KIPP Middle Schools: Impacts on Achievement and Other Outcomes

Final Report

February 27, 2013

Christina Clark Tuttle
Brian Gill
Philip Gleason
Virginia Knechtel
Ira Nichols-Barrer
Alexandra Resch

MATHEMATICA
Policy Research

Mathematica Reference Number:
06441.910

Submitted to:
KIPP Foundation
135 Main Street, Suite 1700
San Francisco, CA 94105
Project Officer: Danielle Stein Eisenberg

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Philip Gleason

KIPP Middle Schools: Impacts on Achievement and Other Outcomes

Final Report

February 27, 2013

Christina Clark Tuttle
Brian Gill
Philip Gleason
Virginia Knechtel
Ira Nichols-Barrer
Alexandra Resch

MATHEMATICA
Policy Research

ACKNOWLEDGMENTS

This study was made possible by the efforts of many individuals and organizations over several years. We sincerely appreciate the willingness of KIPP school principals and other administrative staff to participate in the evaluation, especially their endless patience in providing access to their admissions practices. We would also like to acknowledge the staff of the state and district departments of education who generously made their data available to our team and who provided valuable assistance and guidance.

This report would not have been possible without contributions from many other people at Mathematica. Lisbeth Goble and Emily Dwoyer, the project's survey directors, oversaw all the primary data collection efforts with assistance from Nikkilyn Morrison. They were supported by Laurie Bach, Bina Chandler, Lauren Harris, Felicia Hurwitz, Sumia Ibrahim, Ava Madoff, Kim Mook, Chris Rafferty, Amanda Skaff, staff at Mathematica's Princeton Survey Operations Center, and Information Services staff. Margaret Sullivan was responsible for securing participation from KIPP schools, supported by Emily Dwoyer, Josh Furgeson, and Allison McKie Seifullah. Mary Anne Anderson led the effort to collect administrative data from school districts and states, with assistance from Jessica Jacobson and guidance from Josh Furgeson. Michael Barna provided critical research and programming assistance at several stages of the project, including cleaning school records, implementing matched comparison group analyses, and compiling results. Clare Wolfendale and Alena Davidoff-Gore gave additional programming support during the data-cleaning process. Maureen Higgins provided invaluable support with variable creation, programming for the school and student characteristics tables, and the analysis of factors associated with KIPP impacts; she also helped analyze the results. Ji Kwon-Min helped with data cleaning and variable creation for the measurement of student behaviors and attitudes. Lauren Akers assisted with the analysis and synthesis of the lottery results. Josh Haimson contributed significantly to the study design, and Steve Glazerman provided critical technical review and feedback. Donna Dorsey led the production of the report, which was edited by Laura Bernstein.

Finally, the study and this report benefited greatly from input at various stages from Danielle Eisenberg, Jonathan Cowan, and Steve Mancini at the KIPP Foundation, as well as Carrie Hahnel at The Education Trust—West, Ila Deshmukh Towery at TNTP, and Jason Atwood at Teach For America.

CONTENTS

EXECUTIVE SUMMARY xiii

I INTRODUCTION 1

A. KIPP Network of Schools 1

B. Findings from Prior Research 2

C. Research Questions 3

II STUDY DESIGN 5

A. Overview 5

B. Defining the Sample 6

C. Data Used in the Study 12

D. Analytic Approach 14

III SCHOOL AND STUDENT CHARACTERISTICS 19

A. Who Enters KIPP? 19

B. Are KIPP Students Promoted, and Do They Complete KIPP? 21

C. What Are the Characteristics of KIPP Schools? 22

IV KIPP'S IMPACTS ON TEST SCORES 31

A. How Does KIPP Affect Student Scores on State Assessments? 31

B. Lottery-Based Estimates of KIPP's Impacts on Student Achievement 41

C. How Does KIPP Affect Students' Higher-Order Thinking Skills? 45

V IMPACTS ON STUDENT BEHAVIOR AND ATTITUDES 47

A. How Does KIPP Affect Student Engagement and Effort in School? 47

B. How Does KIPP Affect Educational Expectations and Aspirations? 48

C. How Does KIPP Affect Student Well-Being and Behavior? 50

D. How Does KIPP Affect Satisfaction and Perceptions of School? 51

VI ANALYSIS OF FACTORS ASSOCIATED WITH KIPP IMPACTS 55

A. Do KIPP Middle School Impacts Vary? 56

B. Factors of Interest 59

C. What School-Level Factors Are Related to Impacts? 61

VII CONCLUSIONS AND AREAS FOR FUTURE RESEARCH 67

REFERENCES 71

APPENDIX A: SAMPLE SELECTION AND BASELINE CHARACTERISTICS 73

APPENDIX B: CONSTRUCTING SURVEY OUTCOMES 93

APPENDIX C: SCHOOLS ATTENDED BY LOTTERY WINNERS AND LOTTERY NON-WINNERS 117

APPENDIX D: ANALYTIC METHODS FOR THE MATCHED COMPARISON GROUP ANALYSIS 121

APPENDIX E: ANALYTIC METHODS FOR LOTTERY-BASED ANALYSIS 139

APPENDIX F: VALIDATION OF MATCHING METHODS USING LOTTERY-BASED IMPACT ESTIMATES 151

TABLES

II.1 All KIPP Middle Schools Through 2009-10 7

II.2 Schedule of Lottery Sample Data Collection Activities 12

III.1 Characteristics of Students Who Attend KIPP vs. Feeder vs. All District Schools 20

III.2 KIPP and District Grade Repetition Rates, by Grade 22

III.3 Operational Characteristics of Study and All KIPP Schools 25

III.4 Academic Programming and School Climate at Study and All KIPP Schools 27

III.5 Staff at Study and All KIPP Schools 29

IV.1 Mean Test Score Effects in Mathematics and Reading, Benchmark Model 33

IV.2 Mean Test Score Effects in Science and Social Studies, Benchmark Model 34

IV.3 Impact Estimates on State Assessments for Subset of Oversubscribed KIPP Schools 42

IV.4 Impact Estimates on the TerraNova Test Administered in the Fall of the Third Follow-Up Year 46

V.1 Impacts on Student Motivation and Engagement 49

V.2 Impacts on Education Goals and Aspirations 50

V.3 Impacts on Student Behavior 52

V.4 Impacts on School Experiences and Satisfaction 53

VI.1 Factors Potentially Influencing Charter School Impacts 60

VI.2 Bivariate Relationships Between School Characteristics and KIPP School Impacts 62

VI.3 Multivariate Relationships Between School Characteristics and KIPP School Impacts 63

A.1 Schools and Cohorts Included in QED Sample 76

A.2 List of Potential Covariates for Inclusion in Propensity Score Model 77

A.3 Balance Between KIPP Students and Matched Comparison Students in Year One 79

A.4 Balance Between KIPP Students and Matched Comparison Students in Year Two 80

A.5 Balance Between KIPP Students and Matched Comparison Students in Year Three....80

A.6 Balance Between KIPP Students and Matched Comparison Students in Year Four....81

A.7 Balance Between KIPP Students and Matched Comparison Students with Science Scores....81

A.8 Balance Between KIPP Students and Matched Comparison Students with Social Studies Scores....82

A.9 Lottery Detail for Schools Included in Lottery-Based Analysis....84

A.10 Baseline Equivalence for Lottery Sample (Baseline Sample)....86

A.11 Baseline Equivalence for Lottery Sample (Analytic Sample: Year 1 State Test Scores)....87

A.12 Baseline Equivalence for Lottery Sample (Analytic Sample: Year 2 State Test Scores)....88

A.13 Baseline Equivalence for Lottery Sample (Analytic Sample: TerraNova Test)....89

A.14 Baseline Equivalence for Lottery Sample (Analytic Sample: Parent Survey)....90

A.15 Baseline Equivalence for Lottery Sample (Analytic Sample: Student Survey)....91

B.1 Construction of Principal Survey Outcomes....97

B.2 Construction of Student and Parent Survey Outcomes....106

C.1 Characteristics of Schools Attended by KIPP Lottery Winners and Non-Winners....120

D.1 List of Covariates Included in OLS Model....124

D.2 Test for Selection Effects Prior to KIPP Enrollment....130

D.3 Comparison of KIPP Effects on Subgroups to Effects on Other KIPP Students, Mathematics....133

D.4 Comparison of KIPP Effects on Subgroups to Effects on Other KIPP Students, Reading....134

D.5 KIPP Effects on Hispanics and Students with Low Prior Test Scores....135

D.6 Comparison of Benchmark Impact Model and Alternative Models, Mathematics 136

D.7 Comparison of Benchmark Impact Model and Alternative Models, Reading 137

E.1 Unadjusted Means, Standard Deviations, Sample Sizes, and Reliability of Outcome Measures.....142

E.2 Sensitivity of Impact Estimates to Alternative Models.....145

F.1 Comparison Between Lottery-Based Impact Estimates and Non-Lottery Impact Estimates 156

FIGURES

ES.1 Student Baseline Characteristics: KIPP vs. Feeder Schools xiv

ES.2 Location of KIPP Schools in the Study xvi

ES.3 KIPP Estimated Impacts on Student Achievement xvii

ES.4 KIPP Estimated Impacts on Student Achievement in Percentiles, by Subject.....xviii

IV.1 KIPP Estimated Impacts on Student Achievement in Percentiles, by Subject.....36

IV.2 Percentage of KIPP Schools with Significant Effects in Math, by Year 37

IV.3 Percentage of KIPP Schools with Significant Effects in Reading, by Year.....38

IV.4 Distribution of Reading and Math Impact Estimates After Two Years.....39

IV.5 Distribution of Reading and Math Impact Estimates After Three Years 40

IV.6 Percentage of KIPP Schools with Significant Effects in Science and Social Studies.....40

IV.7 Comparison of Lottery Impact Estimates and Matching Impact Estimates in Math.....43

IV.8 Comparison of Lottery Impact Estimates and Matching Impact Estimates in Reading 44

VI.1 Distribution of School-Level Impact Estimates in Reading.....57

VI.2 Distribution of School-Level Impact Estimates in Math.....57

VI.3 Distribution of School-Level Impact Estimates in Reading (by Region) 58

VI.4 Distribution of School-Level Impact Estimates in Math (by Region).....59

EXECUTIVE SUMMARY

The Knowledge Is Power Program (KIPP) is a rapidly expanding network of public charter schools whose mission is to improve the education of low-income children. As of the 2012–2013 school year, 125 KIPP schools are in operation in 20 different states and the District of Columbia (DC). Ultimately, KIPP's goal is to prepare students to enroll and succeed in college. Prior research has suggested that KIPP schools have positive impacts on student achievement, but most of the studies have included only a few KIPP schools or have had methodological limitations.

This is the second report of a national evaluation of KIPP middle schools being conducted by Mathematica Policy Research. The evaluation uses experimental and quasi-experimental methods to produce rigorous and comprehensive evidence on the effects of KIPP middle schools across the country. The study's first report, released in 2010, described strong positive achievement impacts in math and reading for the 22 KIPP middle schools for which data were available at the time.

For this phase of the study, we nearly doubled the size of the sample, to 43 KIPP middle schools, including all KIPP middle schools that were open at the start of the study in 2010 for which we were able to acquire relevant data from local districts or states. This report estimates achievement impacts for these 43 KIPP middle schools, and includes science and social studies in addition to math and reading. This report also examines additional student outcomes beyond state test scores, including student performance on a nationally norm-referenced test and survey-based measures of student attitudes and behavior.

The average impact of KIPP on student achievement is positive, statistically significant, and educationally substantial. KIPP impact estimates are consistently positive across the four academic subjects examined, in each of the first four years after enrollment in a KIPP school, and for all measurable student subgroups. A large majority of the individual KIPP schools in the study show positive impacts on student achievement as measured by scores on state-mandated assessments. KIPP produces similar positive impacts on the norm-referenced test, which includes items assessing higher-order thinking. Estimated impacts on measures of student attitudes and behavior are less frequently positive, but we found evidence that KIPP leads students to spend significantly more time on homework, and that KIPP increases levels of student and parent satisfaction with school. On the negative side, the findings suggest that enrollment in a KIPP school leads to an increase in the likelihood that students report engaging in undesirable behavior such as lying to or arguing with parents. We describe these findings in more detail in the pages below.

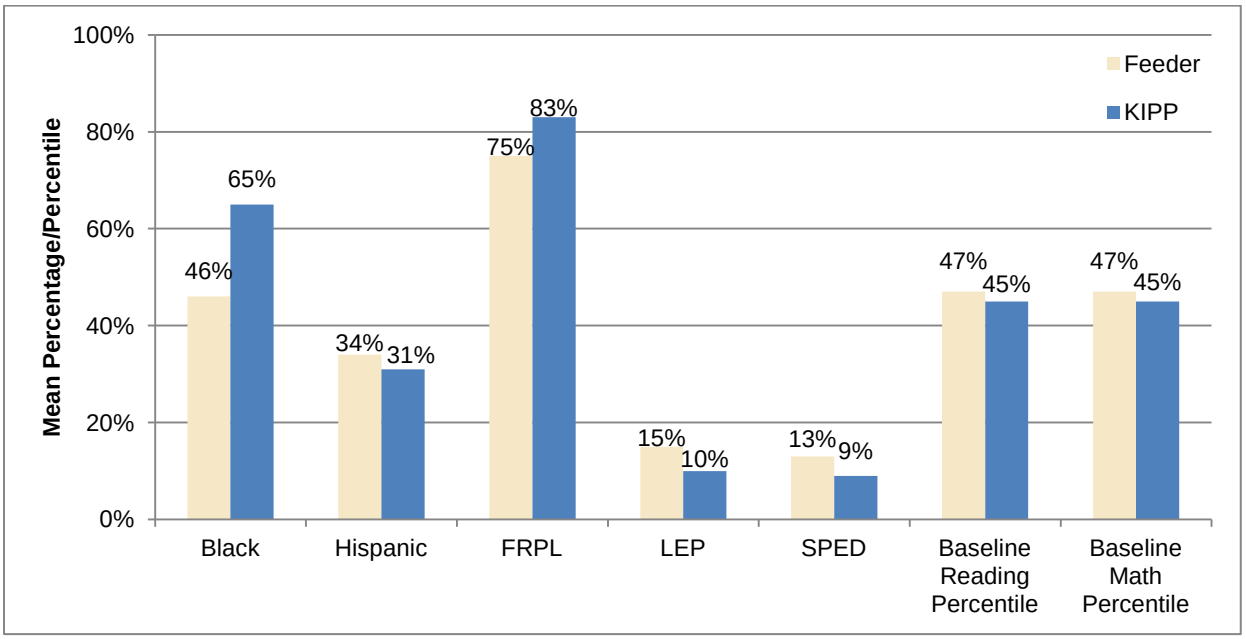
Who Attends KIPP, and How Do KIPP Students Proceed Through Middle School?

