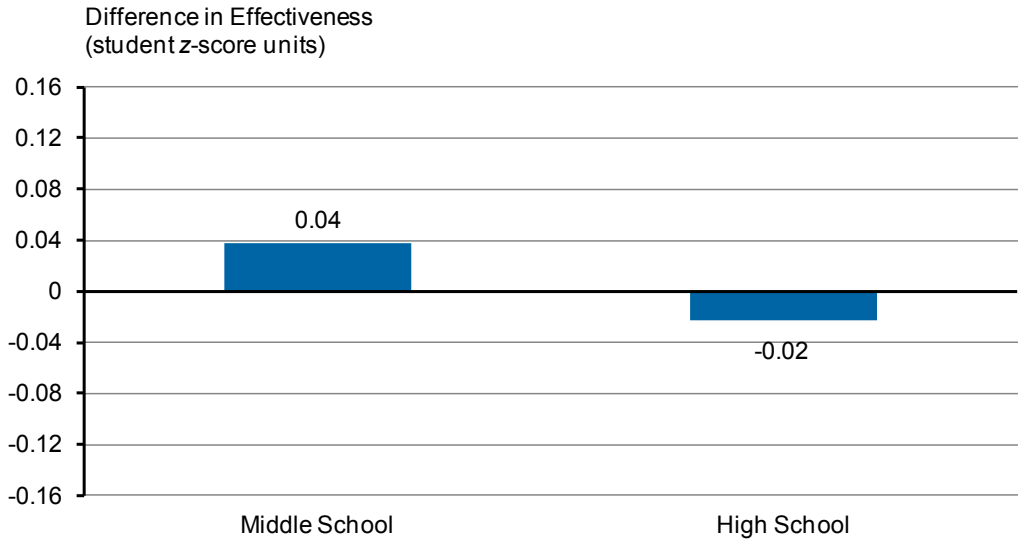
Note: A Teaching Fellow and comparison teacher in the same classroom match are classified as having similar levels of teaching experience if those levels differ in absolute value by no more than two years. The difference in outcomes between groups is adjusted for classroom match fixed effects and all covariates in Appendix Table B.2. The mean outcome for students assigned to Teaching Fellows is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups.

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

3. Subgroups Defined by Middle and High School

We examined differences in effectiveness between Teaching Fellows and comparison teachers separately for middle and high school teachers due to the different types of math courses and student assessments associated with the two grade spans, as discussed in Chapter V. At both the middle and high school levels, the effectiveness of Teaching Fellows was statistically indistinguishable from that of comparison teachers (Figure VIII.5).

Figure VIII.5. Differences in Effectiveness Between Teaching Fellows and Comparison Teachers Within Middle Schools and High Schools



Sources: District administrative records and study-administered Northwest Evaluation Association (NWEA) assessments.

Note: Differences in effectiveness are adjusted for classroom match fixed effects and all covariates in Appendix Table B.2.

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

D. Conclusions

The Teaching Fellows programs aim to recruit and train teachers to teach in low-income school districts. As discussed in Chapter VI, the programs focus their recruitment efforts on both recent college graduates and midcareer professionals and seek applicants who are willing to make a long-term commitment to teaching. The programs are highly selective, with multi-stage application processes, and accept only 13 percent of applicants nationwide. They provide their newly selected teachers with intense five- to seven-week summer training programs as well as mentoring and support after they begin teaching.

Consistent with the Teaching Fellows programs' mission to place teachers in high-need schools, Teaching Fellows in the study were typically placed in urban schools that served economically disadvantaged students, largely from racial and ethnic minority groups. As discussed in Chapter VII, the Teaching Fellows in the study were typically white, with an average of four years of teaching experience and five years of nonteaching work experience, and included roughly equal numbers of men and women. The other secondary math teachers in these same schools were primarily white or black women, with considerably more teaching experience (13 years) and three years of nonteaching work experience, on average. Although the Teaching Fellows teachers took similar numbers of college-level math courses as the teachers from other routes to certification, they were far more likely to have attended a selective undergraduate university and had significantly higher math content knowledge test scores.

The experimental analysis suggests that Teaching Fellows teaching secondary math were neither more nor less effective than comparison teachers. However, this comparison of average effectiveness masked several differences in effectiveness between specific subgroups of teachers:

- **Teaching Fellows were more effective than comparison teachers from less selective AC routes.** Students of Teaching Fellows outperformed students of comparison teachers from less selective AC routes by 0.13 standard deviations.
- **Teaching Fellows were neither more nor less effective than comparison teachers from TC routes.** Students of Teaching Fellows did not perform significantly better or worse than students of comparison teachers from TC routes.
- **Novice Teaching Fellows were more effective than novice comparison teachers.** Students of novice Teaching Fellows (those in their first three years of teaching) outscored students of novice comparison teachers by 0.13 standard deviations. However, on average across all levels of experience, Teaching Fellows were neither more nor less effective than comparison teachers who had similar levels of teaching experience.
- **Teaching Fellows were neither more nor less effective than comparison teachers in both middle and high schools.** Math achievement did not differ significantly between students of Teaching Fellows and students of comparison teachers in either middle or high schools.

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

IX. RELATIONSHIPS BETWEEN TEACHER CHARACTERISTICS AND TEACHER EFFECTIVENESS

When making hiring decisions, school administrators take into account a variety of applicant characteristics other than route to certification that provide information on applicants' likely effectiveness in the classroom. Using nonexperimental analyses, this chapter explores the extent to which various teacher characteristics—including measures of the teachers' academic ability, math knowledge, professional experience, and instructional training—are predictive of teacher effectiveness. This evidence can, in turn, suggest the degree to which differences in effectiveness between teachers from distinct certification routes are related to differences in the types of teachers who were recruited and selected into those routes versus differences in the training and support they received. The findings presented here are correlational relationships between teacher characteristics and student achievement; as such, they can generate hypotheses about—but do not provide conclusive evidence for—the causal effects of these characteristics.

Because each teacher's effectiveness is defined as his or her contributions to student achievement relative to other teachers, our analyses are focused on examining *differences* in effectiveness across teachers and identifying how they are related to differences in the teachers' characteristics. In previous chapters, we found wide variation across classroom matches in the relative effectiveness of the teachers who were compared even after taking into account the teachers' certification routes (see Figures V.1 and VIII.1). This variation suggests that there were other teacher characteristics that were correlated with effectiveness among teachers in the study.

In the remainder of this chapter, we identify characteristics associated with variation in teacher effectiveness (Section A), using the combined sample of all Teach For America (TFA) teachers, Teaching Fellows, and comparison teachers in the study. We then explore whether these characteristics could account for the observed difference in effectiveness between TFA and comparison teachers (Section B). Because Teaching Fellows and comparison teachers did not differ in effectiveness on average, we did not seek to decompose differences in effectiveness between these groups.

A. Analysis of Characteristics That Predict Teacher Effectiveness

We explored the relationships between teacher characteristics and teacher effectiveness by including teacher characteristics in the model of student math achievement used in the experimental analysis from Chapters V and VIII. Because the model controlled for classroom match indicators, only differences in characteristics between teachers in the same match were used to predict differences in achievement between students of matched classes. Achievement differences between matched classes were solely driven by differences in teacher effectiveness, so this regression model was, in effect, an analysis of the association between teacher characteristics and teacher effectiveness.

This analysis examined three broad categories of teacher characteristics: (1) general academic ability; (2) exposure to and knowledge of mathematics; and (3) instructional training and experience. Variables that measured teacher characteristics in the main analysis, along with their sample means and standard deviations, are listed in Appendix Table G.1 (see also Table II.7).

We estimated the association between each teacher characteristic and student math achievement both without and with controls for the other teacher characteristics in the analysis. While findings that were consistent across these two estimation methods were considered to be more reliable, we used findings from models that controlled for other teacher characteristics to assess and interpret the magnitudes of the estimated relationships. From the estimation models, each regression coefficient on a teacher characteristic can be interpreted as the difference in student math achievement that would be predicted from a specified difference in that characteristic between two teachers. Table IX.1 shows these regression coefficients, expressed as predicted differences in achievement between the students of two illustrative teachers; complete results from these models are provided in Appendix Tables G.2 and G.3.

We found that two types of factors were consistently associated with teacher effectiveness, with suggestive evidence for a third factor (Table IX.1):

1. **Teacher effectiveness increased with teacher experience.** Students assigned to a second-year teacher were predicted to score 0.08 standard deviations higher on math assessments than students assigned to a first-year teacher. Among teachers with at least five years of teaching experience, each additional year of teaching experience was associated with an increase of 0.005 standard deviations in student achievement.
2. **Teacher effectiveness declined with increasing amounts of teacher coursework during the school year.** For each additional 10 hours of coursework that teachers took during the school year, the math achievement of their students was predicted to drop by 0.002 standard deviations. These findings imply that a teacher who took an average amount of coursework during the school year, whether for initial certification or any other certification or degree, decreased student math achievement by 0.04 standard deviations relative to a teacher who took no coursework while teaching.
3. **Some evidence suggests that greater math content knowledge was positively associated with teacher effectiveness in high schools, but not in middle schools.** In high schools, we found that teachers' math content knowledge, as measured by their scores on the Praxis II math assessments, was positively associated with student achievement in some, but not all, analyses. Our main analysis indicates that a one standard deviation increase in a high school teacher's Praxis II score—equivalent to rising from the 50th to the 84th percentile of Praxis II scores—was associated with an increase in student math achievement of 0.04 standard deviations (p-value = 0.051). In middle schools, we found no association between teachers' scores on the Praxis II and student achievement.

No other characteristics in the analysis predicted teacher effectiveness to a statistically significant degree.

We discuss next, in greater detail, the specific characteristics that were and were not predictive of teacher effectiveness.

1. General Academic Ability

We measured teachers' general academic ability based on the selectivity of the college or university from which they received their bachelor's degree. There was no statistically

significant difference in effectiveness between teachers from selective colleges or universities and those from all other educational institutions (Table IX.1). Likewise, in sensitivity analyses, we found that teachers from *highly* selective colleges or universities did not differ in effectiveness from teachers whose colleges or universities had lower levels of selectivity (Appendix Table G.5).

Table IX.1. Differences in Student Math Achievement Associated with Differences in Teacher Characteristics

	Without Controlling for Other Teacher Characteristics		Controlling for Other Teacher Characteristics	
Difference in Teacher Characteristic	Predicted Difference in Student Math Achievement (student *z*-score units)	*P*-value	Predicted Difference in Student Math Achievement (student *z*-score units)	*P*-value
Teacher Who Graduated from Selective College or University, Compared with Teacher Who Graduated from Other Institution[a]	0.017	0.543	0.003	0.913
Teacher Who Took More than the Median Number of College-Level Math Courses, Compared with Teacher Who Did Not[b]	-0.022	0.381	-0.027	0.281
Teacher Who Used College-Level Math in a Nonteaching Job, Compared with Teacher Who Did Not	-0.045	0.134	-0.038	0.190
Teacher Whose Score on the Specified Praxis II Test Was Higher by One *z*-Score Unit Than the Score of Another Teacher				
Praxis II Test in Math Content Knowledge	0.017	0.293	0.035	0.051
Praxis II Test in Middle School Math	0.002	0.907	-0.001	0.955
Teacher Who Had More than the Median Number of Hours of Math Pedagogy Instruction During Training, Compared with Teacher Who Did Not[c]	-0.011	0.668	-0.014	0.592
Teacher Who Had More than the Median Number of Days of Student Teaching in Math During Training, Compared with Teacher Who Did Not[d]	-0.001	0.977	-0.003	0.895
Teacher Who Took 10 More Total Hours of Coursework During the School Year than Another Teacher	-0.002*	0.028	-0.002*	0.041
Teacher Who Had One More Year of Teaching Experience Compared with a Teacher in the:				
First Year of Teaching	0.069*	0.038	0.080*	0.016
Second to Fourth Years of Teaching	0.000	0.991	-0.004	0.755
Fifth Year of Teaching or Beyond	0.004*	0.040	0.005*	0.017

Sources: District administrative records, study-administered Northwest Evaluation Association (NWEA) assessments, and Survey of Secondary Math Teachers.

Note: The table shows coefficients on teacher characteristics and corresponding *p*-values from regression models in which student end-of-year math scores are the dependent variable and the set of independent variables consists of an indicator for TFA teachers, an indicator for Teaching Fellows, one or more teacher characteristics, classroom match fixed effects, and all covariates in Appendix Table B.1. Coefficients in the first column of results come from separate models, each of which includes only a single teacher characteristic (or a group of variables representing a single teacher characteristic). Coefficients in the third column of results come from a single model that includes all displayed teacher characteristics simultaneously. Complete model results, including *p*-values, are in Appendix Tables G.2 and G.3.

Table IX.1. (continued)

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

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

2. Exposure to and Knowledge of Math

We did not find strong support for the notion that teachers' familiarity with math—as reflected in the amount of math coursework completed, experience in a math-related nonteaching job, or scores on math knowledge assessments—was related to their effectiveness at teaching math. Across a variety of measures of teachers' exposure to and knowledge of math, no measure predicted teacher effectiveness in a manner that consistently reached statistical significance in multiple estimation approaches.

We found no statistically significant relationships between teacher effectiveness and the amount of college-level math coursework completed. Teachers who had completed more than the median number of math courses—7.5 courses—were statistically indistinguishable in their effectiveness from teachers who had completed less than the median number of courses (Table IX.1). Findings were similar when we measured exposure to math coursework on the basis of teachers' completion of minors, majors, or advanced degrees in math-related subjects (Appendix Table G.6).

The use of math in a nonteaching job is another type of experience through which an individual can acquire greater familiarity with math and its applications. However, among teachers in the study, this type of experience was not associated with effectiveness in the classroom; there was no statistically significant difference in effectiveness between teachers who had used college-level math in a prior nonteaching job and those who had not (Table IX.1).

The most direct measure of the study teachers' knowledge of math was their score from the Praxis II assessment in either Middle School Mathematics (taken primarily by middle school teachers in the study) or Mathematics Content Knowledge (taken primarily by high school teachers in the study). At the middle school level, we found no relationship between teachers' Praxis scores and their effectiveness. At the high school level, we found suggestive evidence that students' achievement was higher when their teachers performed better on the Mathematics Content Knowledge assessment. Our main analysis indicates that a one standard deviation increase in a high school teacher's Praxis II score—equivalent to rising from the 50th to the 84th percentile of Praxis II scores—was associated with an increase in student math achievement of 0.04 standard deviations (Table IX.1; p-value = 0.051 in a model controlling for other teacher characteristics). Across a variety of sensitivity analyses (see Appendix G), the relationship between Praxis scores and high school teachers' effectiveness was statistically significant in some, but not all, analyses. In particular, high school teachers with Praxis II scores in the top half of the score distribution were more effective than those with scores in the bottom half by a statistically significant margin (Appendix Table G.7). Taken together, these findings suggest the possibility that high school math teachers with higher Praxis scores were more effective, but the findings are sensitive to the way in which the analysis is specified.

3. Instructional Training and Experience

Both the formal training that teachers receive and the experience they acquire on the job have the potential to shape their instructional skills. We found no evidence that the measured features of formal training were associated with greater teacher effectiveness; one feature of training—coursework during the school year—predicted *lower* teacher effectiveness. In contrast, various measures of on-the-job teaching experience were positively associated with teacher effectiveness.

Two specific elements of formal training—instruction in math pedagogy and student teaching in math—had no statistically significant relationship with teacher effectiveness (Table IX.1). These results should be interpreted cautiously, as teachers for whom more time had elapsed since their training might have had more difficulty in recollecting the amount of training they received.

Enrollment in coursework—not necessarily limited to math pedagogy—is a broad form of training in which teachers might participate for a variety of reasons, such as to obtain full certification, maintain current certification, or obtain an advanced degree. We found that teachers who took more coursework during the school year were *less* effective by a statistically significant margin. The estimated relationship implied that, for each additional 10 hours of coursework that teachers took during the school year, the math achievement of their students was predicted to drop by 0.002 standard deviations (Table IX.1). Therefore, a teacher who took the average amount of coursework among teachers who took any coursework at all during the school year—about 219 hours—was predicted to lower student math achievement by 0.04 standard deviations relative to a teacher who took no concurrent coursework. The direction of this estimated relationship is consistent with prior correlational evidence for a negative association between coursework taken during the school year and teacher effectiveness (Constantine et al. 2009).

Beyond formal training, on-the-job teaching experience represents a different means of developing instructional skills: learning by doing. We found evidence that additional teaching experience predicted greater effectiveness among two groups of teachers: (1) those who had just started their teaching career, and (2) those who had already accumulated five or more years of teaching experience. First, we considered teachers in their first year of teaching and assessed the extent to which teachers with an additional year of experience—second-year teachers—were predicted to have greater effectiveness. Controlling for other teacher characteristics, students assigned to a second-year teacher were predicted to score 0.08 standard deviations higher on math assessments than students assigned to a first-year teacher (Table IX.1). Second, among teachers in their second, third, or fourth year of teaching, an additional year of experience was associated with neither greater nor lower effectiveness. In other words, a teacher with four years of experience was predicted to be no more or less effective than a teacher with three years of experience. Third, among teachers with at least five years of teaching experience, those with more experience were more effective after controlling for other teacher characteristics; each additional year of teaching experience was associated with an increase of 0.005 standard deviations in student achievement.

To interpret properly the estimated relationships between teacher experience and effectiveness, it is important to recognize that these estimates are based on differences in effectiveness across different teachers—not the evolution of effectiveness for the same teachers

over time. Comparisons of effectiveness between more- and less-experienced teachers capture both the true effects of experience and unobserved differences between the broader group of people who enter teaching and the self-selected group that remains in teaching.

Nevertheless, our estimates accord with the general pattern of experience effects found by prior studies that have used longitudinal data on teachers and students. In particular, our findings are consistent with previous evidence that the largest gain in effectiveness from experience occurs between the first and second years of teaching (Hanushek et al. 2005; Boyd et al. 2006; Kane et al. 2008). Beyond the second year of teaching, gains in effectiveness from additional experience have not been consistently found by previous research. Compared with their own effectiveness in their second year of teaching, teachers in their fifth year or beyond of teaching were no different in effectiveness based on data from an anonymous Texas district (Hanushek et al. 2005) but were more effective based on data from New York City (Kane et al. 2008). Our findings on gains in effectiveness from additional experience beyond the second year of teaching—which showed gains at some experience levels but not at others—share the inconsistency present in previous research.

B. Accounting for the Impact of TFA Teachers

Given that some teacher characteristics were associated with effectiveness, differences in effectiveness between TFA and comparison teachers could potentially be accounted for—that is, predicted—by differences in those characteristics. In our experimental analysis, we found that TFA teachers in the study were, on average, more effective than comparison teachers (Chapter V). Using our estimates for the relationship between teacher characteristics and effectiveness, this section examines the extent to which differences in characteristics between TFA and comparison teachers can account for their difference in effectiveness. Given that this analysis relies on the nonexperimental estimates from Section A, it can produce suggestive, but not conclusive, evidence for the reasons TFA teachers were more effective than comparison teachers. Because we found no average difference in effectiveness between Teaching Fellows and comparison teachers (Chapter VIII), we did not conduct a similar analysis for these teacher groups.

As discussed in Chapter II, two factors were necessary for a characteristic to account for part of the difference in effectiveness between TFA and comparison teachers: (1) the characteristic was associated with teacher effectiveness, and (2) TFA and comparison teachers differed on this characteristic. Based on the magnitude and direction of these two factors, we calculated a predicted difference in effectiveness between TFA and comparison teachers and compared it to the groups' actual difference in effectiveness from Chapter V.

For each characteristic in the analysis, Table IX.2 shows the difference in effectiveness between TFA and comparison teachers, in student *z*-score units, that would have been predicted based only on that characteristic; Appendix Table G.4 contains detailed findings. Although we show results for all characteristics, we primarily considered the characteristics that had a statistically significant association with effectiveness from the findings in Section A of this chapter.

Our central finding is that the observed characteristics, when considered collectively, did not account for any positive portion of the difference in effectiveness between TFA and comparison teachers. On the basis of these characteristics, TFA teachers would have been predicted to be less

effective than comparison teachers, suggesting that these characteristics could not account for why TFA teachers were, in fact, more effective than their counterparts. We discuss these findings in greater detail next.

Table IX.2. Predicted Differences in Effectiveness Between TFA Teachers and Comparison Teachers Based on Observed Teacher Characteristics

Teacher Characteristic	Predicted Difference in Effectiveness Between TFA Teachers and Comparison Teachers Based on the Specified Characteristic (student *z*-score units)
Graduated from Selective College or University[a]	0.001
Number of College-Level Math Courses Taken Is Above Median[b]	0.006
Used College-Level Math in a Nonteaching Job	0.011
Score on Praxis II Test in Math Content Knowledge	0.012
Score on Praxis II Test in Middle School Math	-0.001
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	0.000
Number of Days of Student Teaching in Math During Training Is Above Median[d]	0.000
Hours of Coursework During the Study Year (divided by 10)	-0.005
Has More than One Year of Teaching Experience	-0.026
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	0.010
Number of Additional Years of Teaching Experience Beyond Five Total Years	-0.036
Total Predicted Difference in Effectiveness Between TFA Teachers and Comparison Teachers Based on All Measured Characteristics	**-0.028**
Total Observed Difference in Effectiveness Between TFA Teachers and Comparison Teachers	**0.075**

Sources: District administrative records, study-administered Northwest Evaluation Association (NWEA) assessments, and Survey of Secondary Math Teachers.

Note: For a given characteristic, the entry in the final column is equal to the product of two estimates: (1) the estimated association between the characteristic and student math achievement after controlling for other teacher characteristics and (2) the average difference in the characteristic between TFA teachers and comparison teachers. Complete results are in Appendix Table G.4.

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

TFA = Teach For America.

1. Predictions Based on Each Individual Characteristic

First, we considered whether each characteristic separately could have predicted that TFA teachers would be more effective than comparison teachers. A characteristic could account for part of the positive impact of TFA teachers relative to comparison teachers only under one of two scenarios:

1. If the characteristic had a *positive* association with teacher effectiveness based on the analyses in Section A, then TFA teachers needed to show *more* of the characteristic than comparison teachers.
2. If the characteristic had a *negative* association with teacher effectiveness based on the analyses in Section A, then TFA teachers needed to show *less* of the characteristic than comparison teachers.

The analyses in Section A identified three characteristics for which there was either tentative or consistent evidence of an association with teacher effectiveness: (1) teachers' scores on the Praxis II Mathematics Content Knowledge assessment, (2) amount of coursework taken during the school year, and (3) teaching experience. For each of the three characteristics, we examined how TFA and comparison teachers differed on the characteristic to calculate a predicted difference in effectiveness based on that characteristic alone.

Math content knowledge. Scores on the Praxis II Mathematics Content Knowledge assessment accounted for a positive portion of the difference in effectiveness between TFA and comparison teachers. This is because (1) these Praxis scores were positively associated with teacher effectiveness (Table IX.1) and (2) TFA teachers had higher Praxis scores than comparison teachers (see Chapter IV). Based on this characteristic alone, students of TFA teachers would be predicted to outscore students of comparison teachers by 0.012 standard deviations (Table IX.2). In comparison, the actual, experimentally estimated difference in student math achievement between these groups was 0.075 standard deviations. Therefore, our estimates suggest that the difference in Praxis performance between TFA and comparison teachers accounted for 16 percent of their actual difference in effectiveness.

Coursework during the school year. Patterns of coursework during the school year did not predict a positive impact of TFA teachers relative to comparison teachers (Table IX.2). Although findings from Table IX.1 indicate that coursework during the school year was associated with *lower* effectiveness, TFA teachers took *more* coursework during the school year. Based on these patterns, TFA teachers would have been predicted to be less effective than comparison teachers, when in fact they were more effective.

Teaching experience. On the basis of teaching experience alone, TFA teachers would have been predicted to be less effective than comparison teachers because (1) experience was positively associated with effectiveness and (2) TFA teachers had relatively less experience. Earlier, our findings had indicated that teachers who acquired a second year of teaching experience were more effective than those who had not yet done so (Table IX.1). TFA teachers in our study were less likely than comparison teachers to have acquired a second year of teaching experience. This gap implied that TFA teachers would be predicted to lower student math achievement by 0.026 standard deviations relative to comparison teachers (Table IX.2). Likewise, TFA teachers had, on average, fewer additional years of teaching experience beyond the fifth year, implying a further predicted drop of 0.036 standard deviations in their students' math achievement relative to the students of comparison teachers (Table IX.2). In short, patterns of teaching experience predicted a negative impact of TFA teachers relative to comparison teachers—the opposite of what actually occurred.

2. Predictions Based on All Observed Characteristics

Overall, the observed characteristics could not account for why TFA teachers were actually more effective than comparison teachers. TFA teachers' lower experience levels suggested that they would be less effective than comparison teachers to an extent that would more than offset the other observed characteristics, such as Praxis scores, on which they had an advantage. On net, based on all teacher characteristics in the analysis, students of TFA teachers would have been predicted to score 0.028 standard deviations lower than students of comparison teachers. In fact, students of TFA teachers actually scored 0.075 standard deviations higher than students of comparison teachers.

A potential reason why the characteristics in the analysis did not predict TFA teachers' impact relative to comparison teachers is that the attributes enabling TFA teachers to be more effective might have been different than the ones we measured in the study. For instance, it is possible that TFA's intensive method of screening and evaluating applicants effectively identifies attributes that predict teachers' effectiveness in the classroom but that are difficult for outside researchers to measure. However, given the wide variation in effectiveness among teachers from the same certification route, there are likely to be other teacher characteristics—beyond those considered during the selection of applicants into each route—that are also associated with effectiveness in the classroom and that we did not observe in our study.