To examine the characteristics of the students who enter KIPP schools (typically in 5th grade) we compared the 4th grade characteristics of future KIPP students and their elementary school classmates; that is, non-KIPP students in the same districts attending the same elementary feeder schools from which KIPP middle schools draw students. We also examined patterns of grade repetition and early exit from KIPP schools, as compared with other middle schools nearby.

Data on student characteristics provided little evidence that KIPP "creams" or selectively enrolls higher-performing students, though students entering KIPP are less likely to have received special education services. For most identifiable characteristics, the students entering KIPP schools look much like other students in their neighborhoods: low-achieving, low-income, and non-white.

Nearly all KIPP students (96 percent) are either black or Hispanic, and more than four-fifths (83 percent) are from households with incomes low enough to be eligible for free or reduced-price lunch (FRPL)—percentages that are higher than those of the KIPP students' feeder schools (Figure ES.1). The typical KIPP student scored at the 45th percentile within the district in reading and math prior to entering KIPP, an achievement level significantly lower than the average in their own elementary schools. In contrast, KIPP students are somewhat less likely than students at their feeder schools to have received special education services (9 versus 13 percent) or be classified as having limited English proficiency (LEP, 10 versus 15 percent) when they were in elementary school.

Figure ES.1. Student Baseline Characteristics: KIPP vs. Feeder Schools



Note: All differences are statistically significant at the 0.05 level, two-tailed test.

On average, students do not leave KIPP schools at unusually high rates prior to middle school completion. The proportion of entering students who transfer before 8th grade is identical at KIPP and non-KIPP district schools (37 percent). However, KIPP schools are consistently more likely than local district schools to have students repeat a grade.

How Does KIPP Affect Student Achievement?

We examined KIPP impacts on students' performance on state assessments across four subject areas—reading, math, science, and social studies. We also measured impacts on a nationally norm-referenced test that incorporates items assessing higher-order thinking skills. Our primary method of analysis was a matched comparison group design that produced impact estimates for 41 KIPP schools. This design used propensity score matching techniques to identify a set of non-KIPP district students who, based on their characteristics and achievement trajectories in elementary school, closely resemble KIPP students. Using statistical controls for small remaining differences between the groups, we then compared the achievement trajectories of the KIPP students and comparison students on state assessments in each of the first four years after KIPP entry (typically grades 5–8). Our estimates of KIPP's impact reflect the effect of having ever enrolled at KIPP—students who leave before completing 8th grade remain part of the KIPP "treatment group" after

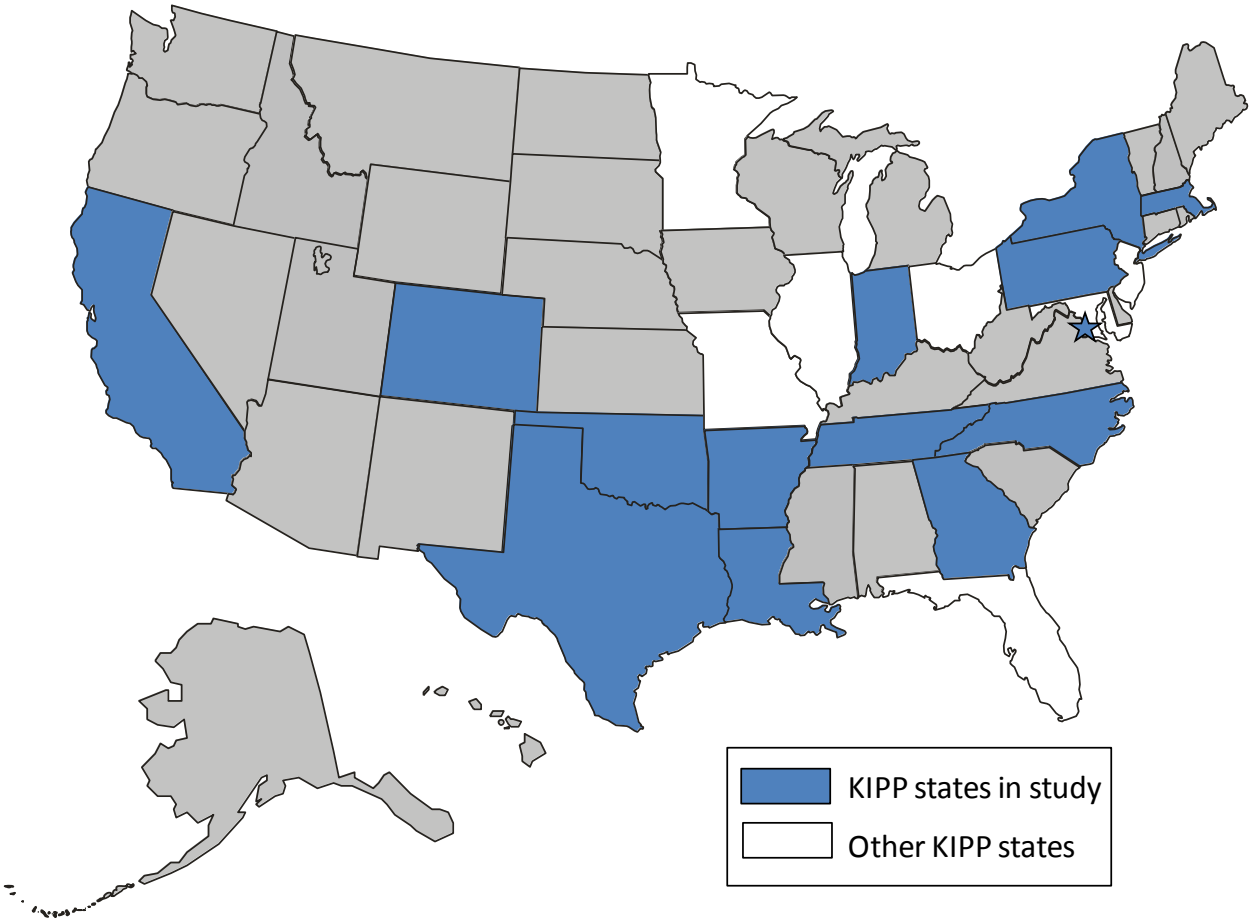
leaving, thereby ensuring that we do not artificially inflate KIPP's estimated impact by focusing only on students who persist at KIPP for four years.

We also used a lottery-based design as an alternative, experimental method of estimating impacts for a subset of 13 KIPP schools (including 2 schools not included in the matched comparison sample of 41 schools). We compared a treatment group of students offered admission to a KIPP school on the basis of receiving a winning draw in the school's randomized admissions lottery with a control group of students who applied to the school and participated in the lottery but who did not receive a winning draw. The lottery design uses random assignment to form treatment and control groups, making it essentially a randomized experiment—the gold standard for estimating impacts. The design guarantees that the treatment group of students is similar to the comparison group on all key characteristics, including baseline test scores and demographics, as well as items that we cannot measure such as motivation and parental support.

Despite the rigor of the lottery design, we cannot use it as our primary approach because most schools do not have enough lottery participants to support the design. Fortunately, the matched comparison design produces estimates of KIPP's achievement impacts that are not significantly different from the experimental estimates. When we apply the matching approach to the same students and schools included in a lottery-based analysis, we find that the impact estimates produced by the two methods are very similar, with no statistically significant differences. The success of the matching approach in replicating the lottery-based results provides more confidence in the results produced by the matching approach with the full set of 41 KIPP schools.

The 41 schools in the matched study comprise a majority of all KIPP middle schools in a majority of the states served by KIPP (Figure ES.2) as of the 2009–2010 school year. At that point, there were 53 KIPP middle schools in operation across 20 states and DC. Another 10 middle schools operated by KIPP had closed or lost their KIPP affiliation by 2010. Of these 63 middle schools operating in 2009–2010 or earlier, we included all KIPP schools (38 operating, 3 closed) located in states and/or school districts that could provide at least three consecutive years of complete, longitudinally linked student-level data for both traditional public and charter schools. For each school in the matching sample, we were able to calculate impacts for between 2 and 10 cohorts per school, with outcomes observed between the 2001-2002 school year and the 2010–2011 school year. These 41 schools are similar to the full population of KIPP middle schools on a variety of operational dimensions and student characteristics, suggesting the possibility of generalizing the matched comparison estimates to the full population of KIPP schools.

Figure ES.2. Location of KIPP Schools in the Study



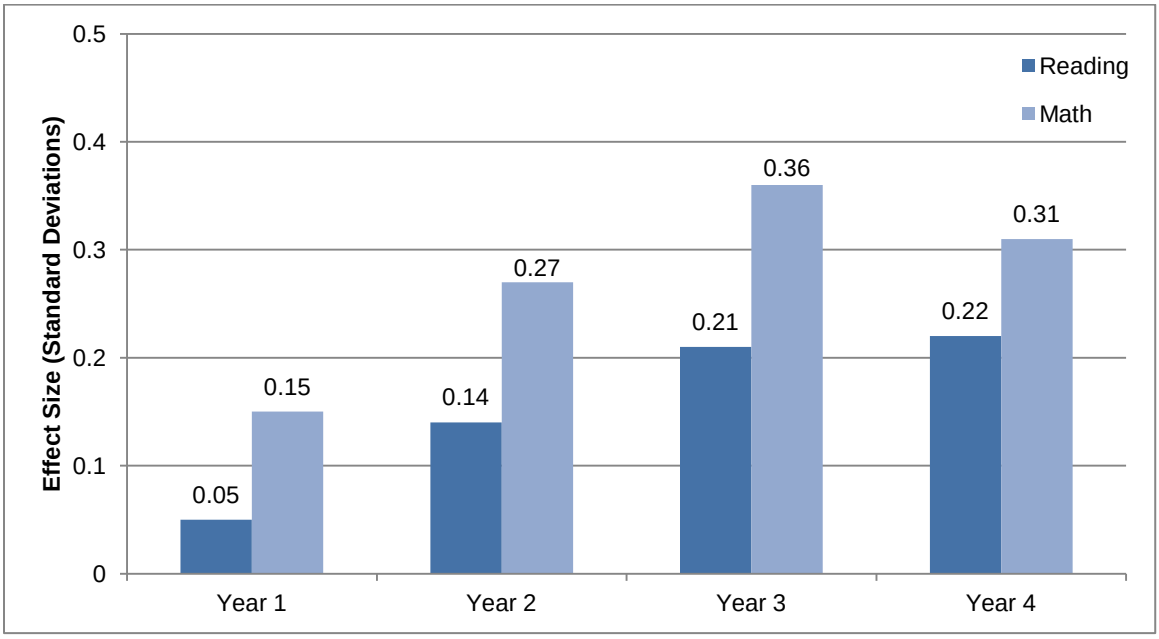
Our impact estimates suggest four key results related to how KIPP affects student achievement:

Key finding 1: KIPP middle schools have positive and statistically significant impacts on student achievement across all years and all subject areas examined.

The estimated effects of KIPP on student achievement are consistently positive. In each of the four years after KIPP entry, KIPP has a statistically significant positive impact on students' performance on state assessments in both reading and math, based on the matched comparison group design (Figure ES.3). The impacts for student subgroups are similar to the average overall impact among all KIPP students. This is true on average across KIPP and for most of the 41 KIPP schools in the matched comparison analysis.

KIPP schools also positively affect student achievement in science and social studies. We measured these impacts in whatever grade states administered tests in these subjects (typically 8th grade). The estimated impacts of KIPP are positive and statistically significant in both science and social studies, and the magnitudes of these effects are similar to the estimated impacts in math and reading after three to four years.

Figure ES.3. KIPP Estimated Impacts on Student Achievement



Note: All impacts are statistically significant at the 0.05 level, two-tailed test.

Key finding 2: The magnitude of KIPP's achievement impacts is substantial.

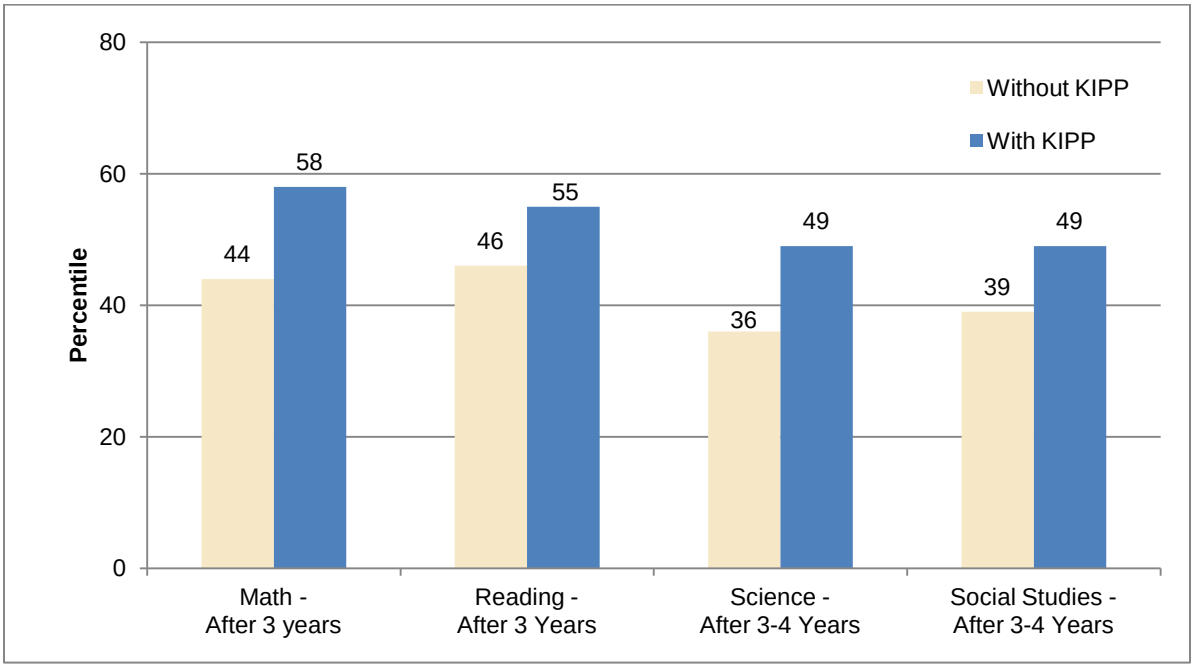
Across the KIPP schools in the analysis sample, average impacts in all subjects are large enough to be educationally meaningful. Three years after enrollment, the estimated impact in math is 0.36 standard deviations, equivalent to moving a student from the 44th to 58th percentile of the district's distribution (Figure ES.4). This impact estimate suggests that the average KIPP middle school produces approximately 11 months of additional learning growth in math for its students after three years (Bloom et al. 2008). The size of the math impact produced by KIPP schools after three years is equivalent to about 40 percent of the local black-white test score gap.

The average impact of KIPP after three years in reading (0.21 standard deviations) is somewhat smaller than that for math—equivalent to moving a student from the 46th to 55th percentile. Compared to national norms, this estimated reading impact represents approximately eight months of additional learning growth (Bloom et al. 2008). The three-year reading impact is equivalent to about 26 percent of the local black-white test score gap in reading.

KIPP's impact in science after three to four years (0.33 standard deviations) is equivalent to moving a student from the 36th to 49th percentile, representing approximately 14 months of additional learning growth. KIPP's impact in social studies after three to four years (0.25 standard deviations) is equivalent to moving a student from the 39th to 49th percentile, representing about 11 months of extra learning growth in social studies. KIPP's science and social studies impacts are equivalent to about a third of the local black-white test score gap in these subjects.

Evidence on the magnitudes of estimated impacts of other charter school management organizations (CMOs) suggests that KIPP is among the highest-performing charter networks in the country (Furgeson et al. 2012).

Figure ES.4. KIPP Estimated Impacts on Student Achievement in Percentiles, by Subject



Note: For math and reading, the figure shows the impact of KIPP on the scores of tests taken three years after enrollment in a KIPP school; for science and social studies, the figure show the impact on scores of tests taken three years after enrollment for some student cohorts and four years after enrollment for other student cohorts. The blue bar represents the mean percentile rank of KIPP students in the relevant analysis sample, relative to local jurisdictions. The beige bar represents this observed mean rank minus the average KIPP impact estimate in each subject. In all four subjects, the difference in percentiles represents an impact that is statistically significant at the 0.05 level, two-tailed test.

Key finding 3: The matched comparison design produces estimates of KIPP's achievement impacts similar to estimates of the same impacts based on an experimental, lottery-based design.

A possible criticism of the matched comparison group design is that we can never be completely certain that we are accounting for unmeasured factors that lead some students to enroll in KIPP schools. It is possible, for example, that students who apply to KIPP differ from other students in their elementary schools with regard to educational motivation. If this characteristic is not captured in prior test scores or other variables in our data set, this omitted student characteristic could lead to bias in our estimates of the KIPP achievement effect. Fortunately, for a subset of schools, we are able to implement a lottery-based design that does not suffer from this limitation.

In the subset of schools in the lottery-based analysis, the estimated impacts of KIPP on student achievement in math and reading are similar to the estimates from the matched comparison design. As mentioned above, this is true when we used the exact same sample of KIPP students and carefully replicated the lottery-based estimates using the matched comparison approach. This is also true when we compared the lottery-based estimates to the original matched comparison group design estimates for those schools, which are based on a larger number of cohorts and students than the lottery-based estimates. In other words, the analysis revealed no evidence of bias in KIPP's estimated achievement impacts based on a matched comparison group design when compared with those based on an experimental, lottery-based design for the subset of KIPP schools for which both

designs are possible. This finding supports our use of the matched comparison group design for generating achievement impact estimates for the broader set of KIPP schools.

Key finding 4: In the lottery sample, average KIPP impacts on a nationally normed, low-stakes test that includes items assessing higher-order thinking skills were similar to impacts on high-stakes state tests.

In the KIPP schools included in the lottery-based analysis, we administered a low-stakes, nationally norm-referenced assessment (the TerraNova, which included constructed response items in the reading component) to test the robustness of the results found on state assessments. The magnitude of the estimated impacts of these KIPP schools on the study-administered test was consistent with the positive point estimates found on the state assessments. However, because a smaller sample of students took the TerraNova, statistical power is limited and the reading estimate does not achieve statistical significance. The math estimate is statistically significant.

This finding is important for two reasons. First, because the test results did not have consequences for students, teachers, or schools, the TerraNova results suggest that the positive impacts of KIPP are not a result of "teaching to the test" on state assessments. Second, TerraNova results taken alongside the positive impacts in science and social studies suggest that KIPP is doing more academically than simply improving students' basic skills in reading and math.

How Does KIPP Affect Student Behavior and Attitudes?

In addition to affecting students' academic achievement, KIPP may influence student behaviors and attitudes related to long-term academic success. For KIPP schools in the lottery sample, we used the experimental design to estimate impacts on various measures of student behavior and attitudes. Notable findings from this analysis include:

- Students enrolled at KIPP spend an additional 35 to 53 minutes on homework per night than they would have in a non-KIPP school, completing an average of more than two hours of homework per night (according to student and parent self-reports) as a result.
- KIPP has no statistically significant effect on a variety of measures of student attitudes that may be related to long-run academic success. The estimated KIPP impacts on indices of student-reported self-control, academic self-concept, school engagement, effort/persistence in school, and educational aspirations are not statistically significant.
- KIPP has no statistically significant effect on several measures of student behavior, including self-reported illegal activities, an index of good behavior, and parent reports of behavior problems. However, KIPP has a negative estimated effect on a student-reported measure of undesirable behavior, with KIPP students more likely to report behaviors such as losing their temper, arguing or lying to their parents, or giving their teachers a hard time.
- Winning an admissions lottery to KIPP has a positive effect on students' and parents' satisfaction with school. In addition, the parents of KIPP students are less likely to report that their child's school is too easy.

Are the Characteristics of KIPP Schools Associated with Impacts?

While most KIPP schools have significant positive impacts on student achievement, some KIPP schools have more positive impacts than others. This raises the question of whether there are particular characteristics of some schools that make them more successful. Ultimately, we would like to understand the conditions under which KIPP schools are most likely to promote the academic achievement of their students so that successful practices and conditions can be replicated.

The factors that drive the success of KIPP schools could not easily be determined in our analysis. Few of the school characteristics we examined are strongly correlated with the estimated impacts of the KIPP schools in the study sample. For example, class size, teacher experience and professional development opportunities are not associated with impacts. The lack of significant correlations between these school characteristics and impacts may be explained, in part, by the limited sample size of 38 schools for which impact estimates and school characteristics were available, affecting our ability to detect small to moderately-sized relationships.

Nonetheless, we identified two factors related to the strength of KIPP schools' impacts on student achievement. One is the approach of the KIPP school toward student behavior and school culture. KIPP's impact on student achievement is larger in schools where principals report a more comprehensive school-wide behavior system. This finding is consistent with the findings of several other recent studies of charter schools (Angrist et al. 2011; Dobbie and Fryer, 2011; Furgeson et al. 2012). Under comprehensive school-wide behavior systems, schools have clearly defined and consistently enforced rewards for good behavior and consequences for negative behavior.

Second, the length of the school day and how time is used are also significantly associated with impacts. All KIPP schools have longer-than-normal school days (with an average KIPP school day of more than nine hours), but some have longer days than others. Overall, average impacts on student achievement are smaller in KIPP schools with a particularly extended school day. This counterintuitive relationship appears to be driven by the fact that, in these schools, the additional time tends to be spent in non-core academic activities. In contrast, average impacts on student achievement are larger in KIPP schools in which relatively more time is spent on core academic activities.

It is difficult to isolate the elements that create a successful KIPP school. This may be because KIPP's approach aims to integrate multiple strategies in concert—which is why KIPP believes that no single factor is responsible for creating a high functioning KIPP school. Nonetheless, the variance in impacts achieved by KIPP schools suggests that there may be operational differences among the schools. More research is needed to identify exactly what makes each school more or less successful than its peers. In future work evaluating the KIPP network's effort to "scale up," we will address this and other key questions in more detail. We will calculate impacts for additional KIPP schools and generate separate impacts by school year (not only by number of years a student is enrolled), giving us a larger sample for analyzing factors that can be correlated to KIPP impacts and the opportunity to observe how the impacts of individual KIPP schools change over time. In addition, this work will enable us to estimate the effectiveness of newer KIPP schools, including elementary and high schools. Finally, as the network matures, researchers will be able to calculate longer-term impacts on students, assessing KIPP's progress towards its goals of seeing more students to and through college.

I. INTRODUCTION

KIPP is a rapidly expanding network of public charter schools whose mission is to improve the educational opportunities available to low-income families. Ultimately, the goal of KIPP is to prepare students to succeed in college and life. The achievement levels of KIPP students, as measured by state and national norm-referenced test scores, are often substantially higher than those of other low-income, minority students. Indeed, the promise seen in KIPP schools, as well as other charter school networks that use a similar approach, has helped place charter schools at the center of the national dialogue around education reform and public schooling.

This is the second report of a national evaluation of KIPP middle schools, which aims to provide the most rigorous and comprehensive evidence on the achievement effects of KIPP middle schools across the country. We focus on middle schools because they serve the grades originally targeted by the KIPP model, and represent the majority of schools within the network. The study's first report, released in 2010 by Mathematica Policy Research, described strong positive achievement impacts in the 22 KIPP middle schools for which data were available at the time. The current report estimates achievement impacts for twice as many KIPP middle schools; includes impact estimates for science and social studies as well as math and reading; examines additional student outcomes beyond state test scores; provides survey data describing the experiences of KIPP students; and attempts to examine whether particular operational features are characteristic of the highest-performing KIPP schools.

A. KIPP Network of Schools

KIPP schools seek to engage students and parents in the educational process, expand the time and effort students devote to their studies, reinforce students' social competencies and positive behaviors, and dramatically improve their academic achievement. KIPP describes its approach as resting on "Five Pillars," publically available on its website:

- **High expectations** for all students to reach high academic achievement, regardless of students' backgrounds
- **Choice and commitment** on the part of students, parents, and faculty to a public, college preparatory education as well as the time and effort required to reach success
- **More time** spent learning, both in academics and extra-curricular activities, each day, week, and year
- **Power to lead** for school principals, who are given the freedom to make budgeting, personnel, and other decisions, in exchange for heightened accountability for student results
- **Focus on results** by regularly assessing student learning and sharing results to drive continuous improvement and accountability

KIPP was founded in 1994 by Mike Feinberg and Dave Levin, two teachers who had recently completed placements with Teach For America. KIPP began with the launch of a fifth-grade public school program in Houston, Texas. In 1995, Feinberg remained in Houston to lead KIPP Academy Middle School and Levin moved to New York City to establish the second KIPP Academy. In 2000,

the KIPP Foundation was established by the two educators in partnership with Doris and Don Fisher, founders of Gap Inc., to support the expansion of the KIPP network. The KIPP Foundation selects and trains school leaders, seeks to identify ways schools can be improved, and provides services to the KIPP network, including legal, real estate, technology, finance, professional development, operations, and communications support.

With the support of the KIPP Foundation, the network has grown dramatically. As of the 2012–13 school year, 125 KIPP public charter schools operate in 20 states and the District of Columbia. From 1994 to 2004, KIPP focused on middle schools only, with all KIPP schools enrolling students in grades 5–8. KIPP now offers instruction at all grade levels: the network includes 37 elementary schools, 70 middle schools, and 18 high schools. Collectively, these schools currently enroll more than 41,000 students.[1] This makes the KIPP network about the same size as the St. Paul, Minnesota or Corpus Christi, Texas school districts. KIPP schools are locally governed, so that the KIPP network is actually comprised of 31 autonomous regional organizations and single-site schools. KIPP Regions encompass a specific metropolitan or geographic area and provide targeted support to their schools on leadership practices, human resources, business operations, technology, and development.

B. Findings from Prior Research

Previous studies of KIPP schools have used methods of varying rigor, typically involving only small numbers of schools, but they have consistently found positive effects of KIPP attendance on student achievement.[2] One criticism of early studies was that researchers did not fully control for the possibility that KIPP might be selectively enrolling higher achieving students, so that positive results are not explained by what KIPP schools do, but which students they enroll. Recently, more rigorous studies in specific geographical areas have provided further evidence that KIPP is having positive effects on the students it serves. In 2008, SRI International published an analysis of student impacts at three San Francisco Bay Area KIPP schools. The authors compared KIPP students at these schools to a matched set of students at traditional public schools. For the three Bay Area schools, the authors found large and statistically significant impacts for both math and reading (with effect sizes ranging from 0.16 to 0.68 standard deviations in reading, and 0.19 to 0.88 standard deviations in math) for students in the first year after enrolling in KIPP middle schools (Woodworth et al. 2008). In the first study of a KIPP school to use randomized admissions lotteries to form treatment and control groups, Angrist et al. (2012) studied the KIPP Academy in Lynn, Massachusetts. Comparing the test scores of admission lottery winners to students who were not initially offered admission, the authors found that a year of KIPP Lynn attendance had a large, statistically significant impact in mathematics (0.35 standard deviations) and a smaller, significant impact in language arts (0.12 standard deviations). A study of charter-school management organizations that included some KIPP schools (Furgeson et al. 2012), using a matched comparison group design, found that KIPP's middle schools in Washington, DC, were producing substantially larger achievement gains than were comparison schools nearby.

In Mathematica's 2010 study of 22 KIPP middle schools that served as a pilot of the quasi-experimental methods used in this report, we used a matched comparison group design to estimate impacts on achievement (Tuttle et al. 2010). One year after students enrolled, we found that on

[1] See http://www.kipp.org/schools (accessed January 14, 2013).

[2] For a thorough review of evidence reported in earlier published studies of KIPP outcomes, see Henig (2008).

average these KIPP schools generated statistically significant impacts of 0.26 standard deviations in math and 0.09 standard deviations in reading. We found substantially larger cumulative impacts three years after enrollment (0.42 standard deviations in math and 0.24 standard deviations in reading), even when students who exited KIPP schools through attrition were kept in the treatment group.