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

REFERENCES

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

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

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.

Barron's Profiles of American Colleges 2003. 25th ed. New York: Barron's Educational Series, 2002.

Benjamini, Yoav, and Yosef Hochberg. "Controlling for the False Discovery Rate: A Practical and Powerful Approach to Multiple Hypothesis Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, 1995, pp. 289–300.

Bloom, Howard S., Lashawn Richburg-Hayes, and Alison Rebeck Black. "Using Covariates to Improve Precision for Studies that Randomize Schools to Evaluate Educational Interventions." *Educational Evaluation and Policy Analysis,* vol. 29, no. 1, March 2007, pp. 30–59.

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.

Carroll, Stephen, Robert Reichardt, and Cassandra Guarino. "The Distribution of Teachers Among California's Districts and Schools." Santa Monica, CA: RAND, October 2000.

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

Clark, Melissa A., Eric Isenberg, and Marykate Zukiewicz. "The Evaluation of the Teach For America Investing in Innovation (i3) Scale–Up: Design Report." Princeton, NJ: Mathematica Policy Research, 2012.

Clark, Melissa A., Daniel W. Player, Alison Wellington, and Sheena McConnell. "An Evaluation of Secondary Math Teachers from Two Highly Selective Alternative Routes to Certification: Design Report." Princeton, NJ: Mathematica Policy Research, December 22, 2009.

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

Clotfelter, Charles, Helen F. Ladd, Jacob Vigdor, and Justin Wheeler. "High-Poverty Schools and the Distribution of Teachers and Principals." *North Carolina Law Review*, vol. 85, 2007, pp. 1345-79.

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

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

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

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

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

Donald, Stephen G., and Kevin Lang. "Inference with Difference-in-Differences and Other Panel Data." *Review of Economics and Statistics*, vol. 89, 2007, pp. 221–233.

Donaldson, Morgaen L., and Susan Moore Johnson. "Teach For America Teachers: How Long Do They Teach? Why Do They Leave?" *Phi Delta Kappan,* vol. 93, no. 2, 2011, pp. 47–51.

Editorial Projects in Education. "Education Counts." Available at http://www.edweek.org/rc/2007/06/07/edcounts.html. Accessed July 18, 2012.

Feistritzer, C.E. "State Policy Trends for Alternative Routes to Teacher Certification: A Moving Target." Presented at the Conference on Alternative Certification: A Forum for Highlighting Rigorous Research, hosted by Learning Point Associates and supported by the Institute of Education Sciences, Washington, DC, 2005.

Gaskill, Peggy E. "Progress in the Certification of Middle Level Personnel." *Middle School Journal*, vol. 33, no. 5, 2002, pp. 33–40.

Gitomer, Drew H., and Yi Qi. "Recent Trends in Mean Scores and Characteristics of Test-Takers on Praxis II Licensure Tests." Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, Policy and Program Studies Service, April 2010.

Glazerman, Steven, and Jeffrey Max. "Do Low-Income Students Have Equal Access to the Highest-Performing Teachers?" NCEE Evaluation Brief. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Document No. PP11-23a, 2011.

Hanushek, Eric A., and Steven Rivkin. "Teacher Quality." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Finis Welch. Vol. 2. Amsterdam: North-Holland, 2006.

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

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

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

Hill, Heather C., Brian Rowan, and Deborah Loewenberg Ball. "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement." *American Educational Research Journal*, vol. 42, no. 2, Summer 2005, pp. 371-406.

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

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

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

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

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis." *Educational Evaluation and Policy Analysis*, vol. 24, no. 1, 2002, pp. 37-62.

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

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

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

National Center for Education Statistics. "Statistical Standards." U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics. Washington DC: U.S. Government Printing Office, 2000. Available at [http://nces.ed.gov/statprog/2002/std4_2.asp]. Accessed July 9, 2013.

Provasnik, Stephen, Patrick Gonzales, and David Miller. *U.S. Performance Across International Assessments of Student Achievement: Special Supplement to The Condition of Education 2009*. Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, 2009.

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.

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

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

Royston, Patrick. "Multiple Imputation of Missing Values: Update of ice." *Stata Journal*, vol. 5, 2005, pp. 527–536.

Royston, Patrick. "Multiple Imputation of Missing Values." *Stata Journal*, vol. 4, 2004, pp. 227–241.

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

Samejima, Fumiko. "A Use of the Information Function in Tailored Testing." *Applied Psychological Measurement*, vol. 1, no. 3, 1977, pp. 233–247.

Samejima, Fumiko. "Estimation of Reliability Coefficients Using the Test Information Function and its Modifications." *Applied Psychological Measurement*, vol. 18, no. 3, 1994, pp. 229-244.

Sass, Tim, Jane Hannaway, Zeyu Xu, David Figlio, and Li Feng. "Value Added of Teachers in High-Poverty Schools and Lower-Poverty Schools." CALDER Working Paper no. 52, National Center for Analysis of Longitudinal Data in Education Research. Washington, DC: Urban Institute, 2010.

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

U.S. Department of Education, Office of Elementary and Secondary Education. *Highly Qualified Teachers / Improving Teacher Quality State Grants: ESEA, Title II, Part A: Non-Regulatory Guidance*. August 3, 2005.

U.S. Department of Education, Office of Postsecondary Education. *Preparing and Credentialing the Nation's Teachers: The Secretary's Eighth Report on Teacher Quality Based on Data Provided for 2008, 2009, and 2010.* Washington, DC, 2011.

U.S. Department of Education, Office of Postsecondary Education, Office of Policy Planning and Innovation. *Meeting the Highly Qualified Teachers Challenge: The Secretary's Annual Report on Teacher Quality.* Washington, DC, 2002.

van Buuren, S., H.C. Boshuizen, and D.L. Knook. "Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis. *Statistics in Medicine*, vol. 18, 1999, pp. 681–694.

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

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

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

APPENDIX A

SUPPLEMENTARY TECHNICAL INFORMATION ON STUDY DESIGN AND DATA COLLECTION

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

In Appendix A, we provide additional information on the study's experimental design, random assignment, district and school recruiting, and data collection and response rates, as a supplement to the information presented in Chapter II.

A. Experimental Design

As discussed in Chapter II, the study's experimental design relied on the random assignment of students within schools to a math class taught by a Teach For America (TFA) or Teaching Fellows teacher or to a comparable class taught by a comparison teacher. The classes between which students were randomly assigned are referred to as a *classroom match*. For classes to match and be eligible for the study, the classes must have covered an eligible math course and been fully comparable, random assignment must not have interfered with the school's scheduling process, and any supplemental math instruction must not have interfered with the study design, as discussed further below.

1. Eligible Math Courses

All middle school math courses were eligible, as were general high school math (such as pre-algebra or remedial math), Algebra I, Geometry, and Algebra II. The study did not include more advanced courses, such as trigonometry or calculus, because appropriate assessments for these subjects were not available.

2. Comparability of Classes

To be eligible for inclusion in a classroom match, classes must have been taught at the same level (for instance, honors or regular). In addition, they must have been taught under similar circumstances, including similar class sizes, same roles for teacher's aides, and same arrangements for inclusion of English language learners and special education students. These restrictions helped to ensure that any differences in student achievement between classes in a match could be attributed to the effectiveness of the teachers rather than to other differences between the classes.

3. No Interference with Scheduling

Secondary school scheduling is a complex process, typically beginning in the middle of the preceding school year and encompassing multiple iterations to determine the number and schedule of courses to be offered, as well as the schedules of individual students. To avoid disrupting the school's normal scheduling process, we allowed matches only if either (1) all classrooms in the match were taught during the same period of the school day or (2) students were assigned to groups with which they took all their classes (sometimes called teams, families, academies, pathways, or schools within schools) and the TFA/Teaching Fellows and comparison teachers in the match taught in separate groups. In the latter case, after we randomly assigned students to groups (and by association, teachers), we allowed schools to determine which period during the school day each student attended math class. The overwhelming majority (95 percent) of students in the study were assigned to concurrent classes, and the remaining 5 percent were assigned to teams.

4. Supplemental Math Did Not Interfere with Study Design

In many schools, students attended supplemental math classes aimed at reinforcing the instruction given in the main math classes that constituted the study sample. In 10 percent of classroom matches in the study, all students enrolled in the study classes were also required to take supplemental math classes; in another 23 percent of matches, specific subsets of students—usually those with low achievement in the prior year—were assigned to supplemental math instruction (Table A.1).

Table A.1. Percentage of Classroom Matches in Which Students Receive Supplemental Math

	Percentage of Matches
No Students Receive Supplemental Math	53
Some Students Receive Some Supplemental Math	23
Regular and supplemental math teacher are of the same type	3
Regular and supplemental math teacher are not necessarily of the same type	21
All Students Receive Supplemental Math	10
Regular and supplemental math teacher are of the same type	2
Regular and supplemental math teacher are not necessarily of the same type	8
No Information on Supplemental Math Available	14
Total Number of Classroom Matches	**228**

Note: Teacher types are Teach For America teachers, Teaching Fellows, and comparison teachers.

Supplemental math instruction had the potential to either reinforce or dilute the contrast between matched classes in the types of teachers to whom students were exposed. The contrast was reinforced within matches in which students received all supplemental math instruction from the same type of teacher (TFA, Teaching Fellow, or comparison) who taught their regular math classes; this arrangement occurred in 5 percent of matches. In the remaining 29 percent of matches with students enrolled in supplemental math instruction, the contrast from the matched classes was diluted because students' supplemental instruction was not necessarily delivered by the same type of teacher who delivered their regular instruction. Examples of the latter scenario included arrangements in which a single math teacher was responsible for all supplemental math instruction, as well as arrangements in which schools had assigned students to supplemental classes before the study team randomly assigned them to the study classes. Notably, in these scenarios, the mix of teachers from whom students received supplemental instruction was still equivalent between the matched study classes. Therefore, any differences in achievement between matched classes were the result of differences in effectiveness between the teachers of those classes.

Before random assignment, we excluded from the study any matches in which arrangements for supplemental math instruction would have prevented us from attributing differences in achievement between matched classes to differences in the types of teachers responsible for those classes. Accordingly, we did not admit matches in which students received all supplemental math instruction from the opposite type of teacher as the one responsible for their regular classes. We also did not admit matches in which assignment to supplemental classes was determined after random assignment (for instance, if lower-performing students were assigned to

a supplemental class midyear), because assignment to supplemental instruction after random assignment could have been correlated with the teacher to whom the students were randomly assigned.

B. Recruitment of Districts, Schools, and Classroom Matches

As discussed in Chapter II, we focused recruitment efforts on districts with large concentrations of secondary math teachers from TFA or a Teaching Fellows program. Figure A.1 illustrates the recruitment of districts and schools into the sample.

Within the 15 districts that agreed to participate in the study, we conducted random assignment for at least one classroom match in each of 109 schools, with a total of 321 matches. Eighty-two of these 109 schools (75 percent, comprising 228 classroom matches) properly implemented random assignment, maintained viable classroom matches, and cooperated with data collection activities—these schools and matches formed the study's research sample. The remaining 27 schools (25 percent) were dropped from the study sample. Twenty-one of these schools (comprising 67 matches) 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 refused to make the requested changes. An additional 6 schools (26 matches) were dropped after random assignment because: (1) students needed to be completely reassigned during the school year (for instance, the school had failed to inform us that it had to reassign all students to different classes at the end of the first semester); (2) a teacher we thought was from TFA or a Teaching Fellows program had actually entered through some other route; (3) classrooms were consolidated due to lower than expected enrollments; (4) a study teacher left the school for administrative reasons unrelated to performance; or (5) the school was unwilling to distribute parental consent forms.

Recruitment of the study sample occurred prior to each of the two school years in which the study was conducted. Table A.2 shows the number of sample members in each year of the study.

In both the TFA and Teaching Fellows study samples, the vast majority of classroom matches consisted of one class taught by a TFA or Teaching Fellows teacher and one class taught by a comparison teacher. In the TFA study sample, 103 of the 111 matches (93 percent) consisted of one class taught by a TFA teacher and one taught by a comparison teacher (Table A.3). The remaining matches included additional classes, taught by either TFA or comparison teachers, with the largest matches including six classes taught by TFA teachers and six taught by comparison teachers. In the Teaching Fellows study sample, 98 of the 118 matches (83 percent) consisted of one class taught by a Teaching Fellow and one taught by a comparison teacher (Table A.4). The largest matches included four Teaching Fellow classes and two comparison classes.

Figure A.1. District and School Recruiting

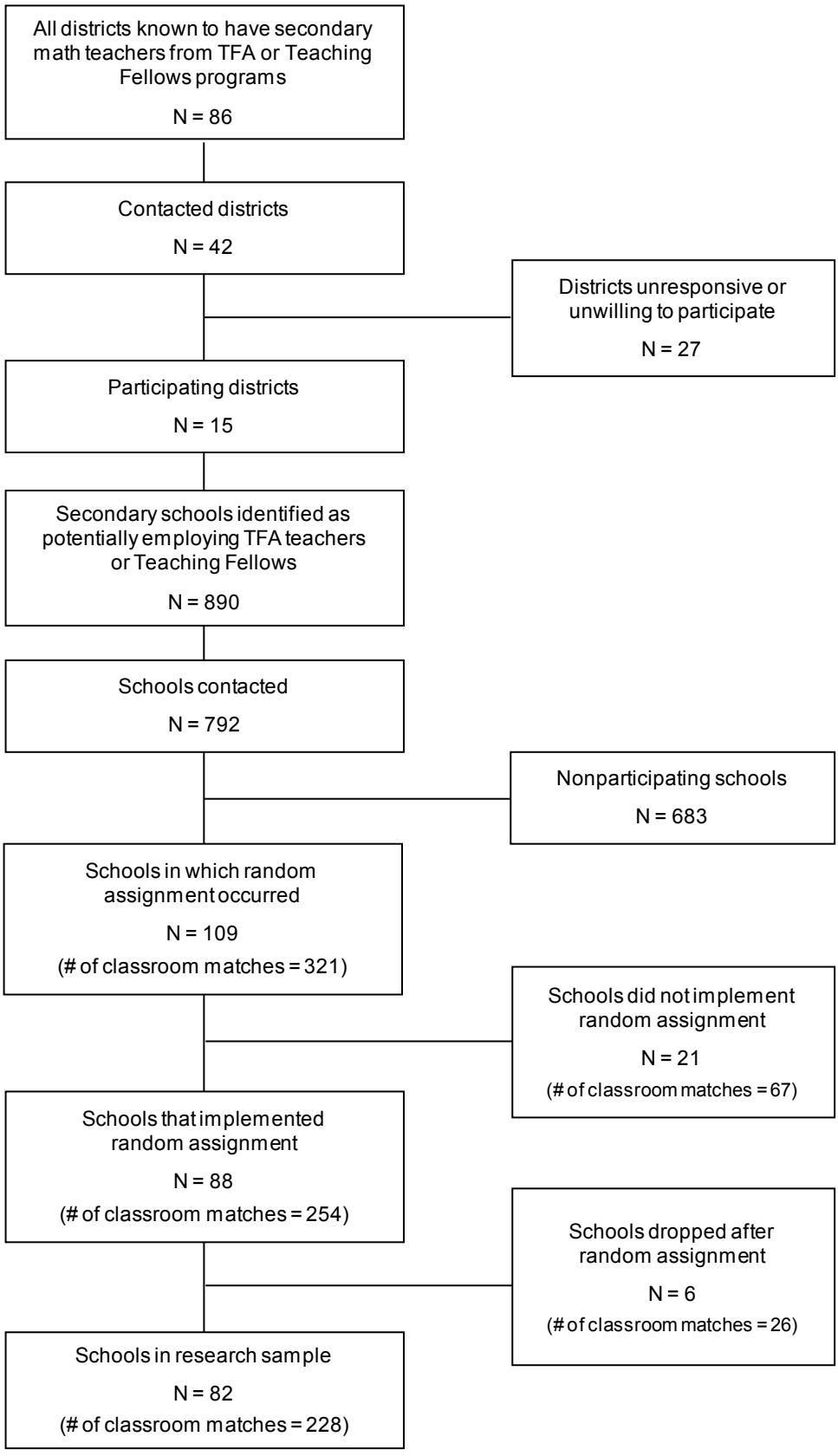


Table A.2. Number of Districts, Schools, Classroom Matches, Teachers, and Classes in the Study, by School Year

	Number of Sample Members		
	2009–2010 School Year	2010–2011 School Year	Both Years Combined
Districts	15	12	15
With TFA teachers in the study	10	8	11
With Teaching Fellows in the study	9	6	9
Schools	65	43	82
With TFA teachers in the study	30	25	45
With Teaching Fellows in the study	40	19	44
Classroom Matches[a]	140	88	228
TFA teachers versus comparison teachers	60	51	111
Teaching Fellows versus comparison teachers	81	37	118
Teachers[a, b]	199	122	287
TFA teachers	36	34	66
Teaching Fellows	55	25	69
Comparison teachers	108	63	152
Compared with TFA teachers	42	36	70
Compared with Teaching Fellows	68	27	84
Classes	329	188	517
Taught by TFA teachers	70	53	123
Taught by Teaching Fellows	93	42	135
Taught by comparison teachers	166	93	259
Compared with TFA teachers	72	53	125
Compared with Teaching Fellows	95	40	135

[a]Counts of classroom matches with TFA teachers and those with Teaching Fellows sum to more than the total number of matches because some matches contained both TFA and Teaching Fellows teachers. Likewise, counts of comparison teachers who were compared with TFA teachers and those who were compared with Teaching Fellows sum to more than the total number of comparison teachers because some comparison teachers were compared with both TFA and Teaching Fellows teachers.

[b]Counts of teachers from the 2009–2010 and 2010–2011 school years sum to more than the total number of teachers in the study because 34 teachers were in both years of the study.

TFA = Teach For America.

Table A.3. Structure of Classroom Matches in the TFA Study Sample

Number of TFA and Comparison Classes in the Classroom Match	Number of Classroom Matches
1 TFA Class, 1 Comparison Class	103
1 TFA Class, 2 Comparison Classes	3
Other Structures[a]	5
Total Number of Classroom Matches	**111**

TFA = Teach For America.

[a]Other structures include matches with 2 TFA classes and 1 comparison class; 3 TFA classes and 3 comparison classes; and 6 TFA classes and 6 comparison classes.

TFA = Teach For America.

Table A.4. Structure of Classroom Matches in the Teaching Fellows Study Sample

Number of Teaching Fellow and Comparison Classes in the Classroom Match	Number of Classroom Matches
1 Teaching Fellow Class, 1 Comparison Class	98
1 Teaching Fellow Class, 2 Comparison Classes	9
2 Teaching Fellow Classes, 1 Comparison Class	3
2 Teaching Fellow Classes, 3 Comparison Classes	3
Other Structures[a]	5
Total Number of Classroom Matches	**118**

[a]Other structures include matches with 2 Teaching Fellow classes and 2 comparison classes; 4 Teaching Fellow classes and 1 comparison class; and 4 Teaching Fellow classes and 2 comparison classes.

C. Selection and Assignment of Students

All students who enrolled in a study class before the start of the school year or in the first month of school were potentially eligible for random assignment and inclusion in the study sample. Initial random assignment occurred in the summers (2009 and 2010) preceding the study school years, as soon as schools were able to provide student lists for assignment. After this initial random assignment, the study team assigned additional students who needed to enroll in a study class through a process referred to as *rolling random assignment*. Ninety-one percent of students in the study were assigned via initial random assignment and 9 percent via rolling random assignment.[25] Next, we describe the two random assignment procedures, the process for verifying that random assignments were properly implemented, and the final student sample.

1. Initial Random Assignment

Initial random assignment was conducted in the study's Sample Management System via an embedded Visual Basic program. We allowed schools 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. We gave schools the opportunity to choose the variables used in stratification, rather than imposing specific stratifiers or requiring schools to choose from a pre-specified list, because different schools had different priorities for the types of student characteristics that had to be balanced between classes. The range of variables on which schools requested stratification included gender, race, ethnicity, academic ability, special education status, English language learner status, feeder school, grade level, age, behavior classification, and whether the student was retained in grade from the previous year. The exact characteristics used to define strata did not have any bearing on the randomization process. Regardless of how the strata were defined, the randomization process sought to equalize, within each stratum, the number of students assigned to the different classes in a classroom match, as described in more detail shortly.

[25] 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 Appendix B.

We also accommodated a limited number of special requests from the school, including grouping students together, assigning students to separate classes, and placing students in a particular class. Students who had to be grouped together or separated could be and were randomly assigned; students who had to be placed in a particular class (or "forced to" that class) were exempted from random assignment and excluded from the research sample. Of the 12,675 students in the study who were randomly assigned, 12 had to be grouped with a peer and 18 had to be separated from a peer. In addition, 46 students were forced to a particular class and exempted from random assignment.

If there were no exemptions from random assignment within 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 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:

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

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 forced to 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 initial random assignment was conducted, late enrolling students were assigned, individually or in small batches, in a process we referred to as rolling random assignment. School staff were provided with a toll-free 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. Assignments were not stratified. Rolling random assignment occurred through the first month of classes; after that time, schools were free to assign students to classes as they chose. Students who enrolled after the first month of school were not included in the study sample.

Because rolling random assignment typically occurred in the first few 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, the rolling random assignment program was constructed so that students would have 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—Teaching Fellows 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 Teaching fellows 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—Teaching Fellows 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 Teaching

Fellows 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 random assignment was conducted, schools were asked to send the study team updated rosters so we could verify that the assignments were properly implemented. If we could not verify assignments or identified students who had been placed with a type of teacher opposite to that of their assigned treatment status, we followed up with the school to resolve any issues—for instance, to request that particular students be placed in accordance with the random assignment results. In some cases, schools agreed to move misplaced students to their correctly assigned classes (and confirmed this move with an updated roster); in other cases, they refused to move these students. We considered random assignment to have been 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. 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. These rosters were used 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.

4. Student Sample

In the 228 matches in the study sample, 13,488 students were on the rosters schools sent during either initial or rolling random assignment (Figure A.2). Forty-six of these students (0.3 percent) were exempted from random assignment and forced to a particular classroom. Of the remaining students, approximately half were assigned to a TFA or Teaching Fellows teacher and half to a comparison teacher. In both treatment and control groups, 6 percent of students who were randomly assigned left the school before the start of the school year and were excluded from the study sample. The remaining 12,675 students (6,350 treatment and 6,325 control) formed the study's *research sample*. Of these students, 5,790 belonged to the TFA study sample and 6,909 belonged to the Teaching Fellows study sample; twenty-four control group students were in both samples because their class was compared with the classes of both TFA and Teaching Fellows teachers (see Chapter II, Table II.4).

As expected due to random assignment, treatment and control group students in the research sample were balanced in terms of most observed baseline characteristics. The treatment and control groups differed by a statistically significant margin on 4 (out of 23) characteristics in the TFA study sample and 5 (out of 23) characteristics in the Teaching Fellows study sample (Tables A.5 and A.6). The largest of these treatment-control differences consisted of a 2 percentage point difference in the percentage female within the TFA study sample, and a 2 percentage point difference in the percentage eligible for free or reduced-price lunch in the Teaching Fellows study sample.