C. Research Questions

With the nationwide expansion of KIPP, the KIPP Foundation, its funders, and other stakeholders are eager to rigorously assess the effectiveness of the program and determine which school practices may be positively related to student outcomes. More specifically, this study addresses six questions about KIPP middle schools across the country:

1. What are the effects of KIPP middle schools on the achievement of their students in math, reading, science, and social studies, up to four years after entering KIPP?
2. How, if at all, do KIPP's effects differ for particular subgroups of students?
3. Do KIPP's effects depend on the particular tests used to measure achievement? In other words, does KIPP affect performance not only on state tests of basic skills but also on low-stakes assessments that include items capturing higher-order thinking skills?
4. Is there evidence that KIPP produces positive effects on higher-order thinking skills beyond basic academic skills?
5. Is there evidence that KIPP affects nonacademic outcomes related to students' engagement, attitudes, and behavior?
6. Are there particular characteristics that distinguish high-performing KIPP schools from lower-performing KIPP schools?

The next chapter discusses the methods we use to measure the characteristics of KIPP schools and estimate impacts. Chapter III describes the characteristics of KIPP schools and students. Chapter IV presents estimates of the impacts of KIPP middle schools on student achievement in reading, math, science, and social studies. Chapter V describes KIPP's effects on other student outcomes, including students' attitudes and behavior. Chapter VI describes our analysis and findings on relationships between the characteristics of KIPP schools and their achievement effects. Finally, in Chapter VII, we summarize our findings and discuss areas for future research.

II. STUDY DESIGN

A. Overview

The goal of this evaluation is to produce the best possible estimate of the average impact of KIPP middle schools on their students' outcomes, relative to outcomes of these same students had they not been able to enroll in a KIPP school. Achieving this goal requires a design that has the greatest possible causal rigor while also representing the largest possible sample of KIPP middle schools.

Because students choose to attend KIPP schools by making an active decision to do so, it is difficult to know whether their observed outcomes are attributable to the effects of the schools or to the underlying characteristics of the students and their families. The best way to rule out the latter explanation, which could lead to selection bias in estimates of charter school impacts, is to use an experimental design in which a student's opportunity to attend the school is determined by a randomized admissions lottery.

However, not all KIPP schools are suitable for lottery-based analysis. Not all KIPP schools are substantially oversubscribed, and even those that are may have particular admission rules and priorities that preclude a lottery-based design. In addition, oversubscribed KIPP schools that are eligible for a lottery-based design may differ in meaningful ways from the average KIPP school. If the more oversubscribed schools also tend to be the most effective (for example), focusing only on this subset of schools would lead us to overstate the effectiveness of the network. We compare the schools in our lottery-based sample to all KIPP middle schools on measurable characteristics in the next chapter to address this issue explicitly, but in general this suggests a need for a more representative sample.

Our study estimates the impact of KIPP middle schools using a hybrid approach that takes advantage of the best features of two distinct methods: (1) an experimental design using randomization based on admissions lotteries at eligible KIPP schools; and (2) a matched comparison group design that compares the outcomes of KIPP students to a group of students with similar observable baseline characteristics at a large number of KIPP schools. The study's lottery-based impact estimates have high causal (internal) validity—that is, because each lottery was random, we know that the differences between the outcomes of lottery winners and non-winners represent the direct impact of receiving an admission offer to a KIPP middle school. However, because lottery data are available only for a small number of KIPP schools, the lottery-based impact findings may not fully represent the impact of KIPP middle schools operating across the country. The study's matched comparison group design can be implemented at a much larger number of KIPP schools and thus has greater external validity.

Fortunately, we can use the lottery-based method to validate the matching approach. We do this by conducting parallel lottery-based and matched comparison group analyses for a subset of KIPP schools and students to examine whether matching methods can successfully replicate lottery-based impact estimates. To the extent that they can, we have greater confidence in the causal validity of the matched comparison group impact estimates as applied to the larger group of KIPP schools and students.

In this study, the matched comparison group analysis, which is based on data from a large number of KIPP middle schools, provides our primary estimates of the impact of KIPP on state test scores. Other student outcomes, however, are available only for students who participated in the admissions lotteries and provided consent to participate in the study. We collected test and survey data not available from districts or states to assess KIPP's impacts on a nationally-normed test of higher-order thinking skills and on the perceptions, attitudes, and behaviors of students and parents. For those outcomes, we were limited to the lottery-based schools, and we used the lottery-based approach to estimate impacts.

B. Defining the Sample

To provide context for the findings, this section describes the samples of KIPP schools and students used in the analyses. Overall, we included 41 schools in the matching analysis, and 13 schools in the lottery-based analysis. Of these 13, 11 are also represented in the matching sample, so there is considerable (but not perfect) overlap in the school samples across the methods.

1. Sample for the Matched Comparison Group Analysis

A key goal for the matching analysis was to include as many schools as possible so that our impacts would comprise a large portion of the KIPP network (Table II.1). Two criteria were used to select KIPP middle schools for the analysis. First, all included schools had to be established in the 2009–10 school year or earlier to ensure that a minimum of two cohorts of students per school would be observed by spring of 2011.[3] Second, the schools had to be located in jurisdictions (states or school districts) that provided at least three consecutive years of complete, longitudinally linked student-level data for both traditional public and charter schools. These data were needed to track individual KIPP and non-KIPP students in the years prior to middle school enrollment, as well as during the middle school. Throughout this report, we use the term "baseline year" to refer to the school year that began one year prior to when a cohort of students first entered middle school at KIPP; the term "pre-baseline year" refers to the point two years before middle school entry.

[3] Throughout the matching analysis, a "cohort" is defined as the group of students who first enrolled in a KIPP middle school at the beginning of a given school year.

Table II.1. All KIPP Middle Schools Through 2009-10

State	City	KIPP School (Year Opened)	KIPP Lottery Analysis School	KIPP Matched Analysis School	KIPP School Not Studied
AR	Helena	Delta College Prep (2002)		X	
CA	Fresno	Fresno (2004)—*closed*			X
	Los Angeles	Academy of Opportunity (2003)	X	X	
		Los Angeles College Prep (2003)	X	X	
	Oakland	Bridge Academy (2002)		X	
	Sacramento	Sacramento Prep (2003)—*closed*			X
	San Diego	Adelante Prep (2003)		X	
	San Jose	Heartwood Academy (2004)			X
	San Francisco	Bayview Academy (2003)		X	
		San Francisco Bay Academy (2003)		X	
	San Lorenzo	Summit Academy (2003)	X		
CO	Denver	Sunshine Peak Academy (2002)		X	
DC	Washington	DC KEY Academy (2001)	X	X	
		DC AIM Academy (2005)		X	
		DC WILL Academy (2006)	X	X	
GA	Atlanta	PATH (2002)—*closed*			X
		Achieve Academy (2003)—*closed*		X	
		WAYS Academy (2003)	X	X	
		Strive Academy (2009)		X	
	East Point	South Fulton Academy (2003)	X		
IL	Chicago	Ascend (2003)			X
		Chicago Youth Village (2003)—*closed*			X
IN	Gary	LEAD College Prep (2006)—*closed*		X	
	Indianapolis	Indianapolis College Prep (2004)		X	
LA	New Orleans	Believe College Prep (2006)		X	
		Central City Academy (2007)		X	
MA	Lynn	Academy Lynn (2004)	X	X	
MD	Annapolis	Harbor Academy (2005)—*closed*			X
	Baltimore	Ujima Village (2002)			X
MN	Minneapolis	Stand (2008)			X
MO	Kansas City	Endeavor (2007)			X
	St. Louis	Inspire (2009)			X
NC	Asheville	Asheville Youth Academy (2002)—*closed*		X	
	Charlotte	Academy Charlotte (2007)			X
	Gaston	Gaston College Prep (2001)		X	

Table II.1 *(continued)*

State	City	KIPP School (Year Opened)		KIPP Lottery Analysis School	KIPP Matched Analysis School	KIPP School Not Studied
NJ	Camden	Freedom (2004)—*closed*				X
	Newark	TEAM (2002)				X
		RISE (2006)				X
NY	Albany	Tech Valley (2005)				X
	Buffalo	Sankofa (2003)—*closed*				X
	New York City	Academy New York (1995)		X	X	
		STAR College Preparatory (2003)			X	
		AMP Academy (2005)			X	
		Infinity Charter (2005)			X	
OH	Columbus	Journey (2008)				X
OK	Oklahoma City	Reach College Preparatory (2002)			X	
	Tulsa	Tulsa College Prep (2005)			X	
PA	Philadelphia	Philadelphia Charter (2003)			X	
		West Philadelphia Prep (2009)			X	
TN	Memphis	Memphis Collegiate Middle (2002)			X	
	Nashville	Academy Nashville (2005)				X
TX	Austin	Austin College Preparatory (2002)		X	X	
		Arts & Letters (2009)				X
	Dallas	TRUTH Academy (2003)		X	X	
	Houston	Academy Middle (1995)		X	X	
		3D Academy (2001)			X	
		Liberation College Prep (2006)			X	
		Spirit College Prep (2006)			X	
		Polaris Academy for Boys (2007)			X	
		Sharpstown College Prep (2007)			X	
		Intrepid Prep (2008)			X	
		Voyage Academy for Girls (2009)			X	
	San Antonio	Aspire Academy (2003)		X	X	
		Total Number of KIPP Middle Schools	63	13	41	20
		Number of KIPP Middle Schools Open by 2009-10	53	13	38	13
		Number of KIPP Middle Schools Closed by 2010	10	0	3	7

As of the 2009–10 school year, there were 53 KIPP middle schools in operation across 20 states and the District of Columbia. Another 10 middle schools operated by KIPP had closed or lost their KIPP affiliation by 2010.[4] Of these 63 schools operating in 2009–10 or earlier, we were able to include 41 KIPP schools (38 operating, 3 closed) in our matched comparison group analysis sample, representing two-thirds of KIPP middle schools potentially eligible for the design.[5] The sample includes all closed KIPP schools located in the jurisdictions that provided data; however, our sample comprises a smaller proportion of KIPP schools that closed (30 percent) than operating KIPP schools (72 percent). This could bias upward our estimates of the average KIPP impact providing impacts are smaller in the closed schools.[6]

For each of the 41 schools in the matching sample, we are able to calculate impacts for between 2 and 10 cohorts per school, with outcomes observed between the 2001–02 and 2010–11 school years. See Appendix A for a detailed description of student cohorts included in the analysis versus all those that ever entered each KIPP school in the study sample.

To select the student-level sample for the matched comparison group analysis, we identified each student who attended a sample KIPP school in either 5th or 6th grade; these students comprised the treatment group. Using a technique called propensity score matching, we then selected from among all students in the same district a comparison group whose demographic characteristics and baseline achievement matched those of the treatment group students. Specifically, we used a propensity score matching procedure known as "nearest neighbor" matching to identify a comparison student within the appropriate grade and year for each KIPP treatment student.[7]

The resulting treatment and matched comparison sample included more than 30,000 students. There are no statistically significant differences between baseline achievement scores of the treatment group of KIPP students and those of the matched comparison group, nor any significant differences on any demographic characteristics in our data. The average baseline z-score in reading

---

[4] The KIPP Foundation has a licensing agreement with all KIPP schools, in which schools using the KIPP name agree to abide by the central principles of KIPP, including academic excellence and financial sustainability. If a school is unable to live up to these principles, whether because of internal circumstances or because of actions taken by the school's authorizer, then the school and KIPP can decide to part ways. The KIPP Foundation has the authority to remove the KIPP name from a school, but the decision to close a school is made solely by the local board of directors.

[5] Six new KIPP middle schools opened during the 2010–11 school year. We were able to include three of these, plus two older schools, in regression-based estimates that did not use propensity score matching. Matching could not be completed for these schools because outcome data was only available for a single cohort of students whose sample size was insufficient for propensity score estimation. With this non-matching approach, we calculated impacts for 43 of 59 schools operating as of 2010–11 (73 percent), and 46 of 69 KIPP schools ever operating (67 percent). Results of this analysis are presented in Appendix D.

[6] To test the potential magnitude of this bias, we calculated average impacts separately for closed and open schools in our sample and assigned those values to schools missing data. If we use these imputed school-specific impacts to estimate KIPP-wide effects for all 63 KIPP schools (including 10 that were closed) that were theoretically eligible for inclusion in the study, the KIPP-wide average impacts would be only slightly lower—by no more than 0.03 SD in magnitude—during the first two years of operation.

[7] We did not allow the same comparison student to be matched to more than one KIPP student in a given cohort (this is known as matching without replacement). Within each jurisdiction, prior to matching, the pool of eligible comparison students was restricted to those whose propensity scores fall within the range of those for KIPP students (that is, we required there to be "common support" in the two groups before identifying comparison matches).

for the treatment group was -0.100 versus -0.095 for the comparison group; in math, the average baseline z-score was -0.135 for the treatment group versus -0.125 for the comparison group. The full details of our propensity score matching and estimation procedures, and detailed tables of baseline equivalence, are presented in Appendix A.

2. Sample for Lottery Analysis

Selecting Schools. As with the matching analysis, we sought to maximize the number of schools included in the lottery analysis. However, the criteria for inclusion in the lottery analysis were more restrictive—to be eligible for the lottery sample, a KIPP school had to (1) be oversubscribed—have more applicants than open seats—for 5th or 6th grade by their scheduled lottery date;[8] (2) conduct a lottery to randomly select students for admissions offers *and* produce a randomly-ordered waitlist of students not selected for admission via the lottery; (3) make subsequent offers of admission to fill additional open seats following the randomly-ordered waitlist, and (4) not exhaust the randomly-ordered waitlist of original lottery participants through the start of the school year. Many KIPP schools ultimately have sizeable waitlists consisting of students who apply to the school after the lottery date. However, these schools could not be included in the lottery analysis simply based on the level of demand, because the design does not support the inclusion of students who apply after the lottery (who may be different from students who apply in time to participate in the lottery).

We recruited two sets of schools into the sample: those with students applying in spring 2008 for admission for the 2008–09 school year (cohort 1) and those with students applying in spring 2009 for admission for the 2009–10 school year (cohort 2). In the fall of 2007 and again in the fall of 2008, we spoke with staff at each operating KIPP middle school to gauge their expected amount of oversubscription for the forthcoming year. In all schools that had conducted a lottery in the past, or expected to do so in the spring of that school year, we worked with staff to track applicants in the period leading up to the admissions lotteries, typically held between February and May. A member of the study team personally attended each lottery to obtain an independently-verified copy of the lottery results and waitlist and to document any stratification used.[9] Finally, throughout the summer and fall, the study team was in regular contact with school staff to obtain updates on offers made to students on the waitlist and to collect rosters.

At the end of the recruitment period for cohort 1, four schools maintained their eligibility for the lottery-based study design. Nine additional schools were eligible for cohort 2, giving us a final sample of 13 KIPP middle schools in the lottery sample (see Table II.1).

---

[8] At the time of recruitment, all KIPP middle schools began with the 5th grade, the "normal" entry point for a KIPP school. However, the combination of attrition and a relatively high rate of grade repetition in 5th grade (discussed later in this chapter) resulted in a number of open slots to be filled at the 6th grade level in many KIPP schools. Oversubscription also tended to be more extensive at the 6th grade level, perhaps due to the fact that this is a typical transition grade to middle school in traditional public schools.

[9] Depending on state legislation, charter schools may be allowed to employ stratification or preferences in their admissions lottery to enable them to better target their intended population of students. In practice, in the KIPP lotteries we observed, some schools stratified on the basis of geography (within school zones, zip codes, and/or districts) and eligibility for free- or reduced-price lunch.

Recruiting Students: Exemptions and Consent. All students who applied to 5th or 6th grade at a participating KIPP school prior to the lottery during one of the study years were initially eligible to be included in the sample, excluding students who applied after the lottery. We also excluded students who were automatically admitted to the school without participating in the lottery, typically those who had a sibling already enrolled in the school. The prevalence of these lottery exemptions varied widely across schools, but on average, 36 percent of open slots were filled with exempt students.

We obtained active parental consent for eligible applicants to participate in the study prior to the schools' admissions lotteries, which ensured that there was no systematic relationship between the likelihood of consent for a given student and whether he or she was offered admission to the school (and thus was in the treatment group) or not offered admission (and thus was in the control group). The average consent rate among lottery participants was 75 percent and was statistically equivalent for treatment and control students (74 percent and 76 percent, respectively).[10]

Defining Treatment Status. We used the lottery and waitlist outcomes to assign students to either the treatment or control group, as appropriate. Treatment status reflects whether or not students' participation in the lottery led them to have an opportunity to attend a KIPP school for the full follow-up period. In particular, the treatment group includes any sample member who received an offer of admission to KIPP for the current school year on the basis of her/his lottery outcome. In most cases, students in the sample who were offered admission based on lottery position (or the lottery position of a sibling who also applied to the school), and prior to a specified cutoff date early in the school year, are included in the treatment group. The control group includes all sample members who were never offered admission on the basis of lottery position or offered admission after this date. Any students offered admission "out-of-order" and who would not otherwise have been made an offer of admission by this date were considered admissions errors and kept in the control group regardless of whether they ultimately attended the KIPP school.[11]

If the lotteries were truly random, we would expect to see few statistically significant differences between lottery winners and non-winners in the mean values of baseline (pre-lottery) student characteristics. Appendix A presents an analysis of baseline equivalence on the characteristics that are included as covariates in the impact analysis for the analytic sample of students with valid outcomes. We found no statistically significant differences between lottery winners and non-winners on baseline achievement, and few statistically significant differences on demographic characteristics.

In sum, this report includes impacts on outcomes for more than two-thirds of the KIPP network operating as of the 2009–10 school year—40 of the 53 KIPP middle schools in 13 of 19 states with KIPP schools, and DC. In Chapter III, we describe the characteristics of KIPP schools in the study and examine similarities and differences between these KIPP schools and those we were not able to study (including those that opened in the 2010–11 and 2011–12 school years) and find that, on most characteristics, the groups of schools are very similar.

---

[10] Typically, schools with lower consent rates were those with less-formal application procedures that were less conducive to incorporating consent material into the application process.

[11] See Appendix A for more detail about defining treatment status and details of individual school lotteries.

C. Data Used in the Study

1. Data for the Matched Comparison Group Analysis

For the matched comparison group analysis, we used de-identified, longitudinally-linked student-level data from jurisdictions (states or districts) hosting at least one KIPP school and able to provide student-level records at the time of data collection. The variables from jurisdictions' administrative data systems included: test scores in reading, mathematics, social studies, and science (where middle school scores represent the primary outcome and elementary school scores a key matching variable and baseline covariate); demographic characteristics, used for matching and as baseline covariates; and schools attended and dates of enrollment, identifying students' exposure to KIPP. Within each jurisdiction, we requested data for all school years beginning with the year prior to the KIPP middle school's first year (to capture baseline data) through the 2010–11 school year. We obtained data from districts for 22 of the 41 schools in the analysis; for the other 19 schools, we obtained records from the state in which the school was located but limited our data to the district (or districts) from which the KIPP school drew students.

2. Data for the Lottery Analysis

Because the study's smaller sample of lottery students provided active consent for participation, we had the opportunity to collect data on other outcomes in addition to state test scores. For the lottery sample, we drew on four sources of data we collected specifically for the study: (1) baseline survey of applicants' parents, (2) nationally-normed, study-administered test of higher-order thinking skills (the TerraNova), (3) follow-up parent survey, and (4) student survey. In addition, we used administrative records collected from states and districts. The data collection structure and schedule are summarized in Table II.2. We dropped sites for specific analyses when we were unable to obtain outcome data for most sample members (described in more detail below).

Table II.2. Schedule of Lottery Sample Data Collection Activities

Activity	Cohort 1	Cohort 2
Baseline Parent Survey	Spring/summer 2008	Spring/summer 2009
Administrative Records		
Baseline	Fall 2008 (covers 2007-08 SY)	Fall 2009 (covers 2008-09 SY)
First follow-up (Year 1)	Fall 2009 (covers 2008-09 SY)	Fall 2010 (covers 2009-10 SY)
Second follow-up (Year 2)	Fall 2010 (covers 2009-10 SY)	Fall 2011 (covers 2010-11 SY)
TerraNova test (Year 3)	Fall 2010	Fall 2011
Follow-Up Parent Survey (Year 2)	Spring 2010	Spring 2011
Follow-Up Student Survey (Year 2)	Spring 2010	Spring 2011

Baseline Survey. Parents whose children applied for admission to KIPP schools participating in the study were asked to complete a baseline survey, via hard copy or telephone. The survey collected demographic and socioeconomic information from parents at the time of application, as well as their reasons for applying to KIPP and information on other schools to which they were

applying. The overall response rate was 82 percent—83 percent among lottery winners and 82 percent among non-winners.

Administrative Records. In coordination with data collection for the matching analysis, records were collected from the states, districts, or schools attended by lottery participants to measure student achievement based on state test scores. These records were obtained for the baseline and pre-baseline years as well as the following two years. We were unable to collect test score information from administrative records for a minimum of half of the sample for three schools. Among members of the resulting analytic sample for state test score outcomes in 10 schools, we have valid administrative records data on test scores for 74 percent of sample members in the first follow-up year (year 1) and 61 percent in the second follow-up year (year 2). In year 1, we obtained valid scores for 78 percent of lottery winners and 72 percent of non-winners in both reading and math. In year 2, we obtained valid scores for 70 percent of lottery winners and 56 percent of non-winners in both reading and math.

TerraNova. We administered a one-time standardized test for all lottery sample students in the fall semester of the third follow-up year. For students promoted on time, the test was administered in the fall of 7th grade (to lottery applicants for 5th grade) and the fall of 8th grade (to applicants for 6th grade). Students were administered the TerraNova 3, Reading Multiple Assessment and Math Survey Exams, Level 17, Form G. There were four purposes in administering the assessment: (1) measure important skills not fully captured by state assessments; (2) measure performance on a nationally norm-referenced but low-stakes test that would not be influenced by "teaching to the test;" (3) provide a consistent achievement outcome across sample schools, which are located in six states; and (4) measure achievement for students who do not take a state assessment (such as those attending private schools) or who take a different state assessment than other students in their original cohort (such as those retained in grade). We were unable to administer tests to at least half of both the treatment and control group at three schools. Among members of the resulting analytic sample for TerraNova outcomes in 10 schools, nearly 70 percent completed the test, including 80 percent of the treatment group and 63 percent of the control group.

Follow-Up Surveys. We administered a short telephone survey to sample members and their parents in spring of 2010 and 2011, during the second year of follow-up for each cohort. The student interviews provided information on students' behavior, both in and out of school, and their attitudes about school. The parent interviews provided information on attitudes about their children's school, assessment of their children's behavior, and reports on their involvement in their children's education and schools. The parent survey response rate was 72 percent—78 percent among lottery winners and 67 percent among non-winners. The response rate on the student survey was 64 percent—71 percent among lottery winners and 58 percent among non-winners. A more detailed description of the outcomes derived from these surveys can be found in Appendix B.

In addition, we used two sources of school-level data to provide context for our analyses (discussed in more detail in Chapter III). These sources, discussed below, provided information for schools attended by students in the lottery sample as well as all KIPP middle schools operating as of 2010–11.

Common Core of Data (CCD) and Private School Survey (PSS): Data from the National Center for Education Statistics (NCES) were used to measure school-level characteristics. These variables included school size, racial/ethnic distribution, and the proportion of students eligible for free and reduced-price lunch.

Principal Survey. We conducted a web-based survey of KIPP middle school principals in spring of 2011. Questions focused on various school features including instructional approaches, operational factors, staff characteristics, and the makeup of the student body. We completed the survey with 55 of 59 KIPP principals (93 percent).

D. Analytic Approach

1. Propensity Score Matching

The validity of our matched comparison group design depends on the ability to eliminate or minimize differences in key characteristics between students who enter KIPP and students in the comparison group who remain in non-KIPP public schools.[12] Our approach achieved this in two ways. First, we used student-level data that included a rich set of student characteristics and multiple years of baseline (prior to KIPP entry) test scores. We used this information to identify a matched comparison group of students who are similar to KIPP students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in elementary school. By matching on more than one year of baseline test score data, we accounted for achievement levels at the time when students applied to KIPP schools as well as pre-KIPP trends in student achievement. After we identified the matched comparison group, the second feature of our approach estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between KIPP students and comparison students. Specifically, the impact estimates adjust for any differences between KIPP students and the matched comparison group pertaining to demographic characteristics or students' prior two years of math and reading test scores.

The combination of propensity-score matching and OLS accounted for differences in observed baseline characteristics and achievement scores between KIPP students and comparison students (in other words, the differences associated with initial selection into KIPP schools). But it remains possible that KIPP students and comparison students differ in unobserved ways that may affect later test scores. There are several other threats to the validity of these impact estimates that we addressed: students moving from KIPP middle schools to other district schools (attrition from KIPP schools), students who are retained in grade, and attrition from the sample.

Attrition from KIPP Schools. The fact that some students depart KIPP schools and return to non-KIPP schools in the surrounding district before the end of 8th grade could potentially introduce selection bias if not appropriately handled. At both KIPP and district schools, students who transfer before the end of middle school tend to be those who are not doing as well academically as those who remain (Nichols-Barrer et al. 2012). In this way, an analysis that only includes persistently enrolled KIPP students would tend to lead to a positive bias in the estimated

[12] Specifically, to produce unbiased impact estimates the design must eliminate differences in student characteristics that could explain academic achievement outcomes and thus be confounded with the treatment of KIPP attendance.

impact of KIPP schools (that is, make KIPP impacts look more positive than they actually are). We addressed this problem by permanently assigning to the treatment group any student who can be found in the records as ever enrolling at KIPP in grades 5 or 6, regardless of whether the student remained in a KIPP school or transferred elsewhere before the end of middle school.[13] In other words, a student who enrolled at KIPP in 5th grade for the 2007–08 school year but left KIPP after completing 6th grade in the 2008–09 school year is included in the treatment group for all four years he or she appears in the data (from 2007–08 to 2010–11, inclusive). By including all students observed attending a KIPP school, regardless of whether they stay through eighth grade, we avoided the problem of overstating the effect of KIPP. Instead, this approach was likely to produce a conservative estimate of KIPP's full impact on students during the years they actually attended KIPP schools. We also conducted an alternative analysis that attempts to estimate the full effect of KIPP on currently enrolled students. Those results are reported in Appendix D.

Grade Repetition. KIPP schools retain students in grades 5 and 6 at a substantially higher rate than do conventional public schools in their local districts (see Chapter III for a full discussion of these findings). This produces a missing data problem for the matching-based analysis of state test scores, as students who repeat a grade do not take the same tests as others in their original cohort. Because KIPP students and comparison students are retained at different rates, our impact estimates could also be biased if we simply excluded all of the retained students from the analysis since we would be excluding a larger proportion of KIPP students. To address this, in the matching analysis of math and reading scores we used information on students' past performance to predict (impute) their outcome scores in the years after retention. For more details on this procedure, as well as a detailed discussion of alternate impact estimates we produced using several other approaches to handle the scores of retained students, see Appendix D.

Analytic Sample Attrition. For a variety of reasons, some students may not have valid data in the year when a given outcome was measured. For example, some students may transfer to a jurisdiction outside of our data catchment area, while others may transfer to local private schools or drop out of school altogether. In a small number of cases, students may simply have missing variable values in a given year or subject.[14] We categorize these cases when students disappear from the analytic sample as out-of-district transfers. If KIPP students transfer out-of-district at a different rate than matched comparison students, it could undermine the validity of impact estimates. But in fact, our matched comparison group did not exit the analytical sample at an appreciably different rate than the study's sample of KIPP students: over the four follow-up years we examined, the difference in sample attrition rates for the two groups is approximately two percentage points

[13] In some locations, our analysis may miss some students who exit very soon after arriving at KIPP. Some of the schools included in our study have day-to-day enrollment records, but others are not so finely grained, creating the possibility of losing students who transfer out before designated student count dates, after which they appear in our administrative records data for surrounding schools.

[14] For example, less than one percent of KIPP students with valid math scores had missing scores in reading one year after entering KIPP.

(see Appendix A).[15] Different analytic sample attrition might occur when students are missing one or more baseline or pre-baseline test scores. To address this we imputed missing baseline data, ensuring that all students with at least one recorded baseline test score remain in the sample. For a detailed discussion of our imputation methods, including results from an alternative set of impact estimates based on data that do not include imputed baseline test scores, see Appendix D.