Figure A.2. Number of Students Involved in Each Stage of Random Assignment and Data Collection

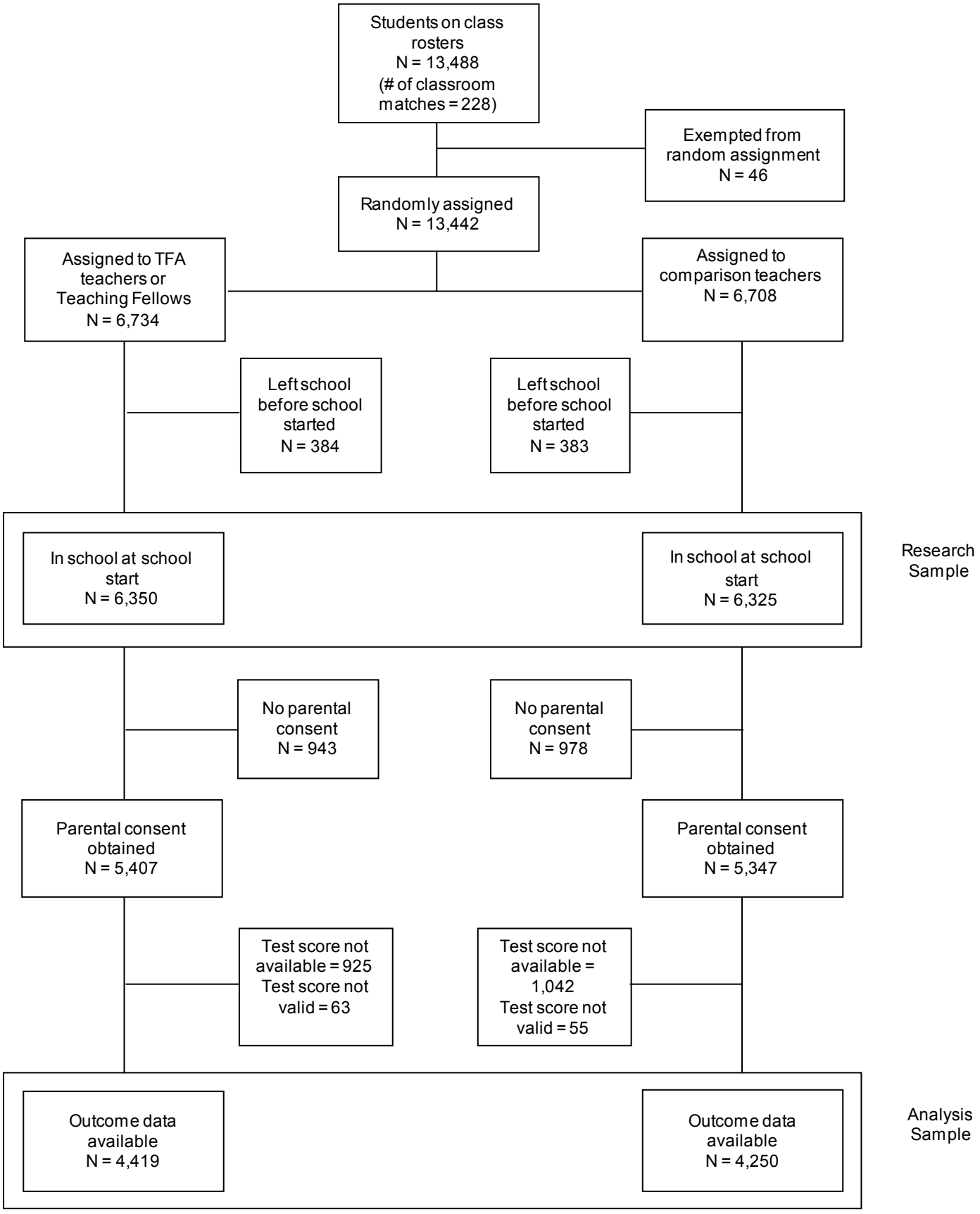


Table A.5. Average Baseline Characteristics of Students Assigned to TFA Teachers or Comparison Teachers (percentages unless otherwise indicated)

Characteristic	Assigned to TFA Teachers	Assigned to Comparison Teachers	Difference	*p*-Value
Baseline Math Achievement				
Baseline math score (average *z*-score)	-0.53	-0.53	-0.01	0.606
Score below 25th percentile of state	40.8	41.0	-0.2	0.775
Score below 50th percentile of state	73.4	73.6	-0.2	0.817
Baseline Reading Achievement				
Baseline reading score (average *z*-score)	-0.53	-0.53	0.00	0.851
Score below 25th percentile of state	41.3	40.6	0.7	0.352
Score below 50th percentile of state	72.6	73.7	-1.1	0.124
Age (average years)	13.62	13.56	0.06**	0.001
Female	48.4	50.3	-1.9**	0.007
Race and Ethnicity				
Asian, non-Hispanic	2.3	1.9	0.4	0.147
Black, non-Hispanic	62.3	63.2	-0.9	0.285
Hispanic	27.6	26.8	0.7	0.262
White, non-Hispanic	6.7	6.7	0.0	0.984
Other race/ethnicity	1.1	1.4	-0.2	0.442
Eligible for Free/Reduced-Price Lunch	88.2	88.7	-0.6	0.473
Limited English Proficient	8.1	8.6	-0.5	0.291
Individualized Education Plan	7.2	6.6	0.6	0.153
Grade Level in Study Year				
6th grade	25.0	25.2	-0.2	0.125
7th grade	27.0	26.5	0.4*	0.024
8th grade	18.8	19.0	-0.2	0.210
9th grade	17.4	17.0	0.5	0.123
10th grade	7.9	8.6	-0.7*	0.027
11th grade	3.4	3.3	0.1	0.655
12th grade	0.3	0.2	0.1	0.243
Number of Students	**2,725**	**2,737**		
Number of Classroom Matches	**111**	**111**		
Number of Teachers	**66**	**70**		
Number of Schools	**45**	**45**		

Source: District administrative records.

Note: Table is based on students in the research sample with parental consent. 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.

*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.

Table A.6. Average Baseline Characteristics of Students Assigned to Teaching Fellows or Comparison Teachers (percentages unless otherwise indicated)

Characteristic	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Difference	*p*-Value
Baseline Math Achievement				
Baseline math score (average *z*-score)	-0.17	-0.17	0.00	0.994
Score below 25th percentile of state	28.7	28.3	0.3	0.679
Score below 50th percentile of state	59.3	59.5	-0.2	0.853
Baseline Reading Achievement				
Baseline reading score (average *z*-score)	-0.11	-0.09	-0.02	0.305
Score below 25th percentile of state	24.7	24.1	0.5	0.591
Score below 50th percentile of state	58.6	59.0	-0.3	0.664
Age (average years)	14.62	14.60	0.02	0.123
Female	52.6	52.1	0.4	0.601
Race and Ethnicity				
Asian, non-Hispanic	8.5	7.6	0.8	0.161
Black, non-Hispanic	53.3	51.8	1.4*	0.034
Hispanic	32.4	34.2	-1.8*	0.014
White, non-Hispanic	5.3	5.5	-0.2	0.646
Other race/ethnicity	0.6	0.8	-0.2	0.145
Eligible for Free/Reduced-Price Lunch	73.5	75.7	-2.1*	0.011
Limited English Proficient	7.2	7.4	-0.2	0.714
Individualized Education Plan	7.1	7.1	0.1	0.901
Grade Level in Study Year				
6th grade	6.6	6.6	0.0	0.958
7th grade	12.0	12.1	-0.1*	0.050
8th grade	14.2	14.1	0.1	0.302
9th grade	29.7	30.0	-0.3	0.571
10th grade	19.4	20.0	-0.6	0.215
11th grade	16.3	15.1	1.2**	0.003
12th grade	1.8	2.1	-0.3	0.217
Number of Students	**2,682**	**2,631**		
Number of Classroom Matches	**118**	**118**		
Number of Teachers	**69**	**84**		
Number of Schools	**44**	**44**		

Source: District administrative records.

Note: Table is based on students in the research sample with parental consent. 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 Teaching Fellows indicator and classroom match indicators, accounting for sample weights and clustering at the teacher level.

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

Parental consent was obtained for 85 percent of research sample students in both the treatment and control groups (Figure A.2). Valid outcome test score data were obtained for 82 percent of consenting students in the treatment group and 79 percent of consenting students in the control group. The 8,669 students with parental consent and valid outcome test score data (4,419 treatment and 4,250 control group members) formed the study's *analysis sample* on which impact estimates were based. As discussed in Chapter II, the treatment and control groups in the analysis samples were also balanced in terms of most baseline characteristics.

The mobility of research sample members determined their level of exposure to the type of teacher to which they were originally assigned. Overall, most students in the research sample (77.2 percent) stayed in their originally assigned class for the full study year (Table A.7). The remaining students moved out of the class. The most common reason for leaving the class was departure from the study school, which occurred for 11.7 percent of research sample members.

Table A.7. Movement of Members of Research Sample During the School Year (percentages unless otherwise indicated)

Mobility Status	All Students in Research Sample	Assigned to TFA Teachers or Teaching Fellows	Assigned to Comparison Teachers
Crossed Over to Study Class with Opposite Teacher Type[a]	2.8	2.9	2.7
Never Crossed Over to Study Class with Opposite Teacher Type[a]	97.2	97.1	97.3
Left study school before end of year	11.7	11.8	11.5
Switched to nonstudy class in same school before end of year	6.7	6.2	7.1*
Switched to another study class with same teacher type before end of year[b]	1.6	1.6	1.7
Stayed in originally assigned class through end of year	77.2	77.4	77.0
Number of Students	**12,675**	**6,350**	**6,325**

Note: Teacher types are TFA teachers, Teaching Fellows, and comparison teachers. Data include both consenting and nonconsenting students. Calculations are unweighted. Tests of differences between TFA or Teaching Fellows classes and comparison classes are based on independent-sample *t*-tests with unequal variances.

[a]Crossovers consist only of cases in which students transferred to another study class with the opposite teacher type (a student in a TFA or Teaching Fellows study class transferring to a comparison class in the study, or vice versa). Transfers to classes or schools outside of the study sample are not classified as cases of crossover because the teacher type of the destination classroom is unknown.

[b]Includes only students who stayed in study classes through the end of the year.

*Difference between TFA or Teaching Fellows classes and comparison classes is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA or Teaching Fellows classes and comparison classes is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Not all students in study classes were randomly assigned. Those who were not randomly assigned were excluded from the study sample, but their presence could theoretically have affected the achievement of randomly assigned students. We requested that schools contact us to randomly assign all students who enrolled in study classes up through the first month of school; however, some students were exempted from random assignment by school request, and schools might have failed to request assignments for other late-enrolling students. After the first month of school, schools were free to place newly enrolling students without random assignment. As of the end of the school year, the percentage of students in study classes who had not been randomly assigned was 14 percent in both treatment and control classes.

We examined the baseline characteristics of nonrandomly assigned students on class rosters at the end of the study year for the TFA and Teaching Fellows study samples, to determine whether particular types of students appeared to be systematically placed with TFA/Teaching Fellows or comparison teachers. In the TFA study sample, students nonrandomly assigned to TFA teachers had lower baseline math scores, were more likely to have baseline reading scores below the 25th percentile in their state, were more likely to be female, and were less likely to be in 9th grade than students nonrandomly assigned to comparison teachers (Table A.8). In the Teaching Fellows study sample, students nonrandomly assigned to Teaching Fellows were more likely to be white and were distributed differently across high school grade levels than students nonrandomly assigned to comparison teachers (Table A.9). There were no other significant differences between treatment and control classes in the characteristics of nonrandomly assigned students.

D. Statistical Power

With an analysis sample of 136 teachers and 4,573 students in the TFA study sample and 153 teachers and 4,116 students in the Teaching Fellows study sample, the study had considerable precision for estimating the average impacts of TFA and Teaching Fellows teachers relative to the comparison teachers with whom they were compared. The minimum detectable effect (MDE) is the smallest true impact for which there would be an 80 percent probability of obtaining a statistically significant estimate. We computed MDEs based on the realized values of key parameters determining the study's precision for the six impact estimates deemed to be of primary policy interest, as discussed in Chapter II—impacts for the full sample of TFA teachers, the subsample compared with TC teachers, and the subsample compared with less-selective AC teachers, as well as the full sample of Teaching Fellows, the subsample compared with TC teachers, and the subsample compared with less-selective AC teachers. The impact of the full sample of TFA teachers relative to comparison teachers was 0.07 standard deviations (Table A.10). That is, if students truly scored at least 0.07 standard deviations higher by being assigned to TFA teachers rather than comparison teachers, then any study with the same design on the same population of teachers would have at least an 80 percent probability of obtaining a statistically significant impact estimate. Similarly, the MDE for the impact of the full sample of Teaching Fellows relative to comparison teachers was 0.09 standard deviations. Given prior tabulations for the average year-to-year gain in math achievement on nationally standardized assessments (Hill et al. 2008), these minimum detectable effect sizes were equivalent to 2.6 to 3.3 months of learning out of a 10-month school year. MDEs for impacts within subgroups defined by the certification route of the comparison teachers ranged from 0.10 to 0.16 standard deviations.

Table A.8. Characteristics of Nonstudy Students on End-of-Year Rosters of Classrooms in the TFA Study Sample (percentages unless otherwise indicated)

Characteristic	In Classes of TFA Teachers	In Classes of Comparison Teachers	Difference	*p*-Value
Baseline Math Achievement				
Baseline math score (average *z*-score)	-0.84	-0.72	-0.12*	0.035
Score below 25th percentile of state	55.4	52.4	3.0	0.435
Score below 50th percentile of state	81.7	80.2	1.6	0.609
Baseline Reading Achievement				
Baseline reading score (average *z*-score)	-0.82	-0.72	-0.10	0.229
Score below 25th percentile of state	58.4	49.1	9.3*	0.018
Score below 50th percentile of state	76.9	77.0	-0.1	0.986
Age (average years)	13.6	13.67	-0.04	0.373
Female	50.0	38.5	11.5**	0.001
Race and Ethnicity				
Asian, non-Hispanic	1.3	1.4	-0.1	0.885
Black, non-Hispanic	70.9	73.0	-2.0	0.367
Hispanic	21.7	20.8	0.9	0.677
White, non-Hispanic	5.2	4.1	1.1	0.379
Other race/ethnicity	1.0	0.8	0.2	0.786
Eligible for Free/Reduced-Price Lunch	87.9	87.7	0.2	0.910
Limited English Proficient	16.1	16.7	-0.6	0.868
Individualized Education Plan	8.6	9.6	-0.9	0.579
Grade Level in Study Year				
6th grade	21.1	21.4	-0.2	0.675
7th grade	36.5	35.0	1.4	0.080
8th grade	21.8	22.5	-0.7	0.163
9th grade	8.7	10.8	-2.2*	0.025
10th grade	8.1	6.8	1.3	0.065
11th grade	2.4	2.3	0.1	0.856
12th grade	1.4	1.1	0.3	0.665
Number of Nonstudy Students with Parental Consent and Any Baseline Data	**343**	**366**		
Percentage of Nonstudy Students with Parental Consent and Any Baseline Data	**85.1**	**85.3**		

Sources: District administrative records and study-administered Northwest Evaluation Association assessments.

Note: Numbers are based on consenting students who do not belong to the research sample but are listed on end-of-year rosters of study classes. Means and percentages are unweighted and adjusted for classroom match fixed effects. *p*-values are based on a regression of the specified characteristic on an indicator for TFA classes and classroom match indicators, accounting for clustering at the teacher level.

*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.

Table A.9. Characteristics of Nonstudy Students on End-of-Year Rosters of Classrooms in the Teaching Fellows Study Sample (percentages unless otherwise indicated)

Characteristic	In Classes of Teaching Fellows	In Classes of Comparison Teachers	Difference	*p*-Value
Baseline Math Achievement				
Baseline math score (average *z*-score)	-0.42	-0.48	0.06	0.394
Score below 25th percentile of state	40.8	38.2	2.6	0.637
Score below 50th percentile of state	76.9	74.7	2.2	0.602
Baseline Reading Achievement				
Baseline reading score (average *z*-score)	-0.45	-0.55	0.09	0.207
Score below 25th percentile of state	38.0	47.4	-9.4	0.052
Score below 50th percentile of state	75.2	72.5	2.7	0.585
Age (average years)	14.3	14.23	0.10	0.412
Female	47.5	47.1	0.4	0.904
Race and Ethnicity				
Asian, non-Hispanic	5.2	6.1	-0.9	0.629
Black, non-Hispanic	42.0	41.4	0.6	0.869
Hispanic	41.9	47.8	-6.0	0.063
White, non-Hispanic	9.5	4.7	4.8*	0.026
Other race/ethnicity	1.5	0.0	1.5	0.065
Eligible for Free/Reduced-Price Lunch	77.6	74.1	3.5	0.295
Limited English Proficient	22.0	20.9	1.2	0.800
Individualized Education Plan	9.1	9.4	-0.3	0.924
Grade Level in Study Year				
6th grade	17.0	17.8	-0.8	0.238
7th grade	12.7	13.3	-0.6	0.179
8th grade	8.2	7.6	0.7	0.264
9th grade	40.6	43.2	-2.5	0.050
10th grade	13.1	9.5	3.6*	0.016
11th grade	2.9	7.6	-4.7*	0.025
12th grade	5.5	1.1	4.3*	0.023
Number of Nonstudy Students with Parental Consent and Any Baseline Data	**263**	**278**		
Percentage of Nonstudy Students with Parental Consent and Any Baseline Data	**60.0**	**63.9**		

Sources: District administrative records and study-administered Northwest Evaluation Association assessments.

Note: Numbers are based on consenting students who do not belong to the research sample but are listed on end-of-year rosters of study classes. Means and percentages are unweighted and adjusted for classroom match fixed effects. *p*-values are based on a regression of the specified characteristic on an indicator for classes of Teaching Fellows and classroom match indicators, accounting for clustering at the teacher level.

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

Table A.10. Realized Values for Key Precision Parameters

Parameter	Value in TFA Study Sample	Value in Teaching Fellows Study Sample
Sample Sizes		
Number of schools	45	44
Number of teachers per school	3.0	3.5
Number of classes per teacher	1.8	1.8
Number of randomly assigned students per class	23.3	25.6
Student Response Rate	0.79	0.60
Teacher-Level Intraclass Correlation Coefficient	0.24	0.52
Teacher-Level R-squared	0.94	0.96
Student-level R-squared	0.41	0.36
Sample Standard Deviation of End-of-Year Math Scores (expressed in standard deviations within the reference population)	0.93	1.08
Minimum Detectable Effect for:		
Full study sample	**0.07**	**0.09**
Subsample for comparisons with TC teachers[a]	**0.10**	**0.10**
Subsample for comparisons with less-selective AC teachers[b]	**0.11**	**0.16**

Note: Minimum detectable effects are expressed in standard deviations of outcome scores within the reference population of the student's assessment.

[a]Within the TFA study sample, calculations are based on a 60 percent subsample of teachers; within the Teaching Fellows sample, calculations are based on a 74 percent subsample of teachers. All other parameters are fixed at their full-sample values.

[b]Within the TFA study sample, calculations are based on a 43 percent subsample of teachers; within the Teaching Fellows sample, calculations are based on a 30 percent subsample of teachers. All other parameters are fixed at their full-sample values.

TFA = Teach for America.

E. Data Reliability and Response Rates

Below we discuss the psychometric properties of the Northwest Evaluation Association (NWEA) assessments administered to high school students in the study and the state assessments taken by middle school students in the study, as well as response rates for students (test score data) and teachers (survey and Praxis data).

1. Psychometric Properties of NWEA Assessments

To measure the math achievement of students in grades 9 to 12 in the study, we administered end-of-course math assessments developed by the NWEA. Based on the content of their math course, students took computer-adaptive assessments in either general high school math, Algebra I, Geometry, or Algebra II. Departing from standard NWEA procedures, which do not time-limit the tests, we imposed a 35-minute time limit on the assessments so they could be administered in a single class period. Although NWEA typically does not score incomplete tests, it provided us with Rasch Unit (RIT) scores and standard errors of measurement (SEMs) on all students, regardless of whether they completed all 50 items. Although imposing a time limit and

accepting incomplete tests were expected to lower the reliability of the test scores, the actual reliability of NWEA scores in the analysis sample remained high. Marginal reliability coefficients—measuring the internal consistency of the assessments (Samejima 1977, 1994)—for all four of the NWEA assessments used in the study were greater than 0.90 in the analysis sample (Table A.11).

Table A.11. Reliability of the NWEA Assessments, by Subject

	Marginal Reliability Coefficient	
Subject	Published by NWEA	In Analysis Sample
General High School Math	0.974[a]	0.927
Algebra I	0.964[a]	0.957
Geometry	0.962[b]	0.969
Algebra II	0.949[b]	0.968

Source: Published reliability coefficients are from NWEA technical documentation; reliability coefficients for analysis sample are from authors' calculations.

[a]Value is the NWEA's published marginal reliability coefficient for test scores taken in 9th grade in the indicated subject.

[b]Value is the NWEA's published marginal reliability coefficient for test scores taken in 10th grade in the indicated subject.

NWEA = Northwest Evaluation Association.

To improve the reliability of the NWEA data, we designated scores as invalid, and treated them as missing in the analysis, for one or more of the following reasons:

1. **The standard error of measurement of the test was greater than 5.5, or the student completed the test in fewer than six minutes.** This rule was recommended by NWEA for assessing test score validity. Four percent or less of each group had scores designated as invalid for this reason (Table A.12).
2. **Students took the wrong assessment**. This may occur when students moved from one study class to another. For example, a student assigned to an Algebra II class may have moved to Algebra I in the middle of the year. In the Algebra I class, this student may have been erroneously administered the Algebra I test. This occurred for less than 2 percent of all tested students in the treatment or control groups.

Overall, approximately 4 percent of NWEA scores were designated as invalid in the treatment or control groups in either the TFA or Teaching Fellows samples for one of these three reasons.

Table A.13 shows statistics on NWEA scores by exam for the analysis sample of students with valid scores. Across all four assessments, the average standard error of measurement ranged from 3.4 to 3.7, the average number of items completed (out of 50) ranged from 37.4 to 43.8, and the average test completion time ranged from 27.6 to 30.6 minutes.

Table A.12. Invalid NWEA Scores, by Treatment Status

Reason Score Was Categorized Invalid	Percentage with Invalid NWEA Scores: Assigned to TFA Teachers or Teaching Fellows	Assigned to Comparison Teachers	Difference	*p*-Value
TFA Study Sample				
Test was in a different subject than the most commonly administered test in the classroom match	0.5	0.6	-0.2	0.673
Standard error of measurement > 5.5 RIT, or test was completed in fewer than 6 minutes	3.9	3.4	0.5	0.628
Any of the specified reasons	4.4	4.1	0.3	0.767
Teaching Fellows Study Sample				
Test was in a different subject than the most commonly administered test in the classroom match	0.8	1.4	-0.6	0.120
Standard error of measurement > 5.5 RIT, or test was completed in fewer than 6 minutes	3.0	2.5	0.4	0.489
Any of the specified reasons	3.8	4.0	-0.2	0.801
Sample Sizes: TFA Study Sample				
Students	**637**	**616**		
Classroom Matches	**28**	**28**		
Teachers	**17**	**16**		
Schools	**14**	**14**		
Sample Sizes: Teaching Fellows Study Sample				
Students	**1,271**	**1,336**		
Classroom Matches	**81**	**81**		
Teachers	**41**	**60**		
Schools	**29**	**29**		

Source: Study-administered NWEA assessments. Numbers are based on students in the research sample who took an NWEA assessment.

Note: *p*-values are based on independent-sample *t*-tests with unequal variances.

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

NWEA = Northwest Evaluation Association; RIT = Rasch Units; TFA = Teach for America.

Table A.13. NWEA Test Statistics in Analysis Sample

	Mean	Standard Deviation	Percentile 5th	25th	50th	75th	95th
General High School Math							
RIT score	217.3	12.7	197.3	206.8	216.1	226.6	240.2
Standard error of measurement	3.4	0.5	2.9	3.0	3.2	3.6	4.4
Number of items completed	43.8	9.5	24.0	35.5	50.0	50.0	50.0
Test completion time (minutes)	27.6	5.8	17.5	22.2	30.4	32.4	34.0
Number of Students in Analysis Sample with Valid NWEA Scores in General High School Math	**38**						
Algebra I							
RIT score	229.7	18.1	197.0	219.6	231.2	241.2	258.3
Standard error of measurement	3.7	0.6	2.9	3.1	3.5	4.1	5.0
Number of items completed	37.4	11.2	19.0	28.0	38.0	50.0	50.0
Test completion time (minutes)	30.6	4.5	20.3	29.6	32.3	33.5	34.4
Number of Students in Analysis Sample with Valid NWEA Scores in Algebra I	**1,488**						
Geometry							
RIT score	238.1	20.2	201.5	226.1	240.3	253.1	268.1
Standard error of measurement	3.5	0.6	2.9	3.0	3.3	3.9	4.8
Number of items completed	40.4	10.1	22.0	32.0	43.0	50.0	50.0
Test completion time (minutes)	29.9	5.2	19.3	28.1	32.1	33.4	34.3
Number of Students in Analysis Sample with Valid NWEA Scores in Geometry	**1,287**						
Algebra II							
RIT score	240.5	20.7	204.1	227.6	240.2	255.0	274.4
Standard error of measurement	3.6	0.6	2.9	3.1	3.5	4.1	4.9
Number of items completed	38.4	10.7	20.0	29.0	39.0	50.0	50.0
Test completion time (minutes)	30.3	4.3	21.5	28.2	31.9	33.4	34.2
Number of Students in Analysis Sample with Valid NWEA Scores in Algebra II	**893**						

Source: Study-administered NWEA assessments.

NWEA = Northwest Evaluation Association; RIT = Rasch Unit.

2. Psychometric Properties of State Math Assessments

To measure the end-of-year math achievement of students in grades 6 through 8 in the study, we obtained their scores on state math assessments. All of these math assessments were criterion-referenced tests—those that measured students' mastery of state-specific curriculum standards. Each state in the study administered a different assessment. Among the districts in the study, the dates on which state math assessments were administered ranged from early March to late May in the spring of the school year.

All state math assessments used in the study had high reliability based on measures of internal consistency calculated from coefficient alpha (Table A.14). Across the eight states represented in the study (labeled A through H in Table A.14 to preserve confidentiality), reliability values ranged from 0.89 to 0.94. These reliability values were close to, but slightly lower than, the reliability values for the NWEA assessment presented in Table A.11; this was expected because of the adaptive nature of the NWEA assessment.

In each state assessment, the precision of the test scores varied across the range of the achievement scale. Test scores were less precise (had higher SEMs) at the low and high ends of the achievement scale compared with scores in the middle of the scale. Given that a key objective of the state assessments was to differentiate students who were proficient on the states' standards from those who were not, larger shares of the test items had a level of difficulty appropriate for differentiating these groups. Therefore, precision was generally highest for scores at or near the proficiency cut point (the point on the achievement scale that divided the nonproficient and proficient scores). As Table A.14 shows, the SEMs at the proficiency cut points were at or near the minimum value for the range of SEMs in each state assessment.

Consequently, the average precision of math scores among middle school students in the study was determined by the distribution of these scores relative to the proficiency cut point. About half (49 percent) of students in the TFA study sample and slightly more than half (56 percent) of students in the Teaching Fellows study sample scored above the proficiency cut point (Table A.15). Therefore, in each sample, the average score of the sample members was no more than 0.12 standard deviations away from the proficiency cut point. This result implies that test scores among middle school students in the analysis sample were centered near the parts of the achievement scale at which the scores were most precise.