As mentioned above, there were no significant differences between our matched comparison group and the treatment group of KIPP students in our sample. If these characteristics fully capture the relevant differences between these two groups (that is, there are no unmeasured differences between the two groups that are directly related to tests during the follow-up period), the resulting analyses will produce unbiased impact estimates for KIPP schools. Previous studies have suggested that applying a combination of propensity-score matching and OLS, as we did here, can succeed in replicating experimental impact estimates in certain contexts (Cook et al. 2008; Bifulco 2012; Fortson et al. 2012; Furgeson et al. 2012). To test whether this is also the case for our sample, we compared the matching results to those obtained from a lottery-based analysis for the sample of 8 schools for which it was possible to use both methods. Results of this exercise are presented in Chapter IV.

2. Lottery Analysis

For the subset of KIPP middle schools in which randomized lotteries created viable treatment and control groups, we present two sets of impact estimates: (1) intent-to-treat (ITT) estimates that rely on treatment status as defined by the random lotteries to estimate the impact of being offered admission to a KIPP middle school and (2) treatment-on-the-treated (TOT) estimates that represent the impact of attending a KIPP middle school.

Our benchmark experimental model is the ITT model, comparing outcomes for the experimental treatment and control groups. We estimated the difference between these two groups using a regression framework that controls for baseline characteristics of sample members. The inclusion of baseline characteristics improves the statistical precision of impact estimates. The baseline covariates include age, gender, race/ethnicity, free lunch status, individualized education program (IEP) status, baseline and pre-baseline test scores, whether the student's primary home language is English, whether the household has only one adult, family income, and mother's education. The impact regression model also included indicators for school, grade, and cohort to account for factors specific to a particular school, differences in the test across grade, and differences across time. In contrast to the matching estimation strategy, a single regression model was used for the lottery-based impact estimates, pooling data from all schools. The difference in outcomes between lottery winners and non-winners is interpreted as the average impact of being offered admission to an oversubscribed KIPP school that was included in the lottery analysis.

[15] The KIPP and comparison group sample attrition rates would have been less comparable without matching. In the jurisdiction-wide sample (prior to matching) there is a significant difference between the cumulative, out-of-district attrition rate of KIPP students (15 percent) and the rate among students in comparison jurisdictions (19 percent). Matching addressed this potential problem by identifying a comparison group with a sample attrition rate that is very similar to that of KIPP students. An explanation of our attrition rate calculation method and descriptive attrition findings can be found in Chapter III.

When estimating the regression model we included weights to adjust for unequal probabilities of selection into the treatment group. These probabilities arise for several reasons. Each school has a different number of available seats and number of applicants, so students in schools with few seats and many applicants would have a relatively low probability of winning the admission lottery. Some schools stratify their lotteries based on characteristics like residential location, whether the student has a sibling in the lottery, and gender, leading to differences in selection probabilities for students with and without these characteristics.

We also imputed values for any missing baseline covariates to allow us to include more students in the analysis. These students with imputed values are those who participated in the lottery and for whom we have valid outcome data, but are missing some of the baseline characteristics. Our imputation procedure is described in more detail in Appendix E.

Because families and students choose whether or not to attend KIPP after winning an admissions lottery, and not all lottery winners ultimately attend KIPP, we cannot simply compare outcomes of KIPP attendees and non-attendees to get an unbiased estimate of attending a KIPP middle school. To generate TOT estimates of the impact of attending a KIPP middle school, we use the outcome of the lottery for each student as an instrumental variable for KIPP attendance. In other words, to obtain TOT estimates we calculated the difference between the outcomes of treatment and control students, and adjusted them to reflect the difference between the proportion of treatment and control students who enroll at KIPP.[16] The same covariates, weights and imputed data are used in the TOT model as in the ITT model. More detail on this estimating strategy can be found in Appendix E.

[16] Control students may end up enrolling at KIPP if they are offered admission after the October cut-off date for assignment to the treatment group (for example, during the second semester), if they apply and are offered admission for the following school year, or in rare cases when they are offered admission out of order off the waitlist.

III. SCHOOL AND STUDENT CHARACTERISTICS

In this chapter, we describe the characteristics of KIPP schools and their students. In particular, we examine two issues—first, average student characteristics at KIPP schools relative to the district schools these students might otherwise attend, assessing whether KIPP students differ notably from students in nearby schools, and second, features and practices of all KIPP schools. In addition to describing KIPP schools, we examine how the KIPP schools in our study samples (both for the lottery and matched comparison analyses) compare to each other and to the full population of KIPP schools. As described in Chapter II, the schools in the lottery and matched comparison analyses are not random and may differ in key aspects from the full population of KIPP schools. This analysis addresses how representative our study samples are relative to the universe of KIPP schools. This will inform the appropriateness of using the matched comparison and lottery results to generalize to the full population.

A. Who Enters KIPP?

To investigate whether KIPP schools attract a different type of student than other district schools, we used student-level school records data to examine baseline characteristics of students who later attended 46 KIPP schools compared to those at feeder district elementary schools[17] and students in the district as a whole (Table III.1).[18] We focused our discussion in this chapter on the comparisons with students at district feeder schools, since students at those schools constitute the population from which KIPP students are most likely to be drawn. In all cases, we measured relevant characteristics of students in the grade immediately preceding KIPP entry (typically 4th grade), so that any differences in a school's classification practices cannot affect the results. Thus, poverty, special education, and English-language learner status are identified before students enter KIPP schools. All noted differences are statistically significant. Key findings include:

- **On average, KIPP schools serve student populations that have high concentrations of black students relative to the elementary schools that feed them.** KIPP schools have a much higher proportion of black students (65 percent) than feeder schools (46 percent). They have a slightly smaller proportion of Latino or Hispanic students (31 percent) than do feeder schools (34 percent).

[17] Because many KIPP schools are located within large urban school districts, the full-district comparison group may include students from neighborhoods that are markedly different from the areas directly served by KIPP. For this reason, we also analyze a more focused comparison group limited to the students who attended one of the subset of district elementary schools (or "feeder" schools) attended by students who eventually enrolled in a KIPP middle school. Our analysis of student characteristics (for both the full-district comparison group and the feeder school comparison group) only used administrative records from grade 4, before any students enrolled in KIPP schools. In each district, data on the comparison groups were limited to student cohorts that contained KIPP students (that is, each district's comparison data did not include observations from years prior to when the relevant KIPP school began accepting new 4th grade applicants).

[18] The sample of KIPP middle schools included in the descriptive analysis comprises the 41 schools included in the matched comparison analysis, as well as five additional schools in the same jurisdictions with insufficient data to calculate impacts. All cohorts that include a 4th grader who went on to attend a KIPP middle school are included in the feeder school and district-wide comparison samples.

Table III.1. Characteristics of Students Who Attend KIPP vs. Feeder vs. All District Schools

	KIPP Students	Students at KIPP Feeder Schools	Students at All District Schools
Latino or Hispanic	0.31 N = 19,289	0.34** N = 2,468,555	0.31 N = 5,768,865
Black	0.65 N = 19,289	0.46** N = 2,468,555	0.41** N = 5,768,865
Female	0.52 N = 19,289	0.49** N = 2,468,555	0.49** N = 5,768,865
Free- or reduced-price lunch	0.83 N = 15,556	0.75** N = 2,007,857	0.70** N = 3,398,487
Special education	0.09 N = 19,272	0.13** N = 2,466,846	0.13** N = 5,766,953
Limited English proficiency	0.10 N = 13,706	0.15** N = 978,067	0.14** N = 3,558,411
Baseline reading score (mean z-score)	-0.11 N = 16,859	-0.05** N = 1,959,083	0.03** N = 4,937,348
Baseline math score (mean z-score)	-0.14 N = 16,745	-0.05** N = 1,970,565	0.03** N = 5,043,050

Note: Values are proportions unless otherwise indicated. Sample sizes in each cell represent the number of students included in the calculation.

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

- **The proportion of female students is slightly higher in KIPP schools than in the elementary schools that feed them.** KIPP schools are 52 percent female compared to 49 percent at feeder schools.
- **KIPP schools have a higher proportion of low-income students but lower proportions of special education students and students with limited English proficiency, compared to feeder schools**. We find that prior to KIPP entry, larger proportions of KIPP students are eligible for FRPL (83 percent) than students at the feeder elementary schools (75 percent). In contrast, prior to KIPP entry, a smaller proportion of students at KIPP schools receive special education services (9 percent) or are classified as having limited English proficiency (10 percent) relative to students at KIPP feeder schools (13 percent and 15 percent, respectively).
- **KIPP students have lower baseline math and reading achievement than students at elementary schools that feed KIPP schools.** On average, students entering KIPP schools have lower scores than their peers at the feeder schools, by 0.09 standard deviations in math and 0.06 standard deviations in reading.

Together, these results provide little evidence to support the claim that KIPP "creams" or selectively enrolls higher-performing students. As mentioned above, students attracted to KIPP differ somewhat from their peers at feeder district elementary schools. On some dimensions KIPP students appear to be more disadvantaged than their peers; KIPP schools serve students that are disproportionately eligible for FRPL and who have lower baseline scores in math and reading than do the district and feeder elementary schools, for example. On other dimensions, these students

might have an advantage—KIPP students are less likely to be receiving special education or be classified as having limited English proficiency prior to enrolling in KIPP schools.

B. Are KIPP Students Promoted, and Do They Complete KIPP?

We also examined student enrollment patterns at KIPP schools regarding attrition and grade repetition. Attrition from KIPP schools is an important potential pathway for student selection. If a KIPP school does not retain a large portion of each entering student cohort, it is possible that lower performing students may be those who exit at a higher rate, which could lead to positive peer effects for the remaining students and would bias the estimated effects if they were based only on the sample of students who remain.[19] Our analysis of attrition includes all KIPP students in our data (including those who entered KIPP after grade 5), and adjusts for the fact that more recent student cohorts are present in the data for a limited number of years after grade 5.[20] We find that the rate of overall attrition from KIPP schools is approximately equivalent, on average, to the attrition rate from district schools—37 percent over three years for both groups.[21]

At a subset of the KIPP schools in our sample, we have also examined student mobility patterns in greater depth. In a prior working paper (Nichols-Barrer et al. 2012), we analyzed KIPP attrition rates among five different subgroups of students. We found that rates of attrition from KIPP schools were significantly lower than the rates of attrition from district schools among black students, black male students, Hispanic students, and students eligible for free or reduced price meals. Among Hispanic males at these KIPP schools, the attrition rate was approximately equivalent to the rate found at schools in local districts.

---

[19] See Nichols-Barrer et al. (2012) for a detailed investigation of this issue. Both our lottery-based and matching approaches prevent this potential bias from being a factor by retaining all students in the treatment group, even if they leave (or, in the case of the lottery-based analysis, never attend) KIPP.

[20] We defined attrition to include school transfers (either in-district or out-of-district) that occur during or immediately after each grade served by KIPP. For a given grade level, the attrition rate is equal to the number of transferring students divided by the total number of students who attended the school in that grade at the beginning of the year. To measure the cumulative attrition rate between grades 5 and 8, we used these grade-specific attrition rates to derive the cumulative probability that a given student will change schools before completing 8th grade. We considered school-specific grade ranges and disregarded school transfers caused by a normal grade progression, such as a move from an elementary school at the end of 5th grade to a middle school in 6th grade. Given KIPP's unique grade span (beginning in 5th grade and ending in 8th grade), comparing the cumulative attrition rates in this manner may overstate the levels of attrition at KIPP relative to other district schools. While some proportion of non-KIPP students also attend schools serving grades 5 through 8 inclusive (e.g., K-8 or K-12 schools), the majority attend an elementary school through 5th grade and then a middle or secondary school the following year. For these students, our definition of attrition does not allow for the possibility of attrition in the year the student completed 5th grade and moved on to 6th grade in another school. In other words, disregarding the "forced" school transfers occurring over the grades covered by our analyses may overlook attrition that would have otherwise occurred.

[21] In addition, we deconstructed this attrition in two ways—within-district attrition, where "movers" leave a given school to attend another school in the same district, and out-of-district attrition, where "leavers" exit a school to attend a private or other school in a different district. We found that KIPP students have slightly higher rates of within-district attrition than district schools (22 versus 18 percent) and lower rates of out-of-district attrition than the district as a whole (15 versus 19 percent). Conversely, when we limit the comparison to middle schools most commonly attended by students from KIPP feeder schools, we found that there are relatively fewer within-district movers and more out-of-district leavers at KIPP (for more details, see Nichols-Barrer et al. 2012).

Our prior working paper also examined the characteristics of students who leave KIPP and the students who transfer into KIPP middle schools in later grades. We found that at both KIPP schools and non-KIPP district schools the students who transfer out tend to be lower-performing than their peers who stay. We also found that, while KIPP schools "backfill," or admit a substantial number of late entrants in grade 6, they admit fewer new students in grades 7 and 8 than do nearby district schools. The late entrants at KIPP schools tend to have higher baseline achievement, are less likely to be male, and are less likely to receive special education services than the rest of the KIPP student body (Nichols-Barrer et al. 2012).

With respect to grade repetition, we found a systematic difference in the frequency with which students repeat a grade at KIPP schools relative to district schools (Table III.2). This pattern is especially evident in 5th and 6th grades, when KIPP's grade repetition rates are much higher than district rates. In grades 7 and 8, the differences between KIPP and other district schools are less pronounced. At KIPP schools, about nine percent of students are retained in grade 5, compared with only two percent at district schools. The proportion retained drops to four percent at KIPP schools in grade 6, three percent in grade 7, and only one percent in grade 8; the proportion of students being retained at district schools remains steady at two percent across all grades. In grade 8, the proportion of students retained at KIPP schools is significantly lower than the percentage at district schools. The higher rates of grade repetition may result from the KIPP instructional model, which generally holds that students should be promoted to the next grade only after they have demonstrated mastery of grade-specific material. Repeating a grade represents a different approach to addressing the needs of underperforming students, which involves a dramatic expansion in instructional time and resources.