Table A.14. Characteristics of State Math Assessments from which End-of-Year Scores Were Used in the Study

			Conditional Standard Error of Measurement			
State	Grade	Reliability	Lowest	Highest	At Proficiency Cut Point	Vertical Scale?
A	6	0.94	N/A	N/A	16	No
B	6	0.94	14	81	15	Yes
B	7	0.93	12	106	14	Yes
B	8	0.94	11	100	12	Yes
C	6	N/A	N/A	N/A	N/A	No[a]
C	7	N/A	N/A	N/A	N/A	No[a]
C	8	N/A	N/A	N/A	N/A	No[a]
D	6	0.89	N/A	N/A	N/A	No
D	7	0.90	N/A	N/A	N/A	No
E	6	0.92[b]	N/A	N/A	9	Yes
E	7	0.92[b]	N/A	N/A	8	Yes
F	6	0.92[c]	2	6	3	Yes
F	7	0.92[c]	2	6	3	Yes
F	8	0.92[c]	2	6	3	Yes
G	6	0.93	6	125	6	No
G	7	0.92	6	129	6	No
G	8	0.93	5	162	5	No
H	6	0.91	22	70	22	Yes
H	7	0.90	21	69	21	Yes
H	8	0.91	21	69	21	Yes

Source: State assessment technical reports.

Note: All statistics pertain to the Spring 2011 administration of the test unless otherwise noted. Reliability values are measures of internal consistency based on coefficient alpha.

[a]The state assessment also produces vertically scaled scores, but those scores were not used in the study.

[b]Statistics are based on the Spring 2010 administration because the state was not included in the study in the 2010–2011 school year.

[c]Statistics are based on the first administration—in Spring 2006—of the test edition used in the study.

N/A = not available from state assessment technical reports.

Table A.15. State Math Assessment Scores of Middle School Students in the Analysis Sample Relative to Proficiency Cut Points

Subsample	Average Difference between End-of-Year Test Score and Proficiency Cut Point (*z*-score units)	Percentage of Students Scoring At or Above Proficiency Cut Point
Teach For America Study Sample	-0.12	49.0
Teaching Fellows Study Sample	0.10	56.3

Source: Test scores from district administrative records; proficiency cutoffs from state education agencies.

3. Response Rates for Students

On average, we had valid outcome test score data (from either state assessments or the NWEA) for 79 percent of students in TFA matches and 60 percent of students in Teaching Fellows matches (Table A.16). In TFA matches, average response rates for the treatment and control groups were similar—they did not differ by more than 2 percentage points for all TFA matches or for the subsets of matches with comparison teachers from a traditional route to certification (TC route) and an alternative route to certification (AC route). In Teaching Fellows matches, average response rates for the treatment and control groups differed by up to 5 percentage points, with higher response rates for students in the treatment group. In the full set of Teaching Fellows matches, the response rate was 61 percent for Teaching Fellows classes and 58 percent for control classes. In matches in which Teaching Fellows and TC teachers were compared, the response rate was 63 percent for Teaching Fellows classes and 58 percent for control classes. In matches in which Teaching Fellows and AC teachers were compared, the response rate was 56 percent for Teaching Fellows classes and 58 percent for control classes. For both TFA and Teaching Fellows matches, response rates were higher at the middle school level than at the high school level.

As shown earlier in Figure A.2, overall student response rates depended on parental consent rates as well as test score availability among consenting students. Overall, on average, we obtained parental consent for 85 percent of students in the research sample, for both the treatment and control groups. We obtained parental consent and valid outcome test score data for 68 percent of the research sample (70 percent of the treatment group and 67 percent of the control group). On average, both consent rates (88 versus 82 percent) and rates of valid outcome test score data (80 versus 57 percent) were higher at the middle school level than at the high school level.

Table A.16. Student Response Rates, by Type of Impact Estimate to Which the Student's Classroom Contributes

Type of Impact Estimate to Which the Student's Classroom Contributes	Percentage of Research Sample Students with Valid End-of-Year Math Scores		
	Assigned to TFA Teachers or Teaching Fellows	Assigned to Comparison Teachers	Total
TFA Teachers Versus All Comparison Teachers	79.5	78.5	79.0
TFA teachers versus TC teachers	75.8	75.4	75.6
TFA teachers versus other AC teachers	84.1	82.6	83.4
Teaching Fellows Versus All Comparison Teachers	61.4	57.8	59.6
Teaching Fellows versus TC teachers	62.9	57.8	60.4
Teaching Fellows versus other AC teachers	55.9	57.8	56.8
Number of Students			
TFA teachers versus all comparison teachers	**2,884**	**2,906**	**5,790**
TFA teachers versus TC teachers	**1,610**	**1,667**	**3,277**
TFA teachers versus other AC teachers	**1,274**	**1,239**	**2,513**
Teaching Fellows versus all comparison teachers	**3,466**	**3,443**	**6,909**
Teaching Fellows versus TC teachers	**2,762**	**2,650**	**5,412**
Teaching Fellows versus other AC teachers	**794**	**793**	**1,587**

Sources: District administrative records and study-administered Northwest Evaluation Association assessments.

Note: Calculations are unweighted.

AC = alternative route to certification; TC = traditional route to certification; TFA = Teach For America.

Compared with students with valid outcome data, students without valid outcome data had much higher rates of absenteeism and differed on some characteristics (Table A.17). On average, for both treatment and control group students, those with valid outcome data had higher baseline test scores (statistically significant only for control group math scores) and were younger, more likely to be eligible for free or reduced-price lunch, and less likely to have an individualized education plan (IEP) (statistically significant only for the control group). Consistent with the fact that students who were frequently absent were less likely to be present for testing (whether for state testing or the study assessment), for both the treatment and control groups, those with valid outcome data had, on average, 11 fewer days of absence than those without valid outcome data and were only about one-fourth as likely (11 versus 44 percent for the treatment group; 10 versus 46 percent for the control group) to have been absent from math class more than 25 percent of the time.

Table A.17. Characteristics of Students With and Without Valid Outcome Data (percentages unless otherwise indicated)

Characteristic	Assigned to TFA Teachers or Teaching Fellows		Assigned to Comparison Teachers	
	Valid Outcome Data	No Valid Outcome Data	Valid Outcome Data	No Valid Outcome Data
Baseline Math Score (average *z*-score)	-0.34	-0.39	-0.33	-0.42**
Baseline Reading Score (average *z*-score)	-0.31	-0.34	-0.30	-0.35
Age (average years)	13.86	14.01**	13.81	13.97**
Female	50.8	49.9	52.2	47.9
Race and Ethnicity				
Asian, non-Hispanic	5.3	5.7	4.8	5.0
Black, non-Hispanic	56.5	53.9	55.9	55.5
Hispanic	31.4	33.0	32.0	31.9
White, non-Hispanic	6.0	6.3	6.2	6.4
Other race/ethnicity	0.9	1.1	1.1	1.3
Eligible for Free/Reduced-Price Lunch	80.4	75.4*	81.9	75.4**
Limited English Proficient	8.0	6.4	7.9	9.6
Individualized Education Plan	6.6	9.7	6.1	9.2*
Grade Level in Study Year				
6th Grade	18.5	17.2	18.5	18.3
7th Grade	21.0	22.4	21.1	20.4
8th Grade	18.9	17.9	18.9	19.0
9th Grade	20.5	24.4*	20.7	22.2
10th Grade	12.4	10.2	12.8	11.4
11th Grade	8.1	5.2	7.5	5.9
12th Grade	0.6	1.9	0.6	2.4*
Number of Days Absent (average days)	11.6	22.7**	11.4	22.1**
Percentage of Days Absent from Math Class				
No more than 25 percent	89.3	56.2**	89.7	54.5**
More than 25 percent	10.7	43.8**	10.3	45.5**
Number of Students[a] (district administrative records)	**4,390**	**937**	**4,220**	**1,040**
Number of Students (teacher-reported attendance data)	**2,952**	**391**	**2,762**	**390**

Sources: Demographic characteristics, baseline test scores, and number of days absent are from district administrative records. Percentage of days absent from math class is from teacher reports.

Note: Numbers are based on students in the research sample with parental consent. Means and percentages are weighted with sample weights and adjusted for classroom match fixed effects. Within each research group (students of TFA or Teaching Fellows teachers and students of comparison teachers), tests of differences between those with and without valid outcome data are based on a regression of the specified characteristic on an indicator for having valid outcome data and classroom match indicators, accounting for sample weights and clustering at the teacher level.

[a]Sample sizes pertain to research sample members with information on any of the listed baseline characteristics from districts' administrative records. Students with end-of-year scores but with no information on any of the listed baseline characteristics are not counted in the sample sizes for this table.

*Difference between those with and without valid outcome data is statistically significant at the 0.05 level, two-tailed test.
**Difference between those with and without valid outcome data is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

4. Response Rates for Teachers

Response rates for the teacher survey and Praxis assessment were higher for TFA and Teaching Fellows teachers than for comparison teachers, and response rates for the teacher survey were generally higher than those for the Praxis (Table A.18). An overwhelming majority (97 percent) of TFA teachers and 90 percent of comparison teachers in the same matches completed the teacher survey, as did 96 percent of Teaching Fellows and 89 percent of comparison teachers in the same matches. Ninety-one percent of TFA teachers and 73 percent of comparison teachers in the same matches had Praxis II math assessment data, as did 94 percent of Teaching Fellows and 81 percent of comparison teachers in the same matches.

Table A.18. Teacher Response Rates and Sample Sizes

	Percentage of Research Sample Teachers with Specified Type of Data		
TFA Study Sample	**TFA**	**Comparison**	**Total**
Percentage of Teachers Responding to Teacher Survey	97.0	90.0	92.7
Percentage of Teachers with Scores on Teacher Assessment (Praxis)	90.9	72.9	84.3
Number of Teachers	**66**	**70**	**287**
Teaching Fellows Study Sample	**Teaching Fellows**	**Comparison**	**Total**
Percentage of Teachers Responding to Teacher Survey	95.7	89.3	92.7
Percentage of Teachers with Scores on Teacher Assessment (Praxis)	94.2	81.0	84.3
Number of Teachers	**69**	**84**	**287**

Note: Calculations are unweighted.

TFA = Teach For America.

For both the teacher survey and the Praxis, nonresponding comparison teachers had significantly more years of teaching experience than responding comparison teachers (Tables A.19 and A.20). There were no significant differences in years of teaching experience between responding and nonresponding TFA teachers and Teaching Fellows. There were also no significant differences between respondents and nonrespondents in route to certification and the level (middle or high) of their school, among both the TFA/Teaching Fellows teachers and comparison teachers.

Table A.19. Characteristics of Teachers, by Response to Teacher Survey (percentages unless otherwise indicated)

Characteristic	TFA or Teaching Fellows Teachers		Comparison Teachers	
	Responded	Did Not Respond	Responded	Did Not Respond
Route to Certification				
TFA	49.2	40.0	n.a.	n.a.
Teaching Fellows	50.8	60.0	n.a.	n.a.
Traditional	n.a.	n.a.	64.7	75.0
Alternative	n.a.	n.a.	35.3	25.0
Years of Teaching Experience	2.9	3.0	11.8	16.9*
School Level[a]				
Middle school	56.2	80.0	51.5	43.8
High school	43.8	20.0	49.3	56.3
Number of Teachers	**130**	**5**	**136**	**16**

Source: Teacher background form.

[a]Percentages of teachers in middle school and high school grades can sum to more than 100 percent because some teachers taught in both levels.

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

TFA = Teach For America; n.a = not applicable.

Table A.20. Characteristics of Teachers, by Availability of Praxis II Scores (percentages unless otherwise indicated)

Characteristic	TFA or Teaching Fellows Teachers		Comparison Teachers	
	Scores Available	Scores Unavailable	Scores Available	Scores Unavailable
Route to Certification				
TFA	48.0	60.0	n.a.	n.a.
Teaching Fellows	52.0	40.0	n.a.	n.a.
Traditional	n.a.	n.a.	62.4	77.1
Alternative	n.a.	n.a.	37.6	22.9
Years of Teaching Experience	3.0	2.4	11.0	16.9**
School level[a]				
Middle school	59.2	30.0	52.1	45.7
High school	40.8	70.0	47.9	57.1
Number of Teachers	**125**	**10**	**117**	**35**

Source: Teacher background form.

[a]Percentages of teachers in middle and high school grades can sum to more than 100 percent because some teachers taught in both levels.

*Difference between those with and without scores is statistically significant at the 0.05 level, two-tailed test.
**Difference between those with and without scores is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America; n.a. = not applicable.

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

APPENDIX B

SUPPLEMENTARY INFORMATION ON ANALYTIC METHODS

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

In Appendix B, we describe the analytic methods used in this report. We first discuss the construction of sample weights (Section A) and the specification of baseline achievement scores (Section B). We then describe the methods used in the contextual analyses (Section C), experimental analyses (Section D), and nonexperimental analyses (Section E).

A. Sample Weights

In the analyses of student data, sample weights were used to account for differences in probabilities of assignment to the treatment group (classes of Teach For America [TFA] or Teaching Fellows teachers) or the control group (classes of comparison teachers) among students in the same classroom match. Random assignment procedures and associated assignment probabilities are described in more detail in Appendix A. Sample weights ensured that even if certain types of students (for instance, late enrollees) had a higher or lower probability of assignment to the treatment group than other types of students, the weighted treatment and control groups within a classroom match were both fully representative of all randomly assigned students in the match.

We constructed sample weights for all students in the research sample—that is, randomly assigned students who did not leave the school before the start of the school year. 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:

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

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.

We then normalized the raw weights so that the sum of the normalized weights within a match equaled the total number of research sample 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. Specifically, we used the following formula:

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

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 research sample students assigned to group g in match k, and N_k is the total number of research sample students in match k.

B. Specification of Baseline Achievement Scores

Baseline achievement scores—that is, math and reading scores that students earned before being assigned to a study class—were key covariates in the experimental and nonexperimental analyses. All baseline scores were expressed as *z*-scores based on the statewide mean and

standard deviation of scores in the grade level and year in which the assessment was administered. This section describes our rules for selecting the grade level from which we obtained each student's baseline scores.

Even within the same classroom match, students differed in the grade level at which they had most recently been assessed by end-of-grade state assessments. For example, in middle school, the grade level that students attended in the immediate prior year differed between grade repeaters and nonrepeaters in the same classroom match. Among high school students, although their last end-of-grade state assessment had typically occurred in 8th grade, several states in the study administered end-of-grade exams in selected high school grades. In those states, high school students in the same classroom match who differed in their current grade level could thereby also differ in the grade level in which they last took an end-of-grade assessment.

Our basic approach to specifying students' baseline scores was to require that baseline scores for all students in the same classroom match be taken from the same grade level. We focused on making the grade level of the baseline scores consistent within classroom matches because only scores within the same matches were compared. The advantage of our approach was that the *z*-scores being compared were thereby based on similar distributions of achievement—distributions pertaining to the same grade level.

Separately for each classroom match, we used the following two steps to identify the grade level from which baseline scores were taken:

1. We calculated the 10th percentile of current grade level within the classroom match. This represented the highest grade level that nearly all (at least 90 percent of) students in the match had reached. We chose not to identify the minimum grade level in the match because this value would have been more vulnerable to data errors, especially those in which a student's recorded grade level was much lower than his or her actual grade level.
2. We identified the highest grade level that was less than the grade level calculated in step 1 and was a grade level at which end-of-grade state exams were administered. Baseline scores for all students in the classroom match were taken from this identified grade level.

We then identified each student's most recent reading and math scores from the grade level specified in step 2 above. These two scores served as the student's baseline scores in the analysis. Cases in which students did not have a score from the specified baseline grade level were treated as missing data. Notably, students differed—even within the same match—in the number of years elapsed since they earned their baseline scores. Therefore, in the impact analyses, we controlled for a set of binary indicators representing different numbers of years elapsed since students earned their baseline scores (see Section D).

We did not control for students' scores from additional grades before the baseline grade level specified in the preceding discussion. Findings by Bloom et al. (2007) indicate that controlling for scores from multiple prior grades would have yielded little to no improvement in precision beyond controlling for scores from a single prior grade. Moreover, higher rates of missing data in scores from additional earlier grades—especially among high school students—would have further limited the precision gains from including these additional variables. Instead, our approach of controlling for scores from multiple subjects—rather than multiple grades—

enabled the use of baseline achievement variables from a single recent baseline grade with lower rates of missing data.

C. Analytic Methods in the Contextual Analyses

We conducted several types of descriptive analyses to develop a fuller understanding of the characteristics, experiences, and institutional environment of the study participants. As discussed next, we focused these analyses on describing the TFA and Teaching Fellows programs, teachers, schools, and students in the study.

1. Descriptive Analyses of TFA and Teaching Fellows Programs

We documented the strategies and methods by which TFA and the Teaching Fellows programs screened and selected candidates, trained and placed their teachers, and supported their teachers during the school year, using data from interviews with program officials. We also examined similarities and differences between TFA and the Teaching Fellows programs as a whole, as well as differences between the individual Teaching Fellows programs in the study.

The information that we collected from program officials included numeric data for several types of program features (for example, how often program staff observed new teachers). When respondents gave a range in response to a question seeking a numeric value, we used the midpoint of the range as the value in the analysis. When respondents reported only a minimum value—for example, that program staff observed new teachers at least once per semester—we used the minimum reported value in the analysis. We gave equal weight to each regional or local program when reporting average characteristics across these programs.

2. Descriptive Analyses of Teacher Characteristics

Documenting the characteristics of each group of study teachers was a key to identifying potential reasons for differences in effectiveness across these groups. On several measures of teachers' professional background and characteristics, we compared (1) TFA teachers with the comparison teachers in the same classroom matches and (2) Teaching Fellows with the comparison teachers in the same classroom matches. For each comparison, we calculated unweighted differences between the groups being compared and tested the statistical significance of the differences.

We also conducted supplemental descriptive analyses in which teachers were weighted so that the TFA/Teaching Fellows and comparison teachers were equally represented in every classroom match. Specifically, for a teacher j belonging to group g (either TFA/Teaching Fellows or comparison), the weight assigned to this teacher, W_j, was

$$W_j = \sum_{k \in M_j} (1/n_{gk}), \quad \text{(B.3)}$$

where M_j denotes the set of matches to which teacher j belonged, and n_{gk} is the number of teachers from group g in match k. For example, if a particular TFA teacher was the only TFA or Teaching Fellows teacher in each of three different matches, then he or she received a weight of 3.0 in this supplemental analysis. If another teacher was one of two comparison teachers in a single match, he or she received a weight of 0.5.

3. Descriptive Analyses of Schools and Students in the Study

We tabulated the average characteristics of schools and students in the study for two main purposes. First, the tabulations characterized the demographic context in which the study was implemented. Second, they helped assess the integrity of random assignment. If random assignment was implemented properly, then there should be few differences in baseline characteristics between students assigned to the treatment group and those assigned to the control group in the same classroom match.

To check the integrity of random assignment, we estimated treatment-control differences on several baseline student characteristics and tested the statistical significance of the differences. These estimates were based on a regression model that controlled for classroom match indicators to ensure that we compared only treatment and control students in the same classroom match. For a given characteristic, z_{ijk}, of student i assigned to teacher j in match k, we used the following ordinary least squares (OLS) regression model:

$$z_{ijk} = \alpha_k + \beta_1 T_{jk} + \varepsilon_{ijk}, \quad \text{(B.4)}$$

where α_k is a classroom match fixed effect; T_{jk} is a binary indicator for being assigned to a TFA or Teaching Fellows teacher (rather than a comparison teacher); and ε_{ijk} is a random error term. The parameter of interest that we estimated, β_1, measures the within-match difference in the specified characteristic between treatment and control group students.

The methods used to estimate Equation (B.4) were identical to those used to estimate the main impacts, discussed later in Section D. In particular, we applied sample weights and calculated standard errors based on the Huber-White sandwich variance estimator, taking into account the clustering of student characteristics at the teacher level (Liang and Zeger 1986). For each characteristic, we used a two-sided *t*-test to assess the statistical significance of the difference between the treatment and control groups.

D. Analytic Methods in the Experimental Analyses

1. Main Estimation Model

In the experimental analyses, we estimated separately the impacts of TFA and Teaching Fellows teachers relative to the comparison teachers with whom they were compared. We used separate samples but identical regression models to estimate the two impacts. We modeled the end-of-year math achievement score, y_{ijk}, of student i who was randomly assigned to teacher j in classroom match k as follows:

$$y_{ijk} = \alpha_k + \beta_1 T_{jk} + \mathrm{X}_{ijk}\beta_2 + \varepsilon_{ijk}, \quad \text{(B.5)}$$

where α_k is a classroom match fixed effect; T_{jk} is a binary indicator for being assigned to a TFA or Teaching Fellows teacher (rather than a comparison teacher); X_{ijk} is a set of baseline student

characteristics specified below; β_1 and β_2 are parameters that we estimated; and ε_{ijk} is a random error term.

The parameter of interest, β_1, is the average difference in effectiveness between TFA or Teaching Fellows teachers and the comparison teachers with whom they were compared. As we did for estimates of treatment-control baseline differences, we conducted the impact estimates in Equation (B.5) using OLS with sample weights, and we calculated sandwich standard errors that accounted for the clustering of student outcomes at the teacher level. By accounting for clustering at the teacher level, we effectively modeled teachers' effects on student achievement as random effects; therefore, the model accommodated the possibility that true differences in effectiveness between TFA/Teaching Fellows teachers and comparison teachers could vary across classroom matches. Moreover, because the sandwich standard errors were robust to any type of correlation in outcomes among students assigned to the same teacher, these standard errors automatically accounted for common influences—for instance, common sets of peers—among students who were assigned to the same classroom as well as scenarios in which students were grouped together in the random assignment process (see Appendix A).

2. Covariates

In the impact estimation, we controlled for several baseline student characteristics. Tables B.1 and B.2 show summary statistics of the covariates in the TFA and Teaching Fellows study samples, respectively.

3. Degrees of Freedom for Hypothesis Tests

We used a two-sided *t*-test to test the null hypothesis, $\beta_1 = 0$. Because sandwich standard errors in finite samples have the potential to lead to overrejection of null hypotheses, we conducted hypothesis tests in the impact analyses based on a *t*-distribution that guarded against overrejection. Following Donald and Lang (2007) and Angrist and Pischke (2009), we used a *t*-distribution with degrees of freedom equal to the number of teachers minus the number of independent variables that varied only at the teacher level, with the teacher-level variables consisting of the treatment status indicator and the classroom match indicators. For these degrees-of-freedom calculations only, we amalgamated classroom matches that shared the same teacher so that the resulting classroom match indicators truly varied only at the teacher level.

4. Handling Missing Data: Main Approach

Our main estimation approach included only students who had valid outcome scores in the analysis sample for the experimental analyses. For randomized controlled trials, simulations in a recent study suggest that impact estimates based only on cases with observed outcome data may have only a small amount of bias (0.05 standard deviations or less) when the outcome data are missing at random among students with the same baseline covariate values (Puma et al. 2009).

Missing covariate data were imputed. For each covariate, we replaced missing values with the mean of the covariate based on research sample students in the same classroom match who had nonmissing values of that covariate. If data on the covariate were missing for all students in a classroom match, missing values were replaced with the covariate mean in the full research sample. In the impact estimation, we controlled for binary indicators that denoted whether a

student had an imputed value for a specified covariate (see Tables B.1 and B.2). Consequently, the impact estimates were unaffected by our choice for the exact numeric values used to replace missing covariate data, so long as these imputed values were constant within classroom matches. Simulations by Puma et al. (2009) have shown that this approach to handling missing covariate data is likely to keep estimation bias at less than 0.05 standard deviations.

Table B.1. Means and Standard Deviations of Covariates Included in Models for Estimating the Impacts of TFA Teachers

	Assigned to TFA Teachers			Assigned to Comparison Teachers				
	Mean	Standard Deviation	N	Mean	Standard Deviation	N	Difference in Means	*p*-Value of Difference
Baseline Achievement								
Baseline math score (*z*-score)	-0.512	0.870	2,073	-0.504	0.853	2,080	-0.008	0.531
Baseline reading score (*z*-score)	-0.514	0.908	2,078	-0.510	0.893	2,090	-0.005	0.734
Demographic Characteristics								
Old for grade	0.073	0.261	2,252	0.064	0.245	2,253	0.009	0.098
Grade is below modal grade in classroom match	0.011	0.113	2,252	0.014	0.117	2,254	-0.003	0.084
Grade is above modal grade in classroom match	0.021	0.143	2,252	0.016	0.127	2,254	0.005	0.160
Retained in same grade between previous and current year	0.022	0.146	2,132	0.024	0.154	2,130	-0.002	0.551
Female	0.486	0.500	2,284	0.500	0.500	2,273	-0.015	0.102
Black, non-Hispanic	0.621	0.487	2,274	0.625	0.484	2,262	-0.004	0.656
Hispanic	0.283	0.452	2,274	0.277	0.448	2,262	0.005	0.471
Non-black, non-Hispanic[a]	0.096	0.300	2,274	0.098	0.297	2,262	-0.002	0.774
Eligible for free/reduced-price lunch	0.899	0.305	1,417	0.905	0.293	1,399	-0.007	0.469
Limited English proficient	0.080	0.276	2,284	0.084	0.277	2,273	-0.004	0.475
Individualized education plan	0.064	0.245	2,284	0.060	0.237	2,273	0.004	0.358
Time Since Baseline Math Test								
1 year[a]	0.814	0.396	2,292	0.808	0.394	2,281	0.005	0.385
2 years	0.072	0.260	2,292	0.082	0.274	2,281	-0.010**	0.006
3 years	0.023	0.161	2,292	0.021	0.143	2,281	0.002	0.470
4 or 5 years	0.001	0.035	2,292	0.000	0.021	2,281	0.001	0.316
Imputation Indicators								
Baseline math score	0.091	0.291	2,292	0.089	0.284	2,281	0.003	0.659
Baseline reading score	0.089	0.288	2,292	0.084	0.278	2,281	0.004	0.400
Old for grade	0.016	0.131	2,292	0.012	0.110	2,281	0.004*	0.018
Current grade	0.016	0.131	2,292	0.012	0.108	2,281	0.004**	0.006
Retained in same grade	0.063	0.252	2,292	0.066	0.249	2,281	-0.003	0.373
Gender	0.003	0.059	2,292	0.003	0.059	2,281	0.000	0.673
Race/ethnicity	0.007	0.086	2,292	0.009	0.094	2,281	-0.002	0.073
Eligible for free/reduced-price lunch	0.377	0.487	2,292	0.377	0.485	2,281	0.000	0.703
Limited English proficient	0.003	0.059	2,292	0.003	0.059	2,281	0.000	0.703
Individualized education plan	0.003	0.059	2,292	0.003	0.059	2,281	0.000	0.703
Total Number of Students in the Analysis Sample			**2,292**			**2,281**		

Source: District administrative records.

Note: Means and differences are adjusted for classroom match fixed effects. *P*-values are based on a regression of the specified variable on a TFA indicator and classroom match indicators, accounting for sample weights and clustering at the teacher level.

[a]Omitted category in regression models.

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

N = number of observations, excluding imputed values; TFA = Teach For America.

Table B.2. Means and Standard Deviations of Covariates Included in Models for Estimating the Impacts of Teaching Fellows

	Assigned to Teaching Fellows			Assigned to Comparison Teachers				
	Mean	Standard Deviation	N	Mean	Standard Deviation	N	Difference in Means	*p*-Value of Difference
Baseline Achievement								
Baseline math score (*z*-score)	-0.111	0.913	1,930	-0.104	0.891	1,795	-0.007	0.636
Baseline reading score (*z*-score)	-0.064	0.984	1,924	-0.048	0.990	1,789	-0.016	0.418
Demographic Characteristics								
Old for grade	0.035	0.174	1,981	0.030	0.172	1,838	0.004	0.175
Grade is below modal grade in classroom match	0.047	0.207	2,106	0.049	0.216	1,964	-0.002	0.581
Grade is above modal grade in classroom match	0.052	0.221	2,106	0.046	0.209	1,964	0.006	0.195
Retained in same grade between previous and current year	0.017	0.124	2,024	0.021	0.144	1,893	-0.004	0.300
Female	0.531	0.499	2,106	0.541	0.498	1,964	-0.010	0.326
Black, non-Hispanic	0.504	0.500	2,106	0.488	0.500	1,964	0.016*	0.047
Hispanic	0.349	0.475	2,106	0.368	0.482	1,964	-0.019*	0.038
Non-black, non-Hispanic[a]	0.147	0.368	2,106	0.144	0.352	1,964	0.002	0.810
Eligible for free/reduced-price lunch	0.737	0.441	2,105	0.759	0.428	1,963	-0.021*	0.017
Limited English proficient	0.077	0.262	2,106	0.072	0.258	1,965	0.005	0.332
Individualized education plan	0.067	0.247	2,106	0.061	0.240	1,964	0.006	0.233
Time Since Baseline Math Test								
1 year[a]	0.607	0.490	2,127	0.614	0.487	1,989	-0.007	0.304
2 years	0.159	0.367	2,127	0.163	0.369	1,989	-0.003	0.425
3 years	0.125	0.339	2,127	0.115	0.319	1,989	0.010**	0.008
4 or 5 years	0.011	0.101	2,127	0.015	0.123	1,989	-0.004	0.116
Imputation Indicators								
Baseline math score	0.098	0.297	2,127	0.094	0.291	1,989	0.004	0.605
Baseline reading score	0.099	0.301	2,127	0.097	0.296	1,989	0.003	0.733
Old for grade	0.067	0.269	2,127	0.065	0.247	1,989	0.002	0.397
Current grade	0.013	0.109	2,127	0.011	0.102	1,989	0.002	0.434
Retained in same grade	0.050	0.221	2,127	0.046	0.209	1,989	0.004	0.417
Gender	0.013	0.109	2,127	0.011	0.102	1,989	0.002	0.434
Race/ethnicity	0.013	0.109	2,127	0.011	0.102	1,989	0.002	0.434
Eligible for free/reduced-price lunch	0.013	0.111	2,127	0.011	0.105	1,989	0.002	0.466
Limited English proficient	0.013	0.109	2,127	0.010	0.100	1,989	0.003	0.312
Individualized education plan	0.013	0.109	2,127	0.011	0.102	1,989	0.002	0.434
Total Number of Students in the Analysis Sample			**2,127**			**1,989**		

Source: District administrative records.

Note: Means and differences are adjusted for classroom match fixed effects. *P*-values are based on a regression of the specified variable on a Teaching Fellows indicator and classroom match indicators, accounting for sample weights and clustering at the teacher level.

[a]Omitted category in regression models.

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

N = number of observations, excluding imputed values.

5. Handling Missing Data: Alternative Approach

In sensitivity analyses reported in Appendix F, we used multiple imputation (Rubin 1987) to replace missing values of the outcome variable and/or covariates in the estimation of Equation (B.5). To generate the imputed data, we conducted multiple imputation by chained equations (MICE) (van Buuren et al. 1999) using data on all research sample students with parental consent. In particular, we specified imputation models in which each variable to be imputed—referred to as a variable of interest—was modeled as a function of predictor variables. The imputation model was linear except when imputing race/ethnicity, an unordered categorical variable, which was imputed with a multinomial logit specification.

The estimated coefficients and standard errors from each imputation model were used to form a posterior distribution for the true coefficients of the imputation model. We made a random draw from this posterior distribution and then used the resulting draw to generate predictions of the variable of interest for all students. For each student with a missing value for the variable of interest, we assigned an imputed value equal to the observed value of the student who had the nearest prediction, in a process known as predictive mean matching. We repeatedly applied this procedure to generate 10 stochastically imputed values for each missing value.

We conducted MICE separately on each of four subsamples: (1) students assigned to TFA teachers, (2) students assigned to comparison teachers who were compared with TFA teachers, (3) students assigned to Teaching Fellows, and (4) students assigned to comparison teachers who were compared with Teaching Fellows. Following standard practice (see, for example, Puma et al. 2009), the imputation model for a given variable of interest used all other variables from Equation (B.5) as predictors. That is, we generated imputed values for the outcome variable based on all covariates, and we generated imputed values for a given covariate based on all other covariates and the outcome variable.

We then estimated Equation (B.5) by using multiply imputed data to replace missing values of (1) covariates only, (2) the outcome variable only, or (3) both the outcome variable and the covariates (see Appendix F). Estimates that used multiply imputed outcome data were no longer limited to analysis sample members; instead, we used all research sample members with parental consent.[26]

6. Estimation Models for Subgroup Analyses

Several of the subgroup analyses assessed the impact of a specified type of TFA/Teaching Fellows teacher compared with two different types of comparison teachers separately. In these cases, we modified Equation (B.5) so that impacts could differ according to the type of comparison teacher. For example, to estimate the impacts of TFA or Teaching Fellows teachers relative to comparison teachers from traditional and alternative certification routes (TC and AC teachers, respectively) separately, we used an estimation model of the form

[26] With Stata, we used the "ice" command (Royston 2004, 2005) to generate multiply imputed data, and we used the "mi estimate" command to produce impact estimates and standard errors with the multiply imputed data based on Rubin's (1987) formulas.

(B.6) $$y_{ijk} = \alpha_k + \delta_1 TC_{jk} + \delta_2 AC_{jk} + \mathrm{X}_{ijk}\beta + \varepsilon_{ijk},$$

where TC_{jk} is an indicator for TC teachers, AC_{jk} is an indicator for AC teachers, and all other variables are defined as in Equation (B.5). The negative of δ_1 and the negative of δ_2 represent the impacts of TFA/Teaching Fellows teachers relative to comparison teachers from, respectively, traditional and alternative routes. We used estimation models of a similar structure when assessing the impacts of TFA/Teaching Fellows teachers with a specified experience level relative to novice and experienced comparison teachers separately.

In other subgroup analyses, we estimated the impacts of TFA/Teaching Fellows teachers relative to comparison teachers separately within middle and high schools. These estimates were based on a variant of Equation (B.5) in which the indicator for TFA/Teaching Fellows teachers was interacted with an indicator for middle schools.

7. Adjustments for Multiple Hypothesis Testing

Because the study tested the statistical significance of several impact estimates—including overall and subgroup-specific impact estimates—the likelihood of finding a difference that was significant at the 5 percent level for any given outcome or subgroup simply due to chance would have been greater than 5 percent if no corrections were made to account for these multiple hypothesis tests. However, adjusting for all hypothesis tests conducted in the study would have severely limited the study's power to detect any impacts at all. To balance these competing factors, we designated the following six impact estimates to be of primary interest, and adjusted significance tests for these estimates to account for multiple hypothesis testing using the approach formulated by Benjamini and Hochberg (1995):

- Impacts of TFA teachers relative to all comparison teachers
- Impacts of TFA teachers relative to TC comparison teachers
- Impacts of TFA teachers relative to AC comparison teachers
- Impacts of Teaching Fellows relative to all comparison teachers
- Impacts of Teaching Fellows relative to TC comparison teachers
- Impacts of Teaching Fellows relative to AC comparison teachers

Across these six impact estimates, our adjustment controlled the false discovery rate—the expected proportion of statistically significant findings that were spurious—at five percent. Because we did not adjust statistical tests in any of the remaining subgroup analyses, there was a greater likelihood that significant findings from these remaining analyses were spurious.

E. Analytic Methods in the Nonexperimental Analyses

1. Model for Estimating Relationships between Teacher Characteristics and Student Achievement

We estimated the relationships between teacher characteristics and student math achievement by augmenting the experimental impact model with measures of teacher characteristics as additional independent variables. The resulting regression model, which we

estimated on the pooled set of observations from the TFA and Teaching Fellows study samples, had the following specification:

$$(B.7) \quad y_{ijk} = \alpha_k + \gamma_1 TFA_{jk} + \gamma_2 Fellow_{jk} + \mathrm{C}_{jk}\gamma_3 + (\mathrm{X}_{ijk} \times TFASam_{ijk})\gamma_4 + (\mathrm{X}_{ijk} \times FellowSam_{ijk})\gamma_5 + \varepsilon_{ijk},$$

where TFA_{jk} and $Fellow_{jk}$ are binary indicators for being assigned to TFA teachers and Teaching Fellows, respectively; C_{jk} is a set of teacher characteristics with coefficients γ_3; $TFASam_{ijk}$ and $FellowSam_{ijk}$ are binary indicators for the TFA and Teaching Fellows study samples, respectively; and all other variables are defined as in Equation (B.5). The teacher characteristics included in C_{jk} are those listed in Table II.7 of Chapter II, and the student-level covariates (X_{ijk}) are the same ones used in the experimental impact models (see Tables B.1 and B.2). Without teacher characteristics in Equation (B.7), estimates of γ_1 and γ_2 would exactly replicate the experimentally estimated impacts of TFA and Teaching Fellows teachers, respectively. As a result, this specification permitted a direct analysis of the extent to which the measured teacher characteristics accounted for the experimental impacts, as discussed later.

The parameters of primary interest were the coefficients in γ_3, which captured the association between teacher characteristics and student math achievement. We estimated variants of Equation (B.7) that included either all teacher characteristics simultaneously or only a single teacher characteristic. In all estimates, we applied sample weights and clustered standard errors at the teacher level.

2. Handling Missing Data in the Measures of Teacher Characteristics

The nonexperimental analyses required more complex methods for handling missing data than the experimental analyses. In the experimental analyses, the independent variable of interest—treatment status—was nonmissing for all students. The other covariates, which had some missing data, were expected to be uncorrelated with the variable of interest due to random assignment, so a relatively simple method—mean imputation along with the inclusion of imputation indicators—could account for missing data in a manner that kept bias low.

In the nonexperimental analyses, the variables of interest—those that measured teacher characteristics—had some missing values. It was therefore important to handle missing data in a manner that preserved the observed distributions of the teacher characteristics and the observed relationships between each characteristic and the other variables in the analysis. Because multiple imputation met these objectives (Schafer and Graham 2002), we applied this procedure to account for missing data in the measures of teacher characteristics.

We imputed missing data in the teacher-level variables through MICE. We used the same general approach to implementing MICE as that used in the sensitivity analyses of the experimental estimates (described in Section D). The procedure generated 10 imputed values for each missing value of a teacher characteristic. Because middle and high school teachers differed in the type of Praxis assessment for which data were needed, we applied MICE separately to four groups of teachers, defined based on state, grade, and entry route: (1) TFA and Teaching Fellows teachers who were supposed to have Praxis II Mathematics Content Knowledge scores;

(2) comparison teachers who were supposed to have Praxis II Mathematics Content Knowledge scores; (3) TFA and Teaching Fellows teachers who were supposed to have Praxis II Middle School Mathematics scores; and (4) comparison teachers who were supposed to have Praxis II Middle School Mathematics scores.

The imputation model for a specified teacher characteristic included all other teacher characteristics as predictors. Every imputation model also included the student outcome variable (y) and the student covariates (X). Because teacher characteristics did not vary within classes, we averaged all student-level variables to the classroom level and estimated the imputation models at the classroom level. However, some teachers in the study taught more than one class; initial imputed values of a particular teacher characteristic could therefore differ across classes taught by the same teacher. To eliminate differences in imputed values for the same teacher within the same imputation round, we took either the mode (for binary variables), median (for ordinal variables), or mean (for continuous variables) of the initial imputed values across classes taught by the same teacher.

3. Specification of Praxis Score Variables

As discussed in Chapter II, the study used two different assessments to measure teachers' knowledge of math concepts: (1) the Praxis II Mathematics Content Knowledge test for high school teachers, and (2) the Praxis II Middle School Mathematics test for middle school teachers. In our main specification, we expressed a teacher's score from a given Praxis II assessment as a z-score based on the mean and standard deviation of scores among comparison teachers who took the same assessment.

To account for the possibility that the relationship between teachers' Praxis II scores and student math achievement could differ by the type of Praxis II assessment that teachers took, we included into Equation (B.7) separate variables for scores from the two Praxis II assessments. The two variables were constructed as interactions between a teacher's z-score and (1) an indicator for whether a teacher was supposed to have taken the Mathematics Content Knowledge assessment, and (2) indicator for whether a teacher was supposed to have taken the Middle School Mathematics assessment. The coefficient on each interaction term captured the relationship between Praxis II scores from the specified assessment and students' math achievement. Because the Praxis type indicators did not vary within classroom matches, there was need to control for these indicators in addition to the classroom match indicators in Equation (B.7).

4. Accounting for the Impacts of TFA Teachers

Based on the estimates of Equation (B.7), we assessed the extent to which differences in effectiveness between TFA and comparison teachers, reported in Chapter V, could be accounted for by the groups' difference in characteristics.[27] Specifically, we calculated the difference in effectiveness between these groups that would be predicted based on each teacher characteristic. This predicted difference was the product of two quantities: (1) the association between the

[27] We did not conduct a parallel analysis for the Teaching Fellows sample because we did not find a statistically significant impact of Teaching Fellows teachers relative to comparison teachers.

characteristic and student achievement and (2) the difference between TFA and comparison teachers with regard to this characteristic.

Formally, for each of the teacher characteristics listed in Table II.7, we calculated the predicted TFA-comparison difference in effectiveness based on that teacher characteristic, as follows. Let $C_{jk}^{(r)}$ be the measure of the rth characteristic for teacher *j*. First, from the estimates of Equation (B.7) that controlled for all teacher characteristics, we obtained the estimated coefficient, $\hat{\gamma}_3^{(r)}$, on $C_{jk}^{(r)}$. Second, to calculate the difference between TFA and comparison teachers on the rth characteristic, we estimated a student-level regression of the form

$$\begin{aligned} C_{jk}^{(r)} &= \alpha_k + \pi_1^{(r)} TFA_{jk} + \pi_2^{(r)} Fellow_{jk} \\ &\quad + (\mathrm{X}_{ijk} \times TFASam_{ijk})\pi_3^{(r)} + (\mathrm{X}_{ijk} \times FellowSam_{ijk})\pi_4^{(r)} + \varepsilon_{ijk} \end{aligned} \tag{B.8}$$

where all variables are defined as in Equation (B.7). The estimated coefficient on the TFA indicator, $\hat{\pi}_1^{(r)}$, represents the difference in the rth characteristic between TFA and comparison teachers. Estimates of Equation (B.8) were based on student-level data rather than teacher-level data in order to provide an exact decomposition of the total TFA impact into explained and unexplained portions, as discussed next.

From the two preceding steps, the predicted difference in effectiveness between TFA and comparison teachers based on the rth characteristic was the product, $\hat{\gamma}_3^{(r)}\hat{\pi}_1^{(r)}$. Mathematically, the predicted TFA-comparison differences in effectiveness based on all measured characteristics, along with the unexplained difference in effectiveness ($\hat{\gamma}_1$) from Equation (B.7), summed to the total TFA-comparison difference in effectiveness from the experimental analyses.

APPENDIX C

SUPPLEMENTARY INFORMATION ON TEACH FOR AMERICA AND TEACHING FELLOWS PROGRAMS

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

This appendix presents supplementary information on Teach For America (TFA) and the Teaching Fellows programs. Figure C.1 lists the 43 TFA regions as of August 2011. Figure C.2 lists TFA's Teaching As Leadership rubric, which includes six principles of leadership, each with specific teacher actions intended to improve student achievement. Figure C.3 lists the 19 Teaching Fellows programs and similar programs affiliated with TNTP as of August 2011. Finally, in support of Chapters III and VI, Table C.1 summarizes how participants in TFA and the Teaching Fellows programs in the study sample were selected, trained, and supported.

Figure C.1. Teach For America Regions, as of August 2011

New England
Connecticut
Greater Boston
Rhode Island

East
Baltimore
D.C. Region
Delaware
Greater Newark
Greater Philadelphia
New York

Southeast
Alabama
Appalachia
Charlotte
Eastern North Carolina
Greater Nashville
Greater New Orleans
Jacksonville
Memphis
Metro Atlanta
Miami-Dade
Mississippi Delta
South Carolina
South Louisiana

Midwest
Chicago
Detroit
Indianapolis
Kansas City
Milwaukee
South Dakota
St. Louis
Twin Cities

Southwest
Dallas-Fort Worth
Houston
New Mexico
Oklahoma
Phoenix
Rio Grande Valley
San Antonio

West
Bay Area
Colorado
Hawai'i
Las Vegas Valley
Los Angeles
Seattle-Tacoma

Source: http://www.teachforamerica.org/where-we-work.

Figure C.2. Teach For America's Teaching as Leadership Rubric

Set Big Goals
- Set ambitious goals for student achievement

Invest Students and Their Families/Influencers in Working Hard to Reach the Big Goals
- Instill "I can" in students
- Instill "I want" in students
- Use role models
- Reinforce academic efforts
- Create a welcoming environment
- Mobilize families and influencers

Plan Purposefully
- Develop assessments
- Create long-term and unit plans (backward design)
- Lesson plan
- Differentiate
- Develop behavioral management plans
- Classroom procedures

Execute Effectively
- Clearly present academic content
- Manage student practice
- Check for understanding
- Reinforce rules and consequences
- Implement time-saving procedures
- Track student performance

Continuously Increase Effectiveness
- Gauge progress and gaps
- Identify contributing student actions
- Identify teacher actions
- Identify underlying factors
- Access learning experiences
- Adjust course

Work Relentlessly
- Persist in the face of challenges
- Expand time and resources
- Sustain energy

Source: http://www.teachingasleadership.org/.

Figure C.3. TNTP's Teaching Fellows Programs, as of August 2011

East Coast
Baltimore City Teaching Residency
DC Teaching Fellows
Delaware Teaching Fellows
New York City Teaching Fellows
Philadelphia Teaching Fellows
Rhode Island Teaching Fellows
The Teacher Academy Residency (Pittsburgh)

Midwest
Chicago Teaching Fellows
Denver Teaching Fellows
Indianapolis Teaching Fellows
Milwaukee Teaching Fellows

South/Southwest
Arizona Teaching Fellows
Georgia Teaching Fellows
Memphis Teaching Fellows
Nashville Teaching Fellows
TeachNOLA (New Orleans)
TEACH Charlotte
Texas Teaching Fellows

West Coast
Oakland Teaching Fellows

Source: http://tntp.org/become-a-teacher/.

Table C.1. Similarities and Differences in TFA and the Teaching Fellows Programs

	TFA	Teaching Fellows Programs
Program Administration and Location	National program; it serves numerous regions around the country	TNTP has set up Teaching Fellows programs in numerous locations around the country; programs have distinct identities, but TNTP oversees them all.
Application Forms and Locality Requests	One form submitted to national program office; applicants may express their interest in being assigned to particular regions; no guarantee that preferences will be met	Each program has its own application form; applicants to programs serving a region with several districts specify locality of interest.
Participant Commitment	2 years	No commitment
Writing Requirements of the Application	Applicants submit a letter of intent explaining, in 500 words or fewer, why they want to join TFA, what they hope to accomplish in the program, and how they would determine their success in the program	Applicants answer three questions (concerning, for example, their interest in the program and their qualifications); each answer should be 200 to 400 words.
Steps in the Application and Selection Processes as of Spring 2010	Online application, telephone interview,[a] in-person assessment	Online application, in-person assessment
Activities in the In-Person Assessment as of Spring 2010	Sample teaching lesson, group discussion, problem-solving activity,[b] one-on-one interview	Sample teaching lesson, writing exercise, group discussion, one-on-one interview
Core Competencies Sought in the Application and Selection Processes	1. Demonstrated leadership and achievement in academic, professional, extracurricular, or volunteer arenas 2. Perseverance and sustained focus in the face of challenges 3. Critical thinking skills 4. Organizational ability 5. Respect for individuals' diverse experiences and ability to work effectively with people from a variety of backgrounds 6. Interpersonal skills to motivate and lead others 7. An understanding of and desire to work relentlessly in pursuit of the organization's vision	1. Critical thinking 2. Achieving ambitious goals 3. Personal responsibility and accountability 4. Commitment to students in high-need schools 5. Dedication to continuous improvement 6. Communication skills 7. Respectful and positive in all situations
Objective Eligibility Criteria for Program Admission and/or District Hiring	1. Hold a bachelor's degree 2. Have a cumulative undergraduate GPA of at least 2.5 3. Meet all applicable state-specified coursework requirements 4. Pass all state-required tests for new teachers 5. Not have been involved in certain criminal proceedings 6. Be a U.S. citizen or national or legal permanent resident	1. Hold a bachelor's degree 2. Not be or have been in process of earning a degree in education or a teaching certificate 3. Meet minimum cumulative undergraduate GPA requirement (varies by program; minimum allowed across programs in the study: 2.5) 4. Meet all applicable state requirements for subject area coursework 5. Pass all state-required tests for new teachers 6. Pass a background check 7. Meet citizenship or residency requirements
Application Review and Selection Process for Deciding Which Candidates to Consider Further and Admit to the Program	Combination of staff judgment and a computerized algorithm	Staff judgment

	TFA	Teaching Fellows Programs
Program Admission and Yield Rates (program-estimated averages over 3 recent years)	50% of all applicants invited to in-person assessment 12% of all applicants offered admission 80% of admitted applicants enrolled	63% of all applicants invited to in-person assessment 13% of all applicants offered admission 77% of admitted applicants enrolled
Pre-Institute Activities	30 to 40 hours of independent study and regional induction of up to one week	25 hours of independent study and a 4-hour orientation meeting
Summer Institute Location	Eight locations around the country; each serves several regional programs	Each program has its own institute(s)
Institute Curriculum	Uniform across locations	Uniform across programs
Other Types of Instructional Activities and Meetings	Uniform across locations; exceptions for new components under development	Variable; adapted to district(s) served
Estimated Hours Spent Receiving Instruction During Institute	109	68
Institute Fieldwork ("practice teaching")	Corps members assigned in groups of 3 to 4 to summer school classrooms, spend estimated 20 hours tutoring small groups, 20 hours leading instruction for whole class	Goal is minimum 15 days in summer school classroom, 5 hours a day; Teaching Fellows are assigned in pairs, take increasing responsibility for instruction, building up to 2 consecutive full days; estimated average total fieldwork: 64 hours
Math Immersion Programs for Individuals Who Might Not Otherwise Qualify to Teach Secondary Math	None at present, but one summer institute offers an online training course to help qualified secondary math teacher candidates hone their math knowledge and skills	Mandatory for certain candidates in three of the Teaching Fellows programs in the study
Removal of Underperforming Participants	Officials do dismiss participants viewed as not well suited for teaching, but reported that data on dismissal rates were not sufficiently reliable (for example, could not definitively distinguish dismissals from departures for other reasons), and would not be made public	Officials do "out-place" participants viewed as not well suited for teaching, but reported it as rarely necessary. Of the approximately 6 percent of Teaching Fellows who did not complete summer institute, fewer than one-fourth were out-placed. Out-placement also occurs during the first year of teaching, but data on frequency were not readily available.
Placement/Hiring for Specific Teaching Jobs	Corps members typically interview with potential hiring principals, but often decisions about where to interview and where corps members will teach are not made by the corps members, but instead made by district, school, or regional program officials	Teaching Fellows typically select, on their own, the schools at which to interview and which hiring offer to accept.
Percentage of Candidates Placed Before Start of School Year (program-estimated averages over 3 recent years)	98% of all candidates 100% of secondary math candidates	92% of all candidates 95% of secondary math candidates
Post-Placement Instruction/Training and Support	Staff provide training on topics such as classroom management, lesson planning, goal setting, student assessment, and pedagogy, guided by corps member's needs. The average amount of training reported across regions was 42 hours. Staff observe corps members in their classrooms; 4 to 5 times over a 2-year period is a commonly reported frequency for these observations. Staff also provide individualized support as needed.	Staff deliver 10 hours of training on practical topics such as classroom management, the use of data to inform instruction, and tailoring instruction for different students. Staff also conduct at least 2 classroom observations of each Teaching Fellow, have at least 2 check-in meetings, and may have additional meetings as needed.