Table III.2. KIPP and District Grade Repetition Rates, by Grade

	KIPP	District
Grade 5	0.09 N = 19,718	0.02** N = 5,671,207
Grade 6	0.04 N = 17,174	0.02** N = 5,464,023
Grade 7	0.03 N = 12,712	0.02** N = 4,910,753
Grade 8	0.01 N = 8,963	0.02** N = 4,219,289

Notes: Grade repetition represents the average proportion of each grade's students who will be retained in the same grade the following year. N = the number of students in the sample (grade repeaters plus non-repeaters).

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

C. What Are the Characteristics of KIPP Schools?

More than a specific set of procedures and practices, KIPP describes itself as a model defined by a core set of operating principles—Five Pillars—described in more detail in Chapter I. Within this framework, individual KIPP principals have broad autonomy to set the direction of their

schools. In this section, we describe the characteristics and practices these principals implement at KIPP schools. We focus on three basic characteristics:[22]

1. **Operational characteristics**. Key features of the way the schools operate, such as the location and size of the schools and the amount of time students spend in school.
2. **Academic climate and school climate.** Nuanced aspects of the schools' culture, such as systems used to manage student behavior at the school and requirements students and parents are asked to meet when students enroll.
3. **Staff.** Characteristics of the staff itself and staffing practices at KIPP schools.

We also examined how KIPP schools in our study samples (those for both lottery and matched comparison analyses) and the KIPP schools not included in either study sample compare to the universe of KIPP schools operating as of the 2009–10 school year.

The schools that formed the study samples for the lottery and matched comparison analyses had to meet specific criteria to be included; as a result they differ from the full population of KIPP schools in measurable ways. By design, to be eligible for the lottery sample, KIPP schools had to be oversubscribed, meaning they had more applicants than available slots for students at the time of the lottery. These schools constituted a minority of KIPP schools at the time of study intake. KIPP schools in the matched comparison analysis had to meet two criteria: (1) be open as of the 2009–2010 school year to ensure that a minimum of two cohorts of students per school would be observed; and (2) be located in jurisdictions (states or school districts) that provided at least three consecutive years of complete, longitudinally linked student-level data for traditional public and charter schools. Comparisons of both these study samples to KIPP schools not included in the study informed the external validity of study estimates; that is, the extent to which estimates from the study are representative of the population of all KIPP schools in operation as of the 2009–10 school year.

In general, there are few differences between KIPP schools in the lottery and matched comparison analyses and the full population of KIPP schools. KIPP schools are generally small and urban, serve a population that is high-minority and high-poverty, and share many key practices such as a lengthy school day and year. Though KIPP schools share many common features, schools in the lottery sample differ from the full population of KIPP schools on some dimensions—notably, lottery schools are a relatively older (in terms of years in operation), and therefore a more established group than the full population of KIPP schools. The analyses do not imply that the lottery schools are better or worse than the full population of KIPP schools, but there are enough differences that it is probably not appropriate to generalize results of the lottery analysis to the full population of KIPP schools. On the other hand, the KIPP schools in the matched comparison analysis are similar to the full population of KIPP schools, suggesting the possibility of generalizing the matched comparison estimates to the full population of KIPP schools. Any generalizations should be made with caution,

[22] We also explored average student characteristics at KIPP schools at the school-level, since the estimates in Section A above were at the student-level. The findings were similar to those in Section A and we found no differences in average student characteristics across the different study groups. The details of these analyses are found in Appendix C.

however, as there may be unobserved differences between KIPP schools in the matched comparison analysis and the full population of KIPP schools.

Below we describe the characteristics of KIPP schools, comparing the matched comparison and lottery samples to the larger population of KIPP schools. All differences are in relation to the full population of KIPP schools. Only differences that reach statistical significance are discussed, except where noted. We were unable to collect comparable data on the full set of these characteristics at neighboring schools. However, using the lottery sample, we can compare a limited number of characteristics of KIPP schools to those of the non-KIPP schools attended by non-winners—in other words, the schools KIPP students would have attended had they not enrolled in KIPP.

1. Operational Characteristics of KIPP Schools

Operational characteristics of KIPP schools include school location, enrollment, average student characteristics, amount of time students spend in school, and other operational features of the school. Key findings are highlighted below, with details shown in Table III.3.

- **KIPP schools are located primarily in urban areas.** Of all KIPP schools, 89 percent are located in large urban areas and 4 percent are in rural areas. The remaining 7 percent are in smaller cities or larger suburbs.
- **KIPP schools tend to be young.** The average KIPP school was six years old in 2010. KIPP schools included in the lottery analysis were about two years older, on average.
- **KIPP schools are typically small.** Average enrollment at KIPP schools is 314 students; at schools in the lottery analysis, enrollment is higher (354 students). Since schools in the lottery analysis also tend to be older (thus serving more grades), this difference is expected.[23] In fact, the difference gets smaller when we looked at enrollment per grade—about 80 students at the average KIPP school and 87 students at KIPP schools in the lottery analysis. The remaining difference may reflect how lottery schools were selected for the sample (to be eligible for the lottery sample, schools must be oversubscribed, and therefore more likely to be at capacity). About four percent of students at KIPP schools enroll mid-year, but less than one percent of students at KIPP lottery schools enroll mid-year. This is also not surprising given the lottery study requirement that the number of applicants exceed the number of available slots at the time of the lottery, which means that the schools in the lottery analysis would be less likely to have open slots for mid-year enrollment. At the average KIPP school, eight percent of students withdrew mid-year, and the typical English/language arts and math class size is 28 students.

[23] Enrollment differences may be related to the fact that KIPP schools in the lottery analysis are older, on average, than the full population of KIPP middle schools. KIPP schools typically begin with a single grade (5th grade for KIPP middle schools) and add a grade in each year of operation until they serve all planned grades (a "grown out" school). As a result, KIPP schools in their first three years of operation generally serve fewer grades and may enroll fewer students per grade as schools work to recruit students.

Table III.3. Operational Characteristics of Study and All KIPP Schools

	All KIPP Schools	KIPP Matched Analysis Schools	KIPP Lottery Analysis Schools
Located in Large Urban Area (Percentage)	88.5	91.9	76.9
Located in a Rural Area (Percentage)	3.8	5.4	0.0
Enrollment			
Total enrollment (mean)	314.1	320.1	353.5*
Enrollment per grade (mean)	80.4	81.3	86.5*
Enrolled students who withdrew (mean percentage)	8.4	8.8	4.4
Enrolled students who enrolled mid-year (mean percentage)	3.9	2.3	0.8*
ELA/math class size (number of students)	27.8	27.6	27.5
Student-teacher ratio	15.7	15.9	15.1
Average Student Characteristics (Mean Percentage)			
Hispanic	27.8	33.2	40.6
White	1.4	1.5	1.4
Black	65.0	60.6	48.9
Receive free lunches	68.0	68.9	64.0
Receive reduced-price lunches	11.7	11.3	12.7
Have an IEP	9.6	9.4	9.2
Limited English proficiency	9.6	10.1	16.3
Time in School			
School day length in hours (mean)	9.2	9.2	9.3
Hours per day spent in core classes (mean)	5.1	5.1	5.2
ELA (mean)	1.6	1.6	1.6
Math (mean)	1.3	1.3	1.3
Science (mean)	1.1	1.1	1.2
History (mean)	1.1	1.1	1.1
Hours per day spent outside of core classes (mean)	4.1	4.1	4.0
School year length in days (mean)	191.5	192.3	195.6
School requires students to attend Saturday school (percentage)	64.0	61.1	69.2
Number of days students attend Saturday school per month (mean)	1.0	0.9	1.0
Average daily attendance (mean)	95.9	96.2	96.3
Operational Factors			
Age of school in 2010 (mean)	5.8	6.2	8.2**
School receives school-wide Title I funding (percentage)	84.0	88.9	100.0**
School serves as its own district (percentage)	43.8	37.1	33.3
School operates within KIPP regional structure (percentage)	75.5	81.6	84.6
Number of Schools in Sample with valid Data	53	38	13

Notes: Data is current as of the 2010-2011 school year. Principal survey responses are supplemented with information from the NCES CCD and 2010 KIPP School Report Card where necessary.

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

- **KIPP students spend considerable time in school.** In keeping with KIPP's "More Time" pillar, students spend a large amount of time in school: on average more than nine hours per day, 192 days per year. This includes one Saturday school day per month at the average KIPP school. For comparison, public schools in the United States have an average school day of 6.6 hours and 180 days in the school year (Snyder and Dillow 2012). Across all KIPP schools, students spend five hours a day in core subjects, including about an hour and a half in English or language arts classes daily, and four hours a day outside of core classes.
- **KIPP schools serve a population of students that is largely minority and high-poverty.** Nearly all students at KIPP schools are minorities: on average, students are 28 percent Hispanic and 65 percent black. Across all KIPP schools, more than two-thirds of students are eligible for free lunch and another 12 percent are eligible for reduced-price lunch.[24]

2. Academic Programming and School Climate at KIPP Schools

We examined features of KIPP schools related to academic programming and school climate, including the manner in which classes are organized at KIPP schools, the proportion of schools not using a math textbook, and the prevalence of enrichment activities and limited-English programming. We also measured behavior and behavior systems, enrollment requirements, and the level of parent involvement. Some of these characteristics may be more consistent across KIPP schools, stemming from core features of the KIPP model, whereas others may vary given the autonomy of individual principals. Although these features are often challenging to measure, they help to describe the nature of the schools' approach. Key findings are highlighted below, with details shown in Table III.4.

- **There are some common programming features at KIPP schools.** Almost three-quarters (74 percent) of principals report that all core classes at KIPP schools include students of mixed ability levels. Virtually all KIPP schools offer a music and/or art program (96 percent), and almost half offer programming for students with limited English proficiency (46 percent). Almost half of KIPP schools (46 percent) report using no math text book in 7th grade, granting teachers and schools considerable flexibility and responsibility to develop their own materials.

[24] Note that these estimates are reported on a school-level, in contrast to student-level estimates comparing the characteristics of students attending KIPP schools to those at district schools. The data for these comparisons is also drawn primarily from survey results (in contrast to the student-level comparisons, which are based on school records data). As a result, we observed small differences between the KIPP-only comparisons provided here and those in the KIPP student-level descriptive analysis presented earlier in the chapter.

Table III.4. Academic Programming and School Climate at Study and All KIPP Schools

	All KIPP Schools	KIPP Matched Comparison Analysis Schools	KIPP Lottery Analysis Schools
Method of Organizing Classes (Percentages)			
All core classes have students with mixed ability levels (ELA or math)	74.0	77.8	83.3
Students loop through multiple grades with teacher	25.5	17.1	8.3
School uses interdisciplinary teaching	21.7	17.1	25.0
School uses paired/team teaching	36.7	30.6	33.3
Primary 7th grade math textbook is "no textbook"	45.5	41.9	55.6
Enrichment Programming Offered (Percentages)			
Talented/gifted program for core subjects	5.9	5.4	0.0
Music and/or art program	96.0	94.6	92.3
Before- or after-school programming	72.3	74.3	91.7*
Individual tutoring	75.0	72.2	75.0
Limited English Proficiency Programming Provided (Percentages)			
Limited English proficiency instruction for students	46.0	45.9	76.9**
Services for parents with limited English skills (interpreters or translations of printed materials)	97.4	100.0	100.0
School Behavior (Means)			
Index of use of school-wide behavior plan[a]	3.3	3.3	3.6**
Percentage of enrolled students expelled from school	0.5	0.4	0.3
Percentage of enrolled students suspended out-of-school	10.7	10.0	10.7
Participation Requirements (Percentages)			
Parents make participation commitments (e.g., interview, orientation session, commitment form)	48.0	61.1**	53.8
Students must sign a responsibilities agreement	76.0	77.8	76.9
Parent involvement			
Index of quality of parent/staff interaction[a] (mean)	3.1	3.1	3.2
Index of amount of parent involvement in school activities [a, b] (mean)	2.1	2.2	2.3
School provides parents weekly or daily notes about their child's progress (percentage)	86.0	83.8	92.3
Number of Schools in Sample with Valid Data	**51**	**37**	**13**

Notes: Data are current as of the 2010-2011 school year.[a] Indices are measured on a scale of 1-4, with higher values representing higher levels.[b] Index has an alpha smaller than 0.7, indicating low reliability. See Appendix B for more information on how indices were created.

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

- **School-wide behavior systems are typical at KIPP schools.** In keeping with the "High Expectations" pillar, which includes an emphasis on "formal and informal rewards and consequences," most principals agree that the three components of a school-wide behavior system are in place at their school. They are: (1) behavioral standards and discipline policies are established and enforced consistently across the entire school; (2) the school has a school-wide behavior code that includes specific positive rewards for students who consistently behave well; and (3) the school has a school-wide behavior code that includes clear consequences for students who violate rules. Agreement is measured on an index scaled from 1–4 based on the extent to which

the school principal agrees that these policies are in place, with one indicating strong disagreement and four indicating strong agreement.[25] Overall, KIPP schools have high scores on this index, with a mean score of 3.3. Principals at schools in the lottery analysis are particularly likely to strongly agree that their school has implemented these policies, with a mean score of 3.6.

- **Students at three quarters of KIPP schools and parents at about half are required to make participation commitments before students enroll, as reported by principals.** The "Choice and Commitment" pillar emphasizes that students and parents have a choice to enroll in a KIPP school and that everyone at the school (leaders, teachers, students and parents) make a commitment to do their part to achieve success. After the admissions lottery determines which students are to be offered admission (if applicable), one way KIPP schools implement this principle is by asking parents and students to sign commitment agreements during a home visit conducted by school staff. Almost half of KIPP principals (48 percent) report that their schools have such participation requirements for parents, and principals at more than three-quarters of schools (76 percent) report that students must sign a responsibilities agreement. Principals at KIPP schools in the matched comparison analysis are significantly more likely to report these participation requirements for parents (61 percent) than principals at all KIPP schools.

3. Staff at KIPP Schools

Next, we examined policies related to staffing at KIPP schools and staff characteristics. Staff members at KIPP schools, particularly principals, are given considerable autonomy to shape school practices. We aimed to understand the characteristics of these staff and the challenges they face related to staffing and operating KIPP schools, as these may influence school culture and operations. Key findings are highlighted below, with details shown in Table III.5.

- **Principals at KIPP schools have limited experience in that particular role; teachers have a bit more experience, on average.** Principals at KIPP schools have an average of 2.5 years of experience as principals (in their current schools or previous schools) and 7.1 years of teaching experience. For context, comparable public school principals have an average of 7 years of principal experience (Battle 2009).[26] Meanwhile, roughly half of teachers at the average KIPP school have 4 or more years of teaching experience. Most KIPP teachers also meet state certification requirements (75 percent, on average).

[25] More information on how this and other indices were created is detailed in Appendix B.

[26] Statistic is for principals serving student populations in which 75 percent or more of students were eligible for free or reduced-price lunches.

Table III.5. Staff at Study and All KIPP Schools

	All KIPP Schools	KIPP Matched Comparison Analysis Schools	KIPP Lottery Analysis Schools
Number of Full-Time Teachers (Mean)	20.6	20.8	23.8**
Staff Experience and Qualifications (Means)			
Experience of principal			
Number of years as principal	2.5	2.7	2.6
Number of years of teaching experience before becoming a principal	7.1	6.7	9.1
Teachers with more than four years of experience (percentage)	50.3	51.3	61.1
Teachers at school with full state certification (percentage)	74.9	74.0	68.9
Principal Time (Means)			
Principal time on work-related activities (hours per week)	74.0	72.9	69.3*
Index of frequency of principal time on problematic issues[a]	2.5	2.4	2.2*
Index of principal satisfaction [a]	3.0	2.9	3.1
Staff Turnover and Vacancies			
Number of principals at the school in the past three years (mean)	2.0	2.1	2.0
Teacher turnover (teachers who left the school during the last school year as percentage of full-time teachers)	21.1	20.1	12.2**
Principal reports difficulty obtaining suitable replacements is a barrier to dismissing poor-performing teachers (percentage)	50.0	51.4	23.1*
Number of teacher vacancies on Oct. 1, 2010 (percentage of full-time teachers)	5.1	5.0	9.4
Among Principals Reporting Difficulty Filling Vacancies, Percentage Reporting Top Three Reasons			
Applicants were not a good fit for school culture/goals	61.0	60.7	66.7
Applicants were not qualified	95.1	96.4	100.0
Vacancies were in a high-need or shortage area	31.7	42.9*	22.2
Compensation			
Midpoint of $ teacher salary range at school (mean)	58,114	57,581	62,550
School provides teacher incentive pay			
In "hard-to-staff" locations (percentage)	17.6	18.9	23.1
In "hard-to-staff" subjects (percentage)	18.0	19.4	23.1
For excellence in teaching (percentage)	51.0	48.6	53.8
Teachers covered by collective bargaining (percentage)	7.8	5.4	7.7
Teacher Coaching (Means)			
Index of intensity of new teacher coaching [b, c]	4.6	4.7	4.5
Index of intensity of experienced teacher coaching [b, c]	4.3	4.4	4.4
Number of Schools in Sample with Valid Data	**53**	**38**	**13**

Notes: Data is current as of the 2010-2011 school year. [a] Indices are measured on a scale of 1-4, with higher values representing higher levels. [b] Indices are measured on a scale of 1-5, with higher values representing higher levels of intensity. [c] Index has an alpha smaller than 0.7, indicating low reliability. See Appendix B for more information on how indices were created.

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

- **Principals at KIPP schools spend a substantial amount of time on the job.** The average KIPP principal reports spending 74 hours per week on work-related activities: over 12 hours per day, six days per week. The amount of time principals report spending on three problematic issues—complaints from parents, conflicts among teachers, and individual teacher complaints—is measured using an index scaled from 1–4, with higher

values of the index indicating that the principal spends more time dealing with problematic issues in these categories. KIPP schools scored a 2.5 on average, indicating principals typically spend at least an hour on each issue between one and five times in a typical month. Principals at lottery schools spend less time on work-related activities (5 fewer hours per week), and also report spending less time on problematic issues. These differences could be driven by principals at lottery schools implementing a more efficient management approach, facing a more limited scope of challenges at the school, or both.

- **On average, KIPP schools lost about a fifth of their teachers during the 2010–2011 school year (for any reason), and the large majority of principals reported difficulty filling vacancies at their schools.** Teacher turnover at KIPP schools was 21 percent during the 2010-2011 school year, but turnover was considerably lower at KIPP schools in the lottery analysis (12 percent).[27] For comparison, a national estimate found annual teacher turnover among all full-time teachers at public schools to be about 15 percent (Keigher, 2010).[28] Most principals (86 percent) reported that teacher vacancies are difficult to fill. Among schools reporting difficulty filling vacancies, 95 percent cite applicants being insufficiently qualified, 61 percent said applicants are not a good fit for the school culture or goals, and 32 percent reported that vacancies were in a high-need or shortage area. KIPP schools in the matched comparison analysis are significantly more likely to report that vacancies are difficult to fill because they are in a high-need or shortage area (43 percent). In keeping with the smaller proportion of vacancies facing KIPP schools in the lottery analysis, these schools are also less likely to cite difficulty finding suitable replacements as a barrier to dismissing poor-performing teachers.

4. Characteristics of KIPP Schools Versus Neighboring Schools

While it is useful to understand the characteristics of KIPP schools on average, any impacts we estimate should ultimately stem from *differences* in students' school experiences at KIPP relative to other schools. It is important to understand the nature and extent of these differences. For the lottery sample, we explored a subset of the characteristics discussed above at the schools attended by lottery winners versus those attended by non-winners. The schools attended by lottery winners—most commonly the KIPP school to which the student applied—differed significantly on a range of characteristics from those attended by non-winners, which are most often traditional public schools. Most notably, lottery winners attend schools that are significantly smaller than those attended by non-winners, and have significantly smaller proportions of white and Latino or Hispanic students. Schools attended by lottery winners have a higher proportion of black students, on average, but this difference is not statistically significant. For more detail on this analysis, see Appendix C.

[27] Teacher turnover is measured as the number of teachers leaving the school (for performance-related or other reasons) during or following the 2010–2011 school year, as a percentage of the total number of full time teachers at the school.

[28] The estimates are similar for teachers in both urban and non-urban schools. Note that this estimate is based on data collected during the 2008-2009 school year and comparisons to the KIPP data (reflective of the 2010-2011 school year) should be made with caution.

IV. KIPP'S IMPACTS ON TEST SCORES

This chapter presents the impact of KIPP schools on student achievement tests, including state assessments in math and reading, which are typically administered annually in the spring of the school year. To combine impact estimates across schools in states with different tests, we standardized these test scores by subject, grade, and year using information from the entire sample of students in each KIPP district (or districts). In this way, each student's score reflects his or her performance relative to a local reference group taking the same exam.[29] In Section A below, we used the study's matching design to calculate KIPP's achievement impacts across a nationwide sample of 41 middle schools. Section B presents lottery-based impact estimates for a smaller number of schools and students, and confirms that (when applied to the same schools and students) impact estimates based on matching produced the same conclusions as those based on randomized admission lotteries. In Section C, we move beyond high-stakes state tests—exams that may have consequences for students, teachers, or schools—and analyze a different measure of academic achievement, presenting KIPP's impact on students' performance on the nationally-normed TerraNova exam administered to students who participated in KIPP admission lotteries. This is a low-stakes test unlikely to be affected by "teaching to the test," and includes items assessing students' higher-order thinking skills.[30]

A. How Does KIPP Affect Student Scores on State Assessments?

Tables IV.1 and IV.2 summarize the estimated impacts of 41 KIPP middle schools on students' state test scores in four subjects, one to four years after students first enter KIPP, based on the analysis using a matched comparison group. In reading and math, both of which are tested annually in our data, we reported a separate impact estimate for each outcome year. Science and social studies are not tested in all schools or every school year; for these two subjects, we analyzed impacts as represented by the latest available middle school score in each jurisdiction. To obtain these results, we used the study's benchmark matching approach to estimate an impact for each of the KIPP schools in our sample. Using these school-specific impact estimates, we then calculated the average KIPP effect across the entire sample of KIPP middle schools.[31]

As explained in Chapter II, we estimated KIPP impacts by comparing the results of KIPP students to those of a matched comparison group (controlling for two years of prior achievement

[29] Each student's scores were converted to z-scores defined relative to the distribution of scores in the relevant jurisdiction, grade, and year. Each z-score represents the number of standard deviations above or below the jurisdiction's mean test score in that subject, grade and year. For more information on this procedure, see Appendix D.

[30] For example, according to information provided by the publisher (McGraw Hill), the reading component of the TerraNova test includes constructed response items expected to take approximately 40 percent of test-takers time. To achieve an advanced level on this component of the test, students must show that they can "recognize literary concepts such as mood, draw conclusions from more challenging text, and make connections between writers' experiences and perspectives…. and provide full justification or support for their answers."

[31] To calculate the average impact, we assigned an equal weight to each KIPP school in the sample. We also examined the sensitivity of the impact estimates to an alternative way of weighting KIPP schools (weighting by sample size), and found that it made little difference. For more details regarding how the average KIPP impacts and standard errors were obtained from school-level impacts, as well as results estimated with alternative weights, see Appendix D.

and various other student characteristics), and addressed complications arising from student attrition out of KIPP schools, grade repetition, and patterns of missing baseline data. As we report later in this chapter, the causal validity of this matching approach is reinforced by its success in replicating randomized lottery-based impact estimates in a subset of KIPP schools (and its success in replicating randomized lottery-based impact estimates in several other studies).

As shown below, the average impacts of KIPP middle schools on student performance on state assessments are positive, statistically significant, and educationally meaningful in all academic subjects we analyzed.[32] These impact estimates suggest several key results for the 41 schools in the matching sample, which are summarized below.

Key finding: The average impacts of KIPP middle schools on student achievement are positive and statistically significant in all of the academic subjects we examined.

After one year, KIPP middle schools have statistically significant positive impacts of 0.15 standard deviations in math and 0.05 standard deviations in reading (Table IV.1). In each later outcome year, we report the cumulative impact associated with being in the KIPP treatment group regardless of whether these treatment group students remained enrolled in KIPP schools. After two years the average estimated impact rises to 0.27 standard deviations in math and 0.14 standard deviations in reading. Impacts in both subjects remain positive and statistically significant in the third and fourth outcome years as well, although the sample of included schools is smaller than in the earlier outcome years.[33] After three years since entry into KIPP, the average impact increases further to 0.36 in math and 0.21 in reading. Impacts remain positive and significant four years after entry into KIPP, with effect sizes of 0.31 in math and 0.22 in reading. Below we discuss how to interpret the magnitude of these impacts.

[32] The term "statistically significant" indicates that there is less than a five percent chance that the observed effect in our sample occurred purely by chance. We use the term "educationally meaningful" to refer to impacts that are larger than 0.12 standard deviations—equivalent to moving a student from the 45th to the 50th percentile in his or her district.

[33] In years 3 and year 4, outcome samples do not include newer KIPP schools, which have not been operating long enough to observe longer-term student outcomes. Specifically, the sample declines from 41 KIPP schools in the first two outcome years to 38 schools in year 3. In year 4, there are 34 schools in the reading impact sample and 28 in the math impact sample. Similarly, the most recent cohorts at all KIPP schools are not included in the year 3 and year 4 outcome samples.

Table IV.1. Mean Test Score Effects in Mathematics and Reading, Benchmark Model

Outcome	Year 1	Year 2	Year 3	Year 4
Math impact	0.15** (0.01)	0.27** (0.01)	0.36** (0.01)	0.31** (0.02)
Number of KIPP schools	41	41	38	28
Reading impact	0.05** (0.01)	0.14** (0.01)	0.21** (0.01)	0.22** (0.01)
Number of KIPP schools	41	41	38	34

Note: Regressions were performed separately for each KIPP middle school in the sample. Reported impacts are an average of equally-weighted impact estimates from regressions of middle school math and reading z-scores on indicator variables for the number of years after a student's enrollment in a KIPP middle school. After grade repetition, students were assigned the same z-score received in the last year prior to retention. The sample consists of students who enter KIPP in grades 5 or 6 matched by jurisdiction and cohort to students who never enroll in KIPP; propensity scores were generated separately by KIPP school, using two years of baseline test scores and all available demographic characteristics. Regression controls include two years of baseline z-scores in math and reading (imputed if one baseline year was missing) as well as dummy variables for demographic characteristics, grade, and cohort. Regressions use robust standard errors (in parentheses) and are clustered on student identifiers.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

Table IV.2 shows KIPP's estimated impact on state tests in science (with a sample of 25 schools) and social studies (with a sample of 19 schools).[34] For each school, the outcome scores were drawn from the highest middle school grade associated with the relevant exam (usually grade 8).[35] In these subjects, the KIPP schools in our sample have positive and statistically significant impacts of 0.33 standard deviations in science and 0.25 standard deviations in social studies.

[34] We did not receive data for these test subjects for 16 schools in science and 22 schools in social studies.

[35] For these two subjects, we used the unadjusted scores of all students who repeated a grade during middle school (i.e., the scores of grade repeaters were not imputed, as they were for math and reading outcomes), regardless of when the test was taken. Thus, if a KIPP student was retained in grade 5 and took the science exam in grade 8, the analysis uses that student's grade 8 score recorded five years after enrolling in KIPP. In addition, we do not have baseline year science or social studies scores for KIPP and comparison group students; we included baseline reading and math scores as covariates in the model instead.

Table IV.2. Mean Test Score Effects in Science and Social Studies, Benchmark Model

Outcome	Impact
Science impact	0.33** (0.02)
Number of KIPP schools	25
Social studies impact	0.25** (0.02)
Number of KIPP schools	19

Note: Regressions were performed separately for each KIPP middle school in the sample. Reported impacts are an average of equally-weighted impact estimates from regressions of middle school math and reading z-scores on indicator variables for the number of years after a student's enrollment in a KIPP middle school. After grade repetition, students were assigned the same z-score received in the last year prior to retention. The sample consists of students who enter KIPP in grades 5 or 6 matched by jurisdiction and cohort to students who never enroll in KIPP; propensity scores were generated separately by KIPP school, using two years of baseline test scores and all available demographic characteristics. Regression controls include two years of baseline z-scores in math and reading (imputed if one baseline year was missing) as well as dummy variables for demographic characteristics, grade, and cohort. Regressions use robust standard errors (in parentheses) and are clustered on student identifiers.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

A variety of alternative analyses (see Appendix D for details) produce results that likewise suggest a pattern of consistently positive and large impacts at KIPP schools. For example, impacts for both reading and math in all four outcome years remain significant and positive under