	TFA	Teaching Fellows Programs
Enrollment in Local Alternative Certification Program Required for Secondary Math Teachers	Yes, if state or district requires enrollment for all new teachers[c]	Yes
Operators of Alternative Certification Programs Attended by Participants	Universities, districts, TNTP, TFA	Universities, districts, TNTP
Estimated Total Hours of Instruction that Secondary Math Teachers Received in Focal Alternative Certification Program	Varies by program; among focal programs in the study, range was 54 to more than 315[d]	Varies by program; among focal programs examined in the study, range was 61 to 250
Financial Support for Participants	All participants receive as many AmeriCorps education awards (up to 2) as they are eligible for; TFA covers all institute and orientation costs; need-based no-interest loans and grants can be used to cover certain other costs associated with joining TFA; in some regions, other sources help cover certification program costs.	Varies by program; some arrange for AmeriCorps education awards, some leverage other sources to help cover certification program costs.

[a]Some applicants were allowed to bypass the telephone interview.

[b]Starting in 2011, this component was moved from the in-person assessment to before the telephone interview or before the in-person assessment.

[c]This was not required for all new secondary math teachers in one of the 10 regional programs in the study.

[d]This was reportedly the amount of instruction in the first two years of a three-year program.

GPA = grade point average; TFA = Teach For America.

APPENDIX D

TEACH FOR AMERICA AND TEACHING FELLOWS TEACHERS COMPARED WITH COMPARISON TEACHERS BY ENTRY ROUTE (ALTERNATIVE OR TRADITIONAL)

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

In Chapters IV and VII of this report, we compared Teach For America (TFA) and Teaching Fellows teachers, respectively, with the full set of comparison teachers with whom they were matched. However, the study's comparison teachers included teachers from both traditional (TC) and alternative (AC) routes to certification, two very different pathways into teaching, and we are interested in the impacts of TFA and Teaching Fellows teachers relative to teachers from TC and AC routes separately as well as combined. To provide context for these impact estimates, in this appendix we separately examine the characteristics of TFA and Teaching Fellows teachers relative to the AC and TC comparison teachers with whom they were matched.

A. TFA Teachers and the Comparison Teachers with Whom They Were Matched

As discussed in Chapter IV, for most of the characteristics we examined, patterns of differences between TFA and the comparison teachers with whom they were matched were similar for comparison teachers from both TC and AC routes. This included demographic characteristics (Table D.1), average number of college-level math courses (Table D.3), Praxis II scores (Table D.4), teaching experience (Table D.5), teacher support during the school year (Table D.8), and teacher satisfaction with teaching at his or her school (Table D.9).

However, there were also some differences in TFA comparisons to teachers from AC and TC routes:

- Although patterns related to educational background were generally similar when comparing TFA teachers with comparison teachers from TC and AC routes, TFA teachers were much less likely to have majored or minored in education than TC comparison teachers, whereas none of the AC teachers in the sample or the TFA teachers to whom they were matched had an education major or minor (Table D.2).
- Although TFA teachers had less nonteaching work experience than comparison teachers from both AC and TC routes, the differences were smaller and not consistently statistically significant when comparing TFA teachers with TC comparison teachers (Table D.5).
- TFA teachers generally reported having received less training and support than TC comparison teachers but received similar or larger amounts of training and support than AC comparison teachers (Table D.6).
- Although TFA teachers were more likely to have taken coursework during the school year compared with comparison teachers from both AC and TC routes, differences between TFA teachers and AC comparison teachers were smaller and were not statistically significant (Table D.7).

B. Teaching Fellows and the Comparison Teachers with Whom They Were Matched

Patterns of differences between Teaching Fellows and the comparison teachers with whom they were matched were also similar for comparison teachers from both TC and AC routes for many characteristics, as discussed in Chapter VII. This included average number of college-level math courses (Table D.12), Praxis II scores (Table D.13), teaching experience (Table D.14), and teacher satisfaction with teaching at his or her school (Table D.17).

There were also some differences in Teaching Fellows comparisons with teachers from AC and TC routes:

- Comparison teachers from AC routes were much more likely to be nonwhite or Hispanic than the Teaching Fellows with whom they were matched, whereas the difference between teachers from TC routes and the Teaching Fellows with whom they were matched was smaller, although still statistically significant (Table D.10).
- Teaching Fellows were more likely to have a degree from a selective college or university than TC comparison teachers, whereas the difference between Teaching Fellows and AC comparison teachers was smaller and not statistically significant (Table D.11).
- Teachers from TC routes were significantly more likely to have majored or minored in secondary math education than the Teaching Fellows with whom they were matched, whereas none of the teachers from AC routes or the Teaching Fellows with whom they were matched had a secondary math education major or minor (Table D.11).
- The amount of nonteaching work experience of Teaching Fellows in the study exceeded that of TC comparison teachers but was statistically indistinguishable from that of AC comparison teachers (Table D.14).
- Although neither difference was statistically significant at the 5 percent level, Teaching Fellows matched to TC comparison teachers reported receiving fewer hours of math pedagogy instruction than the comparison teachers, whereas Teaching Fellows matched to AC comparison teachers reported receiving more hours than the comparison teachers (Table D.15).
- Teaching Fellows completed less student teaching in math, on average, than comparison teachers from TC routes, whereas Teaching Fellows completed more days of student teaching in math than comparison teachers from AC routes, although the latter difference was not statistically significant (Table D.15).

Table D.1. Demographic Characteristics of Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

Characteristic	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Age (average years)	24.2	38.3	-14.1**	0.000	24.7	37.3	-12.6**	0.000
Female	59.0	80.0	-21.0*	0.049	58.6	78.6	-20.0	0.108
Race/Ethnicity								
White, Non-Hispanic	79.5	25.7	53.8**	0.000	72.4	17.9	54.6**	0.000
Nonwhite or Hispanic	20.5	74.3	-53.8**	0.000	27.6	82.1	-54.6**	0.000
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

*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.

Table D.2. Educational Background of Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Bachelor's Degree								
From a highly selective college or university[a]	28.2	8.3	19.9*	0.037	31.0	0.0	31.0**	0.001
From a selective college or university[a]	76.9	29.2	47.8**	0.000	86.2	15.0	71.2**	0.000
Major[b]								
Math	12.8	26.1	-13.3	0.223	0.0	25.0	-25.0*	0.015
Secondary math education	0.0	30.4	-30.4**	0.003	0.0	0.0	0.0	n.a.
Other subject	94.9	65.2	29.7**	0.007	100.0	75.0	25.0*	0.015
Major or minor[b]								
Math	12.8	34.8	-22.0	0.060	10.3	25.0	-14.7	0.207
Secondary math education	0.0	39.1	-39.1**	0.000	0.0	0.0	0.0	n.a.
Other subject	97.4	69.6	27.9**	0.008	100.0	80.0	20.0*	0.033
Graduate Degree	43.6	74.3	-30.7**	0.007	34.5	64.3	-29.8*	0.024
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

[a]Highly selective colleges are those ranked by *Barron's* as most competitive. Selective colleges are those that are ranked by *Barron's* as very competitive, highly competitive, or most competitive.

[b]Percentages might not sum to 100 if some sample members had a degree in more than one subject or had a degree in math education, which counts as both a math-related major and an education major.

[c]Includes statistics, engineering, computer science, finance, economics, and physics/astrophysics.

*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.

Table D.3. Average Number of College-Level Math Courses Taken by Teachers in the TFA Study Sample, by Entry Route

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Total Number of College-Level Math Courses	4.4	8.8	-4.4**	0.003	5.1	7.7	-2.6	0.066
Calculus	1.6	2.2	-0.6	0.166	1.7	1.9	-0.2	0.533
Advanced Algebra	0.5	1.5	-1.0**	0.002	0.4	1.1	-0.8**	0.000
Analysis	0.3	1.0	-0.7**	0.001	0.4	0.5	-0.2	0.544
Advanced Geometry/Topology	0.1	0.7	-0.7**	0.000	0.1	0.5	-0.3	0.061
Probability and Statistics	1.2	1.4	-0.2	0.353	1.1	1.6	-0.5	0.171
Discrete Mathematics	0.2	1.2	-1.0**	0.000	0.2	0.9	-0.7**	0.000
Applied Mathematics	0.5	0.8	-0.3	0.275	1.2	1.2	0.1	0.922
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Table D.4. Praxis II Scores of Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Mathematics: Content Knowledge Test								
Average score	161.3	138.8	22.5	0.124	163.0	141.2	21.8*	0.033
Average score of those who passed[a]	165.8	155.0	10.8	0.328	163.0	150.3	12.7	0.130
Sample size	9	5			6	6		
Middle School Mathematics Test								
Average score	181.3	160.7	20.7**	0.000	177.3	155.6	21.7**	0.000
Average score of those who passed[a]	181.3	168.8	12.5**	0.005	181.0	164.3	16.7**	0.007
Sample size	26	21			22	19		

Source: Praxis II scores from study-administered assessment or provided by the Educational Testing Service.

[a]Praxis passing thresholds from 2011. In states in which this test is not required, score is compared against the average passing threshold across all states that require this 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.

Table D.5. Nonteaching and Teaching Work Experience of Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Nonteaching Work Experience								
Had a nonteaching job after college	15.4	37.1	-21.8*	0.035	20.7	75.0	-54.3**	0.000
Years of experience in nonteaching job (average)[a]	0.2	1.4	-1.2	0.053	0.6	4.3	-3.7**	0.001
Used college-level math in nonteaching job[a]	7.7	20.0	-12.3	0.133	13.8	57.1	-43.3**	0.000
Teaching Experience								
Years of teaching experience (average)	1.9	10.9	-9.1**	0.000	1.7	9.1	-7.3**	0.000
1 or 2 years of teaching experience	79.5	8.6	70.9**	0.000	89.7	10.7	78.9**	0.000
3 to 5 years of teaching experience	20.5	17.1	3.4	0.715	10.3	25.0	-14.7	0.153
More than 5 years of teaching experience	0.0	74.3	-74.3**	0.000	0.0	64.3	-64.3**	0.000
Years of experience teaching math (average)	1.8	8.0	-6.2**	0.000	1.7	7.5	-5.8**	0.000
Fewer than 3 years of math teaching experience	84.6	8.6	76.0**	0.000	89.7	10.7	78.9**	0.000
3 to 5 years of math teaching experience	15.4	28.6	-13.2	0.179	10.3	28.6	-18.2	0.086
More than 5 years of math teaching experience	0.0	62.9	-62.9**	0.000	0.0	60.7	-60.7**	0.000
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

[a]Calculations are based on all teachers, regardless of whether they had a nonteaching job after college.

*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.

Table D.6. Training and Professional Development of Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Average Hours of Math Pedagogy Instruction as Part of Teacher Training[a]	39.0	46.3	-7.4	0.367	36.7	25.6	11.1	0.224
Days of Student Teaching in Math as Part of Teacher Training[b]								
No days	15.4	22.9	-7.5	0.423	27.6	78.6	-51.0**	0.000
1 to 20	46.2	28.6	17.6	0.121	48.3	10.7	37.6**	0.001
More than 20	38.5	48.6	-10.1	0.388	24.1	10.7	13.4	0.187
Average days of student teaching in math	20.0	37.7	-17.7**	0.004	16.6	8.4	8.2	0.124
Hours per Day Spent in Student Teaching in Math as Part of Teacher Training (average)[c]	1.3	3.2	-1.9**	0.000	1.3	3.7	-2.4*	0.036
Average Hours Spent in Math Pedagogy Professional Development During School Year[d]	4.6	11.4	-6.9**	0.000	7.8	10.4	-2.6	0.194
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

[a]Teachers were asked, "As part of your training to become a math teacher, did you receive any instruction in math pedagogy or strategies to teach math?" If so, "In total, how many hours of instruction in math pedagogy or strategies to teach math 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 math 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 in which you went to an elementary or secondary school and taught one or more math lessons to a whole classroom of students?" If so, "On approximately how many days, in total, did you teach at least one full math lesson (at least one class period) to a whole classroom of students during your teacher education/preparation program?" Possible responses were none, 1 to 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 in math, we created a continuous variable equal to zero for teachers who did not do any student teaching in math, 80 for those who did more than 80 days, and the midpoint of the range for all other categories.

[c]Calculated only among those teachers who said they did some student teaching in math.

[d]Teachers were asked, "During this school year, did you attend any professional development classes, workshops, or seminars *provided by the school district* in math pedagogy or strategies to teach math?" If so, "In total, how many hours did you spend attending these professional development classes, workshops, or seminars in math pedagogy or strategies to teach math?" 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.

Table D.7. Coursework Taken During the School Year by Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Overall								
Took Coursework Related to Teaching Job	53.8	11.4	42.4**	0.000	41.4	32.1	9.2	0.478
Total Hours Spent During School Year on Coursework (average)	95.3	22.8	72.5**	0.006	71.2	83.8	-12.6	0.751
Hours spent in class during school year on coursework (average)	43.3	12.3	31.0*	0.010	35.1	35.9	-0.8	0.969
Hours spent out of class during school year on coursework (average)	51.9	10.5	41.5*	0.010	36.1	47.9	-11.8	0.590
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

*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.

Table D.8. Mentoring Received During the School Year by Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Had a Mentor During School Year	56.4	28.6	27.8*	0.015	79.3	28.6	50.7**	0.000
Mentor Assigned by School	51.3	25.7	25.6*	0.023	75.9	28.6	47.3**	0.000
Mentor Assigned by Program	30.8	0.0	30.8**	0.000	34.5	0.0	34.5**	0.000
Average Time Spent Observing Other Teachers (minutes)	84.4	83.7	0.6	0.989	104.3	103.9	0.4	0.994
Average Time Spent Being Observed by Mentor (minutes)	10.2	20.3	-10.1	0.418	26.7	8.6	18.1	0.114
Average Time Spent in Formal Meetings with Mentors (minutes)	143.2	45.5	97.7*	0.017	167.1	81.4	85.6	0.090
Average Time Spent in Informal Meetings with Mentors (minutes)	114.0	90.9	23.1	0.727	85.7	51.8	33.9	0.431
Number of Times Received Written Feedback on Teaching Performance	2.0	0.6	1.4*	0.027	2.4	2.9	-0.4	0.767
Felt that Mentoring Was Very Helpful	43.6	22.9	20.7	0.059	55.2	25.0	30.2*	0.020
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

*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.

Table D.9. Job Satisfaction of Teachers in the TFA Study Sample, by Entry Route (percentages unless otherwise indicated)

	TFA Teachers Matched with Teachers from Traditional Routes to Certification				TFA Teachers Matched with Teachers from Alternative Routes to Certification			
Aspect of Job	TFA Teachers	Teachers from Traditional Routes	Difference	*p*-Value	TFA Teachers	Teachers from Alternative Routes	Difference	*p*-Value
Autonomy and Control Over Own Classroom	82.1	82.9	-0.8	0.929	89.7	89.3	0.4	0.965
Students' Motivation to Learn	43.6	38.2	5.4	0.648	37.9	39.3	-1.4	0.918
Students' Discipline and Behavior	34.2	31.4	2.8	0.804	17.2	39.3	-22.0	0.067
Availability of Resources for Classroom	66.7	73.5	-6.9	0.529	60.7	57.1	3.6	0.791
Recognition/Support from Administration	56.4	70.6	-14.2	0.214	62.1	64.3	-2.2	0.865
Influence Over School Policies and Practices	35.9	61.8	-25.9*	0.027	37.9	64.3	-26.4*	0.047
Opportunities for Professional Development	51.3	85.7	-34.4**	0.001	55.2	78.6	-23.4	0.062
Principal's Leadership and Vision	51.3	68.6	-17.3	0.132	58.6	67.9	-9.2	0.478
Procedures for Performance Evaluation	46.2	71.4	-25.3*	0.027	55.2	75.0	-19.8	0.120
Professional Caliber of Colleagues	61.5	80.0	-18.5	0.082	55.2	71.4	-16.3	0.209
Sample Size	**39**	**35**			**29**	**28**		

Source: Survey of Secondary Math Teachers.

Note: Teachers were asked about their satisfaction with each aspect of their job—possible responses were very dissatisfied, somewhat dissatisfied, somewhat satisfied, and very satisfied. Table shows the percentages of teachers who were somewhat or very satisfied with each aspect of their job.

*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.

Table D.10. Demographic Characteristics of Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
Characteristic	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Age (average years)	33.6	40.1	-6.5**	0.008	32.3	43.3	-11.1**	0.001
Female	52.0	60.0	-8.0	0.415	50.0	50.0	0.0	1.000
Race/Ethnicity								
White, Non-Hispanic	63.3	36.4	26.9**	0.006	73.7	20.0	53.7**	0.000
Nonwhite or Hispanic	36.7	63.6	-26.9**	0.006	26.3	80.0	-53.7**	0.000
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

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

Table D.11. Educational Background of Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Bachelor's Degree								
From a selective college or university[a]	71.4	31.7	39.7**	0.000	73.7	40.0	33.7	0.052
Major[b]								
Math	28.6	48.8	-20.2	0.060	15.8	26.7	-10.9	0.461
Secondary math education	0.0	17.1	-17.1**	0.005	0.0	0.0	0.0	n.a.
Math-related subject[c]	33.3	14.6	18.7*	0.046	36.8	46.7	-9.8	0.578
Other subject	66.7	31.7	35.0**	0.001	68.4	46.7	21.8	0.216
Major or minor[b]								
Math	33.3	58.5	-25.2*	0.021	26.3	33.3	-7.0	0.670
Secondary math education	0.0	34.1	-34.1**	0.000	0.0	0.0	0.0	n.a.
Math-related subject[c]	40.5	22.0	18.5	0.070	36.8	53.3	-16.5	0.353
Other subject	81.0	56.1	24.9*	0.015	73.7	46.7	27.0	0.119
Graduate Degree	80.0	78.2	1.8	0.821	90.5	85.0	5.5	0.605
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

[a]Highly selective colleges are those ranked by *Barron's* as most competitive. Selective colleges are those ranked by *Barron's* as very competitive, highly competitive, or most competitive.

[b]Percentages might not sum to 100 if some sample members had a degree in more than one subject or had a degree in math education, which counts as both a math-related major and an education major.

[c]Includes mathematics, secondary mathematics education, statistics, engineering, computer science, finance, economics, and physics/astrophysics.

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

Table D.12. Average Number of College-Level Math Courses Taken by Teachers in the Teaching Fellows Study Sample, by Entry Route

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Total Number of College-Level Math Courses	9.6	11.1	-1.5	0.332	10.6	10.8	-0.2	0.907
Calculus	2.6	3.1	-0.5	0.089	2.6	2.4	0.2	0.635
Advanced Algebra	1.4	1.7	-0.3	0.261	1.2	1.5	-0.3	0.441
Analysis	0.6	0.9	-0.3	0.192	1.0	1.0	0.0	0.996
Advanced Geometry/Topology	0.6	0.9	-0.2	0.219	0.5	0.7	-0.2	0.506
Probability and Statistics	1.7	1.8	-0.1	0.745	2.0	1.5	0.5	0.284
Discrete Mathematics	1.2	1.2	0.0	0.934	1.4	1.2	0.2	0.572
Applied Mathematics	1.5	1.5	0.0	0.957	1.8	2.5	-0.7	0.460
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

Table D.13. Praxis II Scores of Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Mathematics: Content Knowledge Test								
Average score	157.2	140.9	16.3**	0.007	157.9	132.0	25.9**	0.002
Average score of those who passed[a]	163.9	155.0	9.0	0.100	163.8	150.6	13.1	0.149
Sample size	29	34			14	13		
Middle School Mathematics Test								
Average score	187.6	175.7	11.9*	0.024	183.7	156.3	27.3**	0.002
Average score of those who passed[a]	187.6	177.9	9.6*	0.047	183.7	157.8	25.9**	0.005
Sample size	20	15			6	6		

Source: Praxis II scores from study-administered assessment or provided by the Educational Testing Service.

[a]Praxis passing thresholds from 2011. In states in which this test is not required, score is compared against the average passing threshold across all states that require this 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.

Table D.14. Nonteaching and Teaching Work Experience of Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Nonteaching Work Experience								
Had a nonteaching job after college	68.0	32.7	35.3**	0.000	61.9	70.0	-8.1	0.595
Years of experience in nonteaching job (average)[a]	5.2	1.9	3.3*	0.012	4.8	5.9	-1.1	0.610
Used college-level math in nonteaching job[a]	40.0	20.0	20.0*	0.026	52.4	50.0	2.4	0.883
Teaching Experience								
Years of teaching experience (average)	3.7	12.6	-8.9**	0.000	4.6	14.0	-9.4**	0.000
Years of experience teaching math (average)	3.5	11.2	-7.8**	0.000	3.9	9.8	-5.9**	0.000
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

[a]Calculations are based on all teachers, regardless of whether they had a nonteaching job after college.

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

Table D.15. Training and Professional Development of Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Average Hours of Math Pedagogy Instruction as Part of Teacher Training[a]	47.8	50.4	-2.6	0.727	57.5	39.5	17.9	0.127
Participated in Intensive Math Immersion Program as Part of Teacher Training	46.0	5.5	40.5**	0.000	52.4	0.0	52.4**	0.000
Average Days of Student Teaching in Math as Part of Teacher Training[b]								
No days	8.0	9.1	-1.1	0.843	14.3	60.0	-45.7**	0.002
1 to 20	82.0	14.5	67.5**	0.000	66.7	20.0	46.7**	0.002
More than 20	10.0	76.4	-66.4**	0.000	19.0	20.0	-1.0	0.941
Average days of student teaching in math	10.1	46.7	-36.6**	0.000	14.7	13.4	1.3	0.840
Hours per Day Spent in Student Teaching in Math as Part of Teacher Training (average)[c]	2.0	3.1	-1.1**	0.000	2.7	3.3	-0.6	0.397
Average Hours Spent in Math Pedagogy Professional Development During School Year[d]	8.6	4.4	4.2**	0.003	7.9	9.1	-1.2	0.625
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

[a]Teachers were asked, "As part of your training to become a math teacher, did you receive any instruction in math pedagogy or strategies to teach math?" If so, "In total, how many hours of instruction in math pedagogy or strategies to teach math 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 math 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 in which you went to an elementary or secondary school and taught one or more math lessons to a whole classroom of students?" If so, "On approximately how many days, in total, did you teach at least one full math lesson (at least one class period) to a whole classroom of students during your teacher education/preparation program?" Possible responses were none, 1 to 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 in math, we created a continuous variable equal to zero for teachers who did not do any student teaching in math, 80 for those who did more than 80 days, and the midpoint of the range for all other categories.

[c]Calculated only among those teachers who said they did some student teaching in math.

[d]Teachers were asked, "During this school year, did you attend any professional development classes, workshops, or seminars *provided by the school district* in math pedagogy or strategies to teach math?" If so, "In total, how many hours did you spend attending these professional development classes, workshops, or seminars in math pedagogy or strategies to teach math?" 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.

Table D.16. Coursework Taken During the School Year by Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Overall								
Took Coursework Related to Teaching Job	30.0	23.6	6.4	0.468	28.6	20.0	8.6	0.534
Total hours spent during school year on coursework (average)	65.6	32.1	33.5	0.155	55.3	48.8	6.4	0.865
Hours spent in class during school year on coursework (average)	28.8	14.6	14.2	0.154	25.2	19.5	5.7	0.724
Hours spent out of class during school year on coursework (average)	36.8	17.6	19.2	0.176	30.1	29.3	0.8	0.974
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

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

Table D.17. Job Satisfaction of Teachers in the Teaching Fellows Study Sample, by Entry Route (percentages unless otherwise indicated)

	Teaching Fellows Matched with Teachers from Traditional Routes to Certification				Teaching Fellows Matched with Teachers from Alternative Routes to Certification			
Aspect of Job	Teaching Fellows	Teachers from Traditional Routes	Difference	*p*-Value	Teaching Fellows	Teachers from Alternative Routes	Difference	*p*-Value
Autonomy and Control Over Own Classroom	96.0	89.1	6.9	0.177	95.2	90.0	5.2	0.535
Students' Motivation to Learn	38.0	47.3	-9.3	0.342	28.6	40.0	-11.4	0.454
Students' Discipline and Behavior	36.0	52.7	-16.7	0.086	47.6	35.0	12.6	0.424
Availability of Resources for Classroom	66.0	67.3	-1.3	0.891	66.7	75.0	-8.3	0.568
Recognition/Support from Administration	52.0	70.4	-18.4	0.056	52.4	75.0	-22.6	0.138
Influence Over School Policies and Practices	42.0	56.4	-14.4	0.144	35.0	65.0	-30.0	0.060
Opportunities for Professional Development	62.0	64.8	-2.8	0.769	61.9	68.4	-6.5	0.675
Principal's Leadership and Vision	60.0	76.4	-16.4	0.074	52.4	75.0	-22.6	0.138
Procedures for Performance Evaluation	54.0	61.8	-7.8	0.423	66.7	70.0	-3.3	0.824
Professional Caliber of Colleagues	70.0	83.6	-13.6	0.102	81.0	85.0	-4.0	0.738
Sample Size	**50**	**55**			**21**	**20**		

Source: Survey of Secondary Math Teachers.

Note: Teachers were asked about their satisfaction with each aspect of their job—possible responses were very dissatisfied, somewhat dissatisfied, somewhat satisfied, and very satisfied. Table shows the percentages of teachers who were somewhat or very satisfied with each aspect of their job.

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

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

APPENDIX E

SUPPLEMENTARY INFORMATION ON TEACH FOR AMERICA AND TEACHING FELLOWS TEACHERS COMPARED WITH COMPARISON TEACHERS

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

In this appendix, we examine the sensitivity of our analysis of study teachers' scores on the Praxis II math assessments. As discussed in the main text of the report, response rates on the Praxis II differed between TFA and comparison teachers (Chapter IV) and between Teaching Fellows and comparison teachers (Chapter VII). Ninety-one percent of the TFA teachers in the sample had Praxis II scores, compared with only 73 percent of comparison teachers. Similarly, 94 percent of the Teaching Fellows teachers in the sample had Praxis II scores, compared with only 81 percent of comparison teachers (Appendix Table A.18). In our main analyses, we omitted teachers with missing scores. However, these analyses could be misleading if teachers who did not take the assessment would have systematically scored lower (or higher) than their counterparts who did take the assessment. To examine the sensitivity of our findings, we examined differences in Praxis II scores when we imputed values that would otherwise be missing. We used the same imputed values as those that we used in the nonexperimental analysis of Chapter IX, based on the multiple imputation procedure described in Appendix B.

- Table E.1 presents the sensitivity analysis of estimated differences in Praxis II scores between TFA and comparison teachers.
- Table E.2 presents the sensitivity analysis of estimated differences in Praxis II scores between Teaching Fellows and comparison teachers.

For both the TFA and Teaching Fellows samples, results were similar under the two approaches, although differences between the two groups were somewhat more pronounced. This suggests that teachers with missing Praxis data might have been those who would have scored lower had they taken the test.

Table E.1. Sensitivity of Differences in Praxis II Scores Between TFA and Comparison Teachers in the Study

	TFA Teachers	Comparison Teachers	Difference	*p*-Value
Main Estimates				
Mathematics: Content Knowledge Test				
Average score	162.0	140.1	21.9*	0.010
Sample size	15	11		
Middle School Mathematics Test				
Average score	179.8	158.3	21.6**	0.000
Sample size	45	40		
Multiple Imputation of Missing Scores				
Mathematics: Content Knowledge Test				
Average score	178.9	152.3	26.6*	0.002
Sample size	20	20		
Middle School Mathematics Test				
Average score	174.6	152.3	22.3*	0.000
Sample size	50	58		

Source: Praxis II scores from study-administered assessment or provided by the Educational Testing Service.

[a]Praxis passing thresholds from 2011. In states in which this test is not required, the score is compared against the average passing threshold across all states that require this test.

*Difference between TFA and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between TFA and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

TFA = Teach For America.

Table E.2. Sensitivity of Differences in Praxis II Scores Between Teaching Fellows and Comparison Teachers in the Study

	Teaching Fellows	Comparison Teachers	Difference	*p*-Value
Main Estimates				
Mathematics: Content Knowledge Test				
Average score	157.8	138.5	19.3**	0.000
Sample size	39	47		
Middle School Mathematics Test				
Average score	186.7	170.1	16.5**	0.001
Sample size	26	21		
Multiple Imputation of Missing Scores				
Mathematics: Content Knowledge Test				
Average score	168.1	145.9	22.3*	0.000
Sample size	52	70		
Middle School Mathematics Test				
Average score	163.1	145.9	17.3*	0.001
Sample size	28	25		

Source: Praxis II scores from study-administered assessment or provided by the Educational Testing Service.

[a]Praxis passing thresholds from 2011. In states in which this test is not required, the score is compared against the average passing threshold across all states that require this test.

*Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.05 level, two-tailed test.
**Difference between Teaching Fellows and comparison teachers is statistically significant at the 0.01 level, two-tailed test.

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

APPENDIX F

SUPPLEMENTARY ANALYSES OF THE IMPACTS OF TEACH FOR AMERICA AND TEACHING FELLOWS TEACHERS

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

In Appendix F, we present three types of supplemental findings on the impacts of Teach For America (TFA) and Teaching Fellows teachers relative to comparison teachers: (1) sensitivity analyses that assess the robustness of the main impact estimates, (2) analyses of impacts on student attendance, and (3) subgroup analyses that assess differences in effectiveness between specific types of teachers.

A. Supplemental Findings on the Impacts of TFA Teachers

1. Sensitivity Analyses

We explored various changes, described below, to the estimation sample or regression model for estimating the impacts of TFA teachers relative to comparison teachers. Findings from all specifications were similar to the main findings; estimated impacts of TFA teachers ranged from 0.06 to 0.11 standard deviations and were always statistically significant.

Excluding covariates. Our main estimation model controlled for classroom match indicators and all covariates in Appendix Table B.1. Findings were similar when only classroom match indicators, but no other covariates, were included in the estimation model (Table F.1, model 1).

Averaging impacts across classroom matches. In our main specification, sample weights were normalized so that the weights within each match summed to the number of research sample students in the match. Therefore, matches with more students received more weight in the main impact estimates. An alternative normalization for the sample weights in which every match had the same total weight in the final analysis sample produced similar findings (Table F.1, model 2).

Accounting for missing data and nonresponse. For our main impact estimates, we used an analysis sample that excluded students with missing outcome data; we imputed missing covariate data with match-specific or full-sample variable means and included binary indicators for imputed covariate data in the estimation model. In sensitivity analyses, we produced alternative impact estimates based on (1) all research sample students with parental consent, with missing outcome data replaced by multiply imputed values; (2) the main analysis sample, with missing covariate data replaced by multiply imputed values; and (3) all research sample students with parental consent, with missing data for both outcomes and covariates replaced by multiply imputed values (see Appendix B). All findings from these analyses were similar to the main findings (Table F.1, models 3 through 5).

The main analysis sample also excluded high school students whose scores on the Northwest Evaluation Association (NWEA) assessment were deemed invalid due to an unreasonably short testing duration (fewer than six minutes) or an excessively high standard error of measurement (greater than 5.5 Rasch units). However, findings were similar when these invalid scores were included in the analysis (Table F.1, model 6).

Dropping classroom matches. Our main analysis used all classroom matches for which the participating schools sent class rosters verifying that students had been placed into their randomly assigned classes shortly after random assignment occurred at the beginning of the school year. All matches were retained in the main analysis regardless of how many students switched into or out of the matched classes *after* the beginning of the school year. In a sensitivity

analysis, we dropped matches in which students who entered the study classes through a method other than random assignment constituted more than 10 percent of students enrolling before the end of the first month of school or more than 25 percent of students on end-of-year class rosters; findings from this analysis were similar to the main findings (Table F.1, model 7).

In another sensitivity analysis, we excluded classroom matches in which any students received supplemental math instruction outside of the classes that constituted the match. As discussed in Appendix A, supplemental math classes had the potential to either reinforce or dilute the impacts that students experienced from the classes in the study. After dropping matches with supplemental math classes, the estimated impact of TFA teachers remained positive and statistically significant, with a slightly larger magnitude (0.11 standard deviations) than the main impact estimate (Table F.1, model 8).

Standardizing test scores. For the main analysis, we standardized outcome and baseline scores into *z*-scores based on means and standard deviations in a statewide or national reference population. The reference population was either the statewide student population at the same grade level during the year of the test (if the test was a state assessment) or the national norming sample of the test (if the test was an NWEA assessment). Impact findings were similar when the *z*-scores were standardized based on control group members in the same state and grade during the year of the test (for state assessments) or control group members in the same state who took the same math course (for NWEA assessments) (Table F.1, model 9).

Adjusting for noncompliance. Our main model estimated the intent-to-treat (ITT) impact—the impact of being randomly assigned to a TFA teacher rather than a comparison teacher. However, some students did not stay with the type of teacher to which they were originally assigned, and outcomes for these students did not reflect the effectiveness of their assigned teachers. Only for compliers—those who would have stayed with their assigned type of teacher in any assignment scenario—did the impact of *assignment* to TFA teachers represent the actual impact of being *taught* by TFA teachers. In a sensitivity analysis, we isolated the ITT impact pertaining only to compliers, referred to as the complier average causal effect (CACE), capturing the full difference in effectiveness between TFA and comparison teachers for this subset of students.

To estimate the CACE, we used instrumental variables estimation (Angrist et al. 1996). We estimated a regression for end-of-year math scores in which the independent variable of interest was a binary variable for being enrolled in a TFA teacher's class (rather than a non-TFA teacher's class) based on spring rosters, near the time of testing. An indicator for having been randomly assigned to a TFA teacher served as the instrumental variable for being enrolled in a TFA teacher's class.[28] We controlled for classroom match indicators and all covariates in Appendix Table B.1.

[28] The causal validity of the CACE estimate depends on the validity of the exclusion restriction—the assumption that a student's original assignment could affect his or her outcomes only by influencing the type of teacher (TFA or comparison) whom the student had at the time of testing. The exclusion restriction might not hold for students who crossed over to the opposite type of teacher but who spent a significant amount of time with their originally assigned teacher before crossing over. These students could be affected by *both* their originally assigned teacher and the teacher they had at the time of testing. We do not know the magnitude of this potential bias because

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 CACE. First, we assumed that all students who left the study classes moved to a class taught by the same type of teacher 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.

Consistent with the main results, TFA teachers were more effective than comparison teachers at teaching math to students who complied with their original assignment. Compliers taught by TFA teachers scored 0.08 to 0.10 standard deviations higher than those taught by comparison teachers (Table F.1, models 10 and 11).

2. Analyses of Impacts on Student Absences

A potential channel through which teachers could influence student achievement is by affecting their students' rate of absences. To explore this possibility, we estimated the impact on student absences of TFA teachers relative to comparison teachers. We examined two measures of absences: (1) whether a student had been absent from his or her math class more than 25 percent of the time, as reported by the student's teacher, and (2) the number of days a student was absent from school, as recorded in district administrative records. In the full TFA study sample as well as in middle schools and high schools separately, there was no statistically significant difference between students of TFA teachers and students of comparison teachers in either of the two measures of student absences (Table F.2).

3. Subgroup Analyses

Table F.3 presents detailed results for the subgroup analyses discussed in Chapter V. We show estimates of differences in effectiveness between TFA and comparison teachers within subgroups of matches defined by the certification route of the comparison teacher, the experience level of the TFA and/or comparison teacher, and the level of the school (middle or high) in which the teachers taught. We found positive impacts of TFA teachers within each subgroup examined, ranging from 0.06 to 0.13 standard deviations.

(continued)

we do not have precise information on the exact time when students crossed over. Nevertheless, the magnitude of the potential bias depends on the percentage of students who crossed over. This percentage ranges from 2.5 to 13.0 percent in the TFA study sample, depending on our assumptions for whether students who left the study classes were taught by the same or opposite type of teacher to their originally assigned teacher. The corresponding percentages in the Teaching Fellows study sample range from 2.4 to 5.4 percent.

B. Supplemental Findings on the Impacts of Teaching Fellows

1. Sensitivity Analyses

We explored various changes to the estimation sample or regression model for estimating the impacts of Teaching Fellows relative to comparison teachers. Findings from all specifications were similar to the main findings; across all specifications, estimated impacts of Teaching Fellows were never statistically significant and ranged from -0.03 to 0.03.

The types of sensitivity analyses that we conducted for estimating impacts of Teaching Fellows were identical to the analyses we described in Section A for estimating impacts of TFA teachers. As shown in Table F.4, our basic finding—that Teaching Fellows were statistically indistinguishable from comparison teachers in their effectiveness—held in all sensitivity analyses, including the following:

- A regression model that excluded all covariates except classroom match indicators
- Sample weights that gave equal total weight in the analysis sample to each classroom match
- Use of multiply imputed data to replace missing values of the outcome variable and/or covariates
- Inclusion of NWEA scores that were deemed invalid
- Exclusion of classroom matches with high rates of nonrandom student entry into the study classes
- Exclusion of classroom matches with supplemental math instruction
- Use of achievement *z*-scores based on means and standard deviations in the control group sample
- Adjustment for noncompliance with original random assignment

2. Analyses of Impacts on Student Absences

As in the analysis of TFA teachers' impacts, we estimated impacts on student absences of Teaching Fellows relative to comparison teachers. Students of Teaching Fellows and those of comparison teachers were statistically indistinguishable in the average number of days absent from school, as recorded in district administrative records (Table F.5). However, students of Teaching Fellows were, on average, two percentage points less likely than students of comparison teachers to be chronically absent—absent more than one-fourth of the time—from their math class, as reported by the students' math teachers. The reduction in chronic absenteeism for students of Teaching Fellows was observed in high school but not in middle school.

3. Subgroup Analyses

Table F.6 presents detailed results for the subgroup analyses discussed in Chapter VIII. We show estimates of differences in effectiveness between Teaching Fellows and comparison teachers within subgroups of matches defined by the certification route of the comparison teacher, the experience level of the Teaching Fellow and/or comparison teacher, and the level of

the school (middle or high) in which the teachers taught. Impacts were not statistically significant for the subgroup of Teaching Fellows compared with teachers from traditional routes to certification but were positive and significant for the subgroup compared with teachers from alternative routes to certification (impact = 0.13). Within subgroups defined by whether teachers were novice (in their first three years of teaching) or experienced (in their fourth or more year of teaching), impacts were positive and significant for novice Teaching Fellows compared with novice comparison teachers (impact = 0.13) but were negative and significant for novice Teaching Fellows compared with experienced comparison teachers (impact = -0.10). There was no significant impact of experienced Teaching Fellows compared with experienced comparison teachers, and no significant impact of Teaching Fellows compared with comparison teachers whose teaching experience differed by no more than two years. Moreover, there were no significant impacts among subgroups at the middle and high school levels.

Table F.1. Differences in Effectiveness Between TFA Teachers and Comparison Teachers, Sensitivity Analyses

Model	Mean Student Math Achievement (*z*-scores)		Difference in Math Achievement Between Groups (*z*-scores)		Sample Sizes			
	Assigned to TFA Teachers	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
Main Model[a]	-0.52	-0.60	0.07**	0.000	4,573	111	136	45
Alternative Specifications								
Covariates								
(1) No covariates except classroom match indicators	-0.53	-0.60	0.07**	0.001	4,573	111	136	45
Method of Averaging Impacts Across Classroom Matches								
(2) Weight each classroom match equally	-0.56	-0.64	0.08**	0.000	4,573	111	136	45
Accounting for Missing Data and Nonresponse								
(3) Missing outcome test scores imputed with multiple imputation	-0.57	-0.66	0.09**	0.000	5,462	111	136	45
(4) Missing covariate values imputed with multiple imputation	-0.52	-0.60	0.08**	0.000	4,573	111	136	45
(5) Missing outcome test scores and covariate values imputed with multiple imputation	-0.58	-0.66	0.08**	0.000	5,462	111	136	45
(6) Include NWEA scores that were deemed invalid due to high standard error of measurement or short testing duration	-0.52	-0.60	0.07**	0.000	4,619	111	136	45
Rule for Dropping Classroom Matches								
(7) Drop classroom matches in which percentage assigned nonrandomly in first month > 10% or percentage on final roster who had entered nonrandomly > 25%	-0.43	-0.49	0.06*	0.012	3,434	78	112	39
(8) Drop classroom matches with supplemental math classes[b]	-0.53	-0.63	0.11**	0.000	2,460	58	72	30

Table F.1 *(continued)*

Model	Mean Student Math Achievement (*z*-scores)		Difference in Math Achievement Between Groups (*z*-scores)		Sample Sizes			
	Assigned to TFA Teachers	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
Standardization of Test Scores								
(9) Compute *z*-scores based on control group mean and standard deviation within each state and subject	0.08	-0.01	0.09**	0.000	4,573	111	136	45
Adjustment for Students who Switched to a Different Type of Teacher than their Originally Assigned Teacher								
(10) Complier average causal effect, assuming that all students who left study classes stayed with the type of teacher with which they were last observed	-0.51	-0.59	0.08**	0.000	4,573	111	136	45
(11) Complier average causal effect, assuming that all students who left study classes switched to the opposite teacher type	-0.43	-0.53	0.10**	0.000	4,573	111	136	45

Sources: District administrative records and study-administered NWEA assessments.

Note: The difference between groups is adjusted for classroom match fixed effects and all covariates in the specified model. The mean outcome for students assigned to TFA teachers is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups.

[a]The main model adjusts for the baseline student characteristics shown in Appendix Table B.1. Missing values of covariates are imputed with sample means in the same classroom match; if the sample mean of a covariate in a particular classroom match is missing, then missing values are imputed with the sample mean in the full research sample. Covariate imputation indicators are included in the regression model. Outcome data are not imputed—students with missing values of outcome data are excluded from the analysis. Weights are scaled so that the weights for research sample members in each classroom match sum to the number of research sample members in the match. Students with missing outcome test scores or invalid NWEA scores are excluded. All classroom matches in which random assignment was implemented are included. *Z*-scores are computed using state means and standard deviations by grade level for middle school students (who took state assessments) and national means and standard deviations by subject for high school students (who took the study-administered NWEA assessments). The main model does not adjust for students who switched to a different type of teacher than their originally assigned teacher.

[b]This model also retains classroom matches for which there is no information on participation in supplemental math classes.

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

NWEA = Northwest Evaluation Association; TFA = Teach For America.

Table F.2. Impact on Student Absences of TFA Teachers Relative to Comparison Teachers

Measure of Student Absences	Rate of Student Absences		Difference Between Groups		Sample Sizes			
	Assigned to TFA Teachers	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
All Students								
Absent from Math Class More Than 25 Percent of the Time (percentage)	13.1	12.3	0.8	0.528	3,149	87	111	41
Number of Days Absent from School (average)	10.8	11.0	-0.1	0.658	4,601	103	124	39
Middle School Students								
Absent from Math Class More Than 25 Percent of the Time (percentage)	9.3	9.3	0.1	0.964	2,184	67	85	28
Number of Days Absent from School (average)	9.6	9.3	0.3	0.197	3,352	80	99	29
High School Students								
Absent from Math Class More Than 25 Percent of the Time (percentage)	21.3	18.9	2.4	0.477	965	20	26	13
Number of Days Absent from School (average)	14.6	15.6	-1.0	0.283	1,249	23	25	10

Sources: Percentage of days absent from math class is based on teacher reports; days absent from school are based on district administrative records.

Note: Means and differences are adjusted for classroom match fixed effects and all covariates listed in Appendix Table B.1.

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

TFA = Teach For America.

Table F.3. Differences in Effectiveness Between TFA Teachers and Comparison Teachers, by Subgroup

Type of Comparison	Mean Student Math Achievement (*z*-scores)		Difference in Math Achievement Between Groups (*z*-scores)		Sample Sizes			
	Assigned to TFA Teachers	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
Subgroups Defined by Certification Route of Comparison Teacher								
TFA versus TC comparison teachers	-0.52	-0.58	0.06*	0.025	2,477	58	82	34
TFA versus AC comparison teachers	-0.52	-0.62	0.09**	0.001	2,096	53	58	17
Difference between subgroups			-0.03	0.351	4,573	111	136	45
Subgroups with Novice TFA Teachers								
Novice TFA versus novice comparison teachers	-0.24	-0.32	0.08**	0.006	710	18	23	10
Novice TFA versus experienced comparison teachers	-0.59	-0.66	0.07**	0.001	3,642	85	107	36
Difference between subgroups			0.01	0.728	4,352	103	127	43
Subgroups with Experienced TFA Teachers (suppressed[a])								
Subgroups Defined by School Level								
Middle school	-0.47	-0.52	0.06**	0.007	3,373	83	103	31
High school	-0.69	-0.82	0.13**	0.002	1,200	28	33	14
Difference between subgroups			-0.07	0.090	4,573	111	136	45

Sources: District administrative records and study-administered Northwest Evaluation Association assessments.

Note: The difference between groups is adjusted for classroom match fixed effects and all covariates in Appendix Table B.1. The mean outcome for students assigned to TFA teachers is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups. Hypothesis tests within subgroups defined by the certification route of the comparison teacher were adjusted for multiple hypothesis testing, as described in Appendix B; the remaining hypothesis tests in this table were not adjusted for multiple hypothesis testing.

[a]Due to small sample sizes, estimates did not meet the minimum level of precision required for a subgroup estimate to be presented in this report.

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

AC = alternative route to certification; TC = traditional route to certification; TFA = Teach For America.

Table F.4. Differences in Effectiveness Between Teaching Fellows and Comparison Teachers, Sensitivity Analyses

Model	Mean Student Math Achievement (*z*-scores)		Difference in Math Achievement Between Groups (*z*-scores)		Sample Sizes			
	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
Main Model[a]	-0.39	-0.39	0.00	0.956	4,116	118	153	44
Alternative Specifications								
Covariates								
(1) No covariates except classroom match indicators	-0.40	-0.39	-0.01	0.820	4,116	118	153	44
Method of Averaging Impacts Across Classroom Matches								
(2) Weight each classroom match equally	-0.55	-0.52	-0.03	0.147	4,116	118	153	44
Accounting for Missing Data and Nonresponse								
(3) Missing outcome test scores imputed with multiple imputation	-0.51	-0.50	-0.01	0.815	5,313	118	153	44
(4) Missing covariate values imputed with multiple imputation	-0.39	-0.39	0.01	0.800	4,116	118	153	44
(5) Missing outcome test scores and covariate values imputed with multiple imputation	-0.51	-0.50	0.00	0.963	5,313	118	153	44
(6) Include NWEA scores that were deemed invalid due to high standard error of measurement or short testing duration	-0.38	-0.39	0.01	0.736	4,188	118	153	44
Rule for Dropping Classroom Matches								
(7) Drop classroom matches in which percentage assigned nonrandomly in first month > 10% or percentage on final roster who had entered nonrandomly > 25%	-0.26	-0.28	0.03	0.390	3,246	91	113	32
(8) Drop classroom matches with supplemental math classes[b]	-0.38	-0.37	-0.01	0.751	3,208	95	125	37

Table F.4 *(continued)*

Model	Mean Student Math Achievement (*z*-scores)		Difference in Math Achievement Between Groups (*z*-scores)		Sample Sizes			
	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
Standardization of Test Scores								
(9) Compute *z*-scores based on control group mean and standard deviation within each state and subject	0.02	0.02	0.00	0.884	4,116	118	153	44
Adjustment for Students who Switched to a Different Type of Teacher than their Originally Assigned Teacher								
(10) Complier average causal effect, assuming that all students who left study classes stayed with the type of teacher with which they were last observed	-0.38	-0.38	0.00	0.955	4,116	118	153	44
(11) Complier average causal effect, assuming that all students who left study classes switched to the opposite teacher type	-0.34	-0.34	0.00	0.955	4,116	118	153	44

Sources: District administrative records and study-administered NWEA assessments.

Note: The difference between groups is adjusted for classroom match fixed effects and all covariates in the specified model. The mean outcome for students assigned to Teaching Fellows is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups.

[a]The main model adjusts for the baseline student characteristics shown in Appendix Table B.2. Missing values of covariates are imputed with sample means in the same classroom match; if the sample mean of a covariate in a particular classroom match is missing, then missing values are imputed with the sample mean in the full research sample. Imputation indicators are included in the regression model. Weights are scaled so that the weights for research sample members in each classroom match sum to the number of research sample members in the match. Students with missing outcome test scores or invalid NWEA scores are excluded. All classroom matches in which random assignment was implemented are included. *Z*-scores are computed using state means and standard deviations by grade level for middle school students (who took state assessments) and national means and standard deviations by subject for high school students (who took the study-administered NWEA assessments). The main model does not adjust for students who switched to a different type of teacher than their originally assigned teacher.

[b]This model also retains classroom matches for which there is no information on participation in supplemental math classes.

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

NWEA = Northwest Evaluation Association.

Table F.5. Impact on Student Absences of Teaching Fellows Relative to Comparison Teachers

Measure of Student Absences	Rate of Student Absences		Difference Between Groups		Sample Sizes			
	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
All Students								
Absent from Math Class More Than 25 Percent of the Time (percentage)	17.2	19.7	-2.4*	0.047	3,346	96	124	38
Number of Days Absent from School (average)	18.0	17.7	0.3	0.405	5,061	115	149	43
Middle School Students								
Absent from Math Class More Than 25 Percent of the Time (percentage)	5.1	3.9	1.1	0.288	968	25	36	12
Number of Days Absent from School (average)	12.0	11.7	0.3	0.551	1,627	37	53	17
High School Students								
Absent from Math Class More Than 25 Percent of the Time (percentage)	22.6	26.5	-3.9*	0.020	2,378	71	89	28
Number of Days Absent from School (average)	20.9	20.7	0.2	0.557	3,434	78	97	28

Sources: Percentage of days absent from math class is based on teacher reports; days absent from school are based on district administrative records.

Note: Means and differences are adjusted for classroom match fixed effects and all covariates listed in Appendix Table B.2.

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

Table F.6. Differences in Effectiveness Between Teaching Fellows and Comparison Teachers, by Subgroup

Type of Comparison	Mean Student Math Achievement (*z*-scores)		Difference in Math Achievement Between Groups (*z*-scores)		Sample Sizes			
	Assigned to Teaching Fellows	Assigned to Comparison Teachers	Estimate	*p*-Value	Students	Classroom Matches	Teachers	Schools
Subgroups Defined by Certification Route of Comparison Teacher								
Teaching Fellows versus TC comparison teachers	-0.36	-0.32	-0.03	0.248	3,268	92	113	33
Teaching Fellows versus AC comparison teachers	-0.50	-0.63	0.13*	0.014	902	30	46	19
Difference between subgroups			-0.17**	0.005	4,116	118	153	44
Subgroups with Novice Teaching Fellows								
Novice Teaching Fellows versus novice comparison teachers	-0.40	-0.53	0.13**	0.004	354	12	17	7
Novice Teaching Fellows versus experienced comparison teachers	-0.63	-0.53	-0.10**	0.002	1,153	38	53	19
Difference between subgroups			0.24**	0.000	1,475	47	68	23
Subgroups with Experienced Teaching Fellows								
Experienced Teaching Fellows versus novice comparison teachers (suppressed[a])								
Experienced Teaching Fellows versus experienced comparison teachers	-0.27	-0.30	0.03	0.446	2,408	65	80	26
Teaching Fellows Versus Comparison Teachers Whose Levels of Teaching Experience Differ by No More than Two Years	-0.17	-0.20	0.03	0.399	1,283	33	46	17
Subgroups Defined by School Level								
Middle school	-0.35	-0.39	0.04	0.377	1,610	37	53	17
High school	-0.41	-0.39	-0.02	0.472	2,506	81	101	29
Difference between subgroups			0.06	0.253	4,116	118	153	44

Sources: District administrative records and study-administered Northwest Evaluation Association assessments.

Note: The difference between groups is adjusted for classroom match fixed effects and all covariates in Appendix Table B.2. The mean outcome for students assigned to Teaching Fellows is calculated as the unadjusted mean outcome for students assigned to comparison teachers plus the adjusted difference in outcomes between the two groups. Hypothesis tests within subgroups defined by the certification route of the comparison teacher were adjusted for multiple hypothesis testing, as described in Appendix B; the remaining hypothesis tests in this table were not adjusted for multiple hypothesis testing.

Table F.6. *(continued)*

[a]Due to small sample sizes, estimate did not meet the minimum level of precision required for a subgroup estimate to be presented in this report.

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

AC = alternative route to certification; TC = traditional route to certification.

APPENDIX G

SUPPLEMENTARY FINDINGS ON FACTORS ASSOCIATED WITH TEACHER EFFECTIVENESS

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

In Appendix G, we present supplemental tables that document the relationships between teacher characteristics and student math achievement from the nonexperimental analyses. The tables in Section A contain detailed results that form the basis for the nonexperimental findings reported in Chapter IX. The tables in Section B contain sensitivity analyses that assess the robustness of the main nonexperimental findings.

A. Detailed Results from the Main Nonexperimental Analyses

1. Teacher Characteristics in the Analysis

Table G.1 lists the measures of teachers' academic ability, exposure to and knowledge of math, instructional training, and teaching experience that constituted the main set of teacher-level variables included in the analysis. Means and standard deviations are provided for each characteristic.

2. Estimates of Relationships Between Teacher Characteristics and Student Achievement

To assess the relationships between teacher characteristics and student math achievement, we estimated a student-level regression for students' end-of-year math scores. Independent variables consisted of classroom match indicators, separate indicators for Teach For America (TFA) and Teaching Fellows teachers, all student-level baseline covariates, and measures of teacher characteristics (see Appendix B).

Earlier in the report, in Chapter IX, Table IX.1 reported coefficients on the teacher characteristics from the regression estimates. In this appendix, Tables G.2 and G.3 display these coefficients again but also report their corresponding *p*-values as well as additional coefficients and *p*-values pertaining to the TFA and Teaching Fellows indicators. Each of the models in Table G.2 included a single teacher characteristic (or a group of variables measuring a single teacher characteristic, such as teaching experience). The single model shown in Table G.3 included all of the main teacher-level variables.

The estimated coefficients on the teacher characteristics in these models represented associations rather than causal effects. For instance, if any of the teacher characteristics included in the models were correlated with some other teacher characteristics that also affected student math achievement but were omitted from the models (for instance, teacher enthusiasm), the true causal effect of that omitted variable on the outcome would be attributed to the included characteristics, leading to biased estimates of the causal effects of each included characteristic. Nonetheless, these estimates could provide useful descriptive information on the association between each included teacher characteristic and student math achievement.

As discussed in Chapter IX, evidence from both Tables G.2 and G.3 indicated that students' math achievement was negatively associated with the amount of coursework that their teachers took during the school year and positively associated with two of three measures of teacher experience. We found no other statistically significant relationships between teacher characteristics and student achievement.

3. Accounting for the Difference in Effectiveness Between TFA and Comparison Teachers

We used the estimated relationships between teacher characteristics and student math achievement from Table G.3 to ascertain the difference in effectiveness between TFA and comparison teachers that could be predicted based on each characteristic. These predicted differences in effectiveness represented the portion of the groups' actual difference in effectiveness that could be accounted for by the observed characteristics.[29]

Table G.4 shows results from all steps in this analysis. The first column of entries repeats the coefficients from Table G.3. The second and third columns show the estimate and *p*-value for the difference in each characteristic between TFA and comparison teachers. The final column provides the predicted difference in effectiveness between TFA and comparison teachers based on each characteristic, calculated as the product of the first and second columns.

As we discussed in Chapter IX, the observed characteristics collectively did not account for any positive portion of the difference in effectiveness between TFA and comparison teachers. After summing the predicted TFA-comparison differences in effectiveness based on all observed characteristics, students of TFA teachers were predicted to score 0.028 standard deviations below students of comparison teachers. In fact, however, the experimental estimates from Chapter V found that students of TFA teachers actually scored 0.075 standard deviations higher than students of comparison teachers.

B. Sensitivity Analyses of the Relationships Between Teacher Characteristics and Student Achievement

Starting from the main nonexperimental regression model (Table G.3) that included all teacher characteristics, we explored several modifications to this model to determine the robustness of the main findings. We used alternative measures of particular characteristics, modeled the relationships between teacher characteristics and student achievement with different functional forms, and included additional characteristics in the regression model.

1. Alternative Measures of College Selectivity

Our main measure of college selectivity was a binary indicator for whether a teacher had received a bachelor's degree from a selective college or university. In alternative models, we replaced this variable with an indicator for bachelor's degrees from highly selective colleges or universities, either alone or in conjunction with another indicator for bachelor's degrees from selective but not highly selective colleges or universities (Table G.5). Consistent with the main findings, no measures of college selectivity were associated with student math achievement.

[29] We did not conduct a parallel analysis for the Teaching Fellows sample because we did not find a statistically significant average difference in effectiveness between Teaching Fellows and comparison teachers.

2. Alternative Measures of Exposure to Math Coursework

In the main model, we measured exposure to math coursework with a binary indicator for whether a teacher was in the top half of the study teachers in terms of the number of college-level math courses completed. We conducted sensitivity analyses that explored whether teacher effectiveness was associated with having college minors, college majors, or advanced degrees in math or math-related subjects, with math-related subjects defined as either including (Table G.6, models 1 and 2) or excluding (Table G.6, models 3 and 4) secondary math education. As in the main analyses, we found no statistically significant relationships between teachers' exposure to math coursework and the achievement of their students in these sensitivity analyses.

3. Alternative Measures of Praxis Achievement

Our main specification modeled a linear relationship between teachers' Praxis II scores, expressed as *z*-scores, and student math achievement. In sensitivity analyses, we replaced the continuous Praxis variables with either (1) indicators for whether teachers' Praxis II scores were above the median for study teachers who took the same assessment (Table G.7, model 1); or (2) indicators for whether teachers' Praxis II scores met the passing thresholds for certification in their states (Table G.7, model 2).[30]

Like the main analyses, the sensitivity analyses did not find any association between the Praxis II scores of middle school teachers and their students' achievement. However, high school teachers whose Praxis II scores (on the Mathematics Content Knowledge assessment) exceeded the median were more effective, by a statistically significant degree, than those whose scores were below the median. Earning a passing score on the Mathematics Content Knowledge assessment was not statistically related to teacher effectiveness.

4. Inclusion of Teacher Demographic Characteristics

The main nonexperimental model did not include teacher demographic characteristics because these variables were not central to assessing whether teachers' academic ability and professional background were predictive of teacher effectiveness. In a sensitivity analysis, we augmented the main nonexperimental model with additional variables that measured teachers' demographic characteristics, including their gender, age, and race, as well as a variable indicating whether the teacher and student were of the same race (black or non-black) (Table G.8). This alternative specification yielded similar conclusions to the main findings—namely, that students' achievement was positively associated with some measures of their teachers' experience and negatively associated with the amount of concurrent coursework in which their teachers were enrolled. In addition, the relationship between teachers' scores on the Praxis II Mathematics Content Knowledge assessment and student math achievement reached statistical significance, with higher Praxis scores predicting higher student achievement. We also found that female teachers were more effective than their male counterparts, but no other teacher demographic characteristics were predictive of student math achievement.

[30] If the teacher's state did not require the specified Praxis test, the passing threshold was specified as the average threshold across all states that required the test.

Table G.1. Summary Statistics of Teacher Characteristics Examined in the Nonexperimental Analysis

Teacher Characteristic	Variable Structure	Mean	Standard Deviation
Graduated from Selective College or University[a]	Dichotomous	0.542	0.498
Number of College-Level Math Courses Taken Is Above Median[b]	Dichotomous	0.477	0.499
Used College-Level Math in Nonteaching Job	Dichotomous	0.315	0.464
Score on Praxis II Test in Math Content Knowledge (*z*-score)	Continuous	0.176	0.707
Score on Praxis II Test in Middle School Math (*z*-score)	Continuous	0.287	0.917
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	Dichotomous	0.467	0.499
Number of Days of Student Teaching in Math During Training Is Above Median[d]	Dichotomous	0.343	0.475
Hours of Coursework During the School Year (divided by 10)	Continuous	5.728	11.964
Has More Than One Year of Teaching Experience	Dichotomous	0.876	0.330
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	Continuous	1.770	1.292
Number of Additional Years of Teaching Experience Beyond Five Total Years	Continuous	3.824	7.053
Number of Students		**8,669**	
Number of Classroom Matches		**228**	
Number of Teachers		**287**	
Number of Schools		**82**	

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

Table G.2. Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Only a Single Teacher Characteristic

Teacher Characteristic	Regression Coefficient	*p*-Value
Model 1		
Entered teaching through TFA	0.075**	0.000
Entered teaching through Teaching Fellows program	0.001	0.956
Model 2		
Entered teaching through TFA	0.067**	0.004
Entered teaching through Teaching Fellows program	-0.004	0.880
Graduated from selective college or university[a]	0.017	0.543
Model 3		
Entered teaching through TFA	0.070**	0.000
Entered teaching through Teaching Fellows program	-0.003	0.924
Number of college-level math courses taken is above median[b]	-0.022	0.381
Model 4		
Entered teaching through TFA	0.062**	0.004
Entered teaching through Teaching Fellows program	0.014	0.616
Used college-level math in nonteaching job	-0.045	0.134
Model 5		
Entered teaching through TFA	0.068**	0.004
Entered teaching through Teaching Fellows program	-0.007	0.798
Score on Praxis II test in Math Content Knowledge (*z*-score)	0.017	0.293
Score on Praxis II test in Middle School Math (*z*-score)	0.002	0.907
Model 6		
Entered teaching through TFA	0.075**	0.000
Entered teaching through Teaching Fellows program	0.001	0.967
Number of hours of math pedagogy instruction during training is above median[c]	-0.011	0.668
Model 7		
Entered teaching through TFA	0.075**	0.000
Entered teaching through Teaching Fellows program	0.001	0.970
Number of days of student teaching in math during training is above median[d]	-0.001	0.977
Model 8		
Entered teaching through TFA	0.080**	0.000
Entered teaching through Teaching Fellows program	0.001	0.983
Hours of coursework during the school year (divided by 10)	-0.002*	0.028
Model 9		
Entered teaching through TFA	0.126**	0.000
Entered teaching through Teaching Fellows program	0.029	0.359
Has more than one year of teaching experience	0.069*	0.038
Number of additional years of teaching experience beyond two total years (until teacher has five total years of experience)	0.000	0.991
Number of additional years of teaching experience beyond five total years	0.004*	0.040
Number of Students	**8,669**	
Number of Classroom Matches	**228**	
Number of Teachers	**287**	
Number of Schools	**82**	

Table G.2 *(continued)*

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: Each model represents a separate regression in which student end-of-year math scores are the dependent variable and the displayed variables are the independent variables. All models also include classroom match fixed effects and all covariates in Appendix Table B.1. Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

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

TFA = Teach For America.

Table G.3. Estimated Associations Between Teacher Characteristics and Student Math Achievement in a Model that Includes Multiple Teacher Characteristics: Main Specification

Teacher Characteristic	Regression Coefficient	*p*-Value
Entered Teaching Through TFA	0.102*	0.014
Entered Teaching Through Teaching Fellows Program	0.019	0.609
Graduated from Selective College or University[a]	0.003	0.913
Number of College-Level Math Courses Taken Is Above Median[b]	-0.027	0.281
Used College-Level Math in Nonteaching Job	-0.038	0.190
Score on Praxis II Test in Math Content Knowledge (*z*-score)	0.035	0.051
Score on Praxis II Test in Middle School Math (*z*-score)	-0.001	0.955
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	-0.014	0.592
Number of Days of Student Teaching in Math During Training Is Above Median[d]	-0.003	0.895
Hours of Coursework During the School Year (divided by 10)	-0.002*	0.041
Has More Than One Year of Teaching Experience	0.080*	0.016
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	-0.004	0.755
Number of Additional Years of Teaching Experience Beyond Five Total Years	0.005*	0.017
Number of Students	**8,669**	
Number of Classroom Matches	**228**	
Number of Teachers	**287**	
Number of Schools	**82**	

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: The model in the table represents a single regression in which student end-of-year math scores are the dependent variable and the displayed variables are the independent variables. The model also includes classroom match fixed effects and all covariates in Appendix Table B.1. Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

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

TFA = Teach For America.

Table G.4. Extent to Which Teacher Characteristics Explain the Difference in Effectiveness Between TFA Teachers and Comparison Teachers

Teacher Characteristic	Regression Coefficient from Model that Includes Multiple Teacher Characteristics (Table G.3)	Average Difference in Characteristic between TFA Teachers and Comparison Teachers		Predicted Difference in Effectiveness Between TFA Teachers and Comparison Teachers (*z*-score units)
		Estimate	*p*-Value	
Graduated from Selective College or University[a]	0.003	0.450**	0.000	0.001
Number of College-Level Math Courses Taken Is Above Median[b]	-0.027	-0.233**	0.001	0.006
Used College-Level Math in a Nonteaching Job	-0.038	-0.289**	0.000	0.011
Score on Praxis II Test in Math Content Knowledge (*z*-score)	0.035	0.354**	0.001	0.012
Score on Praxis II Test in Middle School Math (*z*-score)	-0.001	0.780**	0.000	-0.001
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	-0.014	0.012	0.884	0.000
Number of Days of Student Teaching in Math During Training Is Above Median[d]	-0.003	-0.046	0.538	0.000
Hours of Coursework During the School Year (divided by 10)	-0.002*	2.641	0.160	-0.005
Has More Than One Year of Teaching Experience	0.080*	-0.327**	0.000	-0.026
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	-0.004	-2.301**	0.000	0.010
Number of Additional Years of Teaching Experience Beyond Five Total Years	0.005*	-7.324**	0.000	-0.036
Total Predicted Difference in Effectiveness Between TFA Teachers and Comparison Teachers Based on All Characteristics				**-0.028**
Total Observed Difference in Effectiveness Between TFA Teachers and Comparison Teachers				**0.075**
Number of Students	**8,669**			
Number of Classroom Matches	**228**			
Number of Teachers	**287**			
Number of Schools	**82**			

Table G.4 *(continued)*

Source: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: Differences in characteristics between TFA and comparison teachers are estimated from a student-level regression of the indicated teacher characteristic on a TFA indicator, classroom match fixed effects, and all covariates in Appendix Table B.1. Therefore, estimated differences might not be identical to those calculated from teacher-level data in Chapter IV. Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

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

TFA = Teach For America.

Table G.5. Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Multiple Teacher Characteristics: Specifications with Alternative Measures of College Selectivity

	Model 1		Model 2	
Teacher Characteristic	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value
Entered Teaching Through TFA	0.111**	0.004	0.104*	0.012
Entered Teaching Through Teaching Fellows Program	0.021	0.546	0.017	0.635
Graduated from Highly Selective College or University[a]	-0.042	0.338	-0.034	0.511
Graduated from Selective but Not Highly Selective College or University[a]			0.013	0.648
Number of College-Level Math Courses Taken Is Above Median[b]	-0.027	0.266	-0.027	0.276
Used College-Level Math in Nonteaching Job	-0.037	0.192	-0.037	0.188
Score on Praxis II Test in Math Content Knowledge (*z*-score)	0.038*	0.025	0.037*	0.035
Score on Praxis II Test in Middle School Math (*z*-score)	0.002	0.909	0.002	0.904
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	-0.017	0.488	-0.017	0.495
Number of Days of Student Teaching in Math During Training Is Above Median[d]	0.003	0.916	0.002	0.933
Hours of Coursework During the School Year (divided by 10)	-0.002*	0.027	-0.002*	0.030
Has More Than One Year of Teaching Experience	0.080*	0.016	0.081*	0.015
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	-0.002	0.863	-0.003	0.844
Number of Additional Years of Teaching Experience Beyond Five Total Years	0.005*	0.016	0.005*	0.019
Number of Students	**8,669**		**8,669**	
Number of Classroom Matches	**228**		**228**	
Number of Teachers	**287**		**287**	
Number of Schools	**82**		**82**	

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: Each model represents a separate regression in which student end-of-year math scores are the dependent variable and the displayed variables are the independent variables. All models also include classroom match fixed effects and all covariates in Appendix Table B.1. Missing values of teacher-level variables are imputed with multiple imputation.

[a]Highly selective colleges are those ranked by *Barron's* as most competitive. Selective but not highly selective colleges are those ranked by *Barron's* as very competitive or highly competitive.

[b]Teacher at the median took 7.5 college-level math courses.

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

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

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

TFA = Teach For America.

Table G.6. Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Multiple Teacher Characteristics: Specifications with Measures of Academic Degrees in Math or Math-Related Subjects

	Model 1: Broader Definition of Math-Related Subject		Model 2: Broader Definition of Math-Related Subject		Model 3: Narrower Definition of Math-Related Subject		Model 4: Narrower Definition of Math-Related Subject	
Teacher Characteristic	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value
Entered Teaching Through TFA	0.096*	0.020	0.089*	0.038	0.110**	0.005	0.105*	0.011
Entered Teaching Through Teaching Fellows Program	0.008	0.812	0.005	0.879	0.023	0.517	0.019	0.614
Graduated from Selective College or University[a]	-0.004	0.899	0.007	0.851	-0.003	0.931	0.005	0.883
Has College Minor in Math or Math-Related Subject[b]	-0.044	0.320	-0.024	0.631	-0.037	0.370	-0.021	0.717
Has College Major in Math or Math-Related Subject[b]	-0.038	0.243	-0.035	0.429	-0.015	0.627	-0.009	0.826
Has Advanced Degree in Math or Math-Related Subject[b]	0.001	0.984	0.002	0.955	-0.068	0.147	-0.068	0.156
Has College Minor in Math or Math-Related Subject from Selective College or University[a, b]			-0.053	0.455			-0.032	0.748
Has College Major in Math or Math-Related Subject from Selective College or University[a, b]			-0.008	0.892			-0.009	0.868
Used College-Level Math in Nonteaching Job	-0.031	0.293	-0.033	0.257	-0.029	0.310	-0.028	0.320
Score on Praxis II Test in Math Content Knowledge (*z*-score)	0.034	0.056	0.034	0.061	0.029	0.143	0.029	0.158
Score on Praxis II Test in Middle School Math (*z*-score)	0.003	0.812	0.004	0.777	-0.001	0.944	0.000	0.979
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	-0.017	0.521	-0.016	0.539	-0.005	0.840	-0.005	0.830
Number of Days of Student Teaching in Math During Training Is Above Median[d]	0.002	0.951	-0.002	0.934	-0.004	0.891	-0.006	0.829
Hours of Coursework During the School Year (divided by 10)	-0.002*	0.041	-0.002*	0.035	-0.002*	0.039	-0.002*	0.039
Has More Than One Year of Teaching Experience	0.076*	0.024	0.077*	0.023	0.079*	0.017	0.083*	0.015
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	-0.001	0.911	-0.003	0.829	-0.004	0.789	-0.005	0.719

Table G.6 *(continued)*

Teacher Characteristic	Model 1: Broader Definition of Math-Related Subject		Model 2: Broader Definition of Math-Related Subject		Model 3: Narrower Definition of Math-Related Subject		Model 4: Narrower Definition of Math-Related Subject	
	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value
Number of Additional Years of Teaching Experience Beyond Five Total Years	0.005*	0.025	0.005*	0.032	0.005*	0.013	0.005*	0.020
Number of Students	**8,669**		**8,669**		**8,669**		**8,669**	
Number of Classroom Matches	**228**		**228**		**228**		**228**	
Number of Teachers	**287**		**287**		**287**		**287**	
Number of Schools	**82**		**82**		**82**		**82**	

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: Each model represents a separate regression in which student end-of-year math scores are the dependent variable and the displayed variables are the independent variables. All models also include classroom match fixed effects and all covariates in Appendix Table B.1. Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]In the broader definition, math-related subjects include mathematics, secondary mathematics education, statistics, engineering, computer science, finance, economics, and physics/astrophysics. The narrower definition includes all of the aforementioned subjects except secondary math education.

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

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

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

TFA = Teach For America.

Table G.7. Estimated Associations Between Teacher Characteristics and Student Math Achievement in Models that Include Multiple Teacher Characteristics: Specifications with Alternative Measures of Praxis Achievement

	Model 1		Model 2	
Teacher Characteristic	Regression Coefficient	*p*-Value	Regression Coefficient	*p*-Value
Entered Teaching Through TFA	0.102**	0.010	0.109**	0.006
Entered Teaching Through Teaching Fellows Program	0.009	0.796	0.026	0.448
Graduated from Selective College or University[a]	0.000	0.990	0.004	0.885
Number of College-Level Math Courses Taken Is Above Median[b]	-0.035	0.160	-0.023	0.357
Used College-Level Math in Nonteaching job	-0.038	0.179	-0.044	0.136
Score on Praxis II Test in Math Content Knowledge Is Above Median	0.114**	0.001		
Score on Praxis II Test in Middle School Math Is Above Median	0.008	0.784		
Score on Praxis II Test in Math Content Knowledge Is at Least as High as State's Passing Threshold[c]			0.050	0.258
Score on Praxis II Test in Middle School Math Is at Least as High as State's Passing Threshold[c]			-0.030	0.466
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[d]	-0.007	0.774	-0.008	0.752
Number of Days of Student Teaching in Math During Training Is Above Median[e]	-0.010	0.683	-0.006	0.834
Hours of Coursework During the School Year (divided by 10)	-0.002	0.055	-0.002*	0.035
Has More Than One Year of Teaching Experience	0.088**	0.007	0.079*	0.020
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	-0.002	0.873	-0.002	0.859
Number of Additional Years of Teaching Experience Beyond Five Total Years	0.005**	0.006	0.004*	0.040
Number of Students	**8,669**		**8,669**	
Number of Classroom Matches	**228**		**228**	
Number of Teachers	**287**		**287**	
Number of Schools	**82**		**82**	

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: Each model represents a separate regression in which student end-of-year math scores are the dependent variable and the displayed variables are the independent variables. All models also include classroom match fixed effects and all covariates in Appendix Table B.1. Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]Teacher at the median took 7.5 college-level math courses.

[c]If the teacher's state does not require the indicated Praxis test, then the passing threshold is specified as the average threshold across all states that require this test.

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

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

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

TFA = Teach For America.

Table G.8. Estimated Associations Between Teacher Characteristics and Student Math Achievement in a Model that Includes Multiple Teacher Characteristics: Specification with Teacher Demographic Characteristics

Teacher Characteristic	Regression Coefficient	*p*-Value
Entered Teaching Through TFA	0.107*	0.017
Entered Teaching Through Teaching Fellows Program	0.016	0.630
Graduated from Selective College or University[a]	0.001	0.969
Number of College-Level Math Courses Taken Is Above Median[b]	-0.033	0.186
Used College-Level Math in Nonteaching Job	-0.035	0.230
Score on Praxis II Test in Math Content Knowledge (*z*-score)	0.042*	0.019
Score on Praxis II Test in Middle School Math (*z*-score)	0.010	0.515
Number of Hours of Math Pedagogy Instruction During Training Is Above Median[c]	-0.016	0.509
Number of Days of Student Teaching in Math During Training Is Above Median[d]	-0.003	0.903
Hours of Coursework During the School Year (divided by 10)	-0.002*	0.044
Has More Than One Year of Teaching Experience	0.084*	0.015
Number of Additional Years of Teaching Experience Beyond Two Total Years (until teacher has five total years of experience)	-0.008	0.543
Number of Additional Years of Teaching Experience Beyond Five Total Years	0.005*	0.036
Teacher is Female	0.061*	0.014
Teacher's Age (in years)	-0.001	0.705
Teacher is Black	0.032	0.369
Teacher Has Same Black/Nonblack Status as Student	-0.021	0.322
Number of Students	**8,669**	
Number of Classroom Matches	**228**	
Number of Teachers	**287**	
Number of Schools	**82**	

Sources: District administrative records, study-administered Northwest Evaluation Association assessments, and Survey of Secondary Math Teachers.

Note: The model in the table represents a single regression in which student end-of-year math scores are the dependent variable and the displayed variables are the independent variables. The model includes classroom match fixed effects and all covariates in Appendix Table B.1. Missing values of teacher-level variables are imputed with multiple imputation.

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

[b]Teacher at the median took 7.5 college-level math courses.

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

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

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

TFA = Teach For America.

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


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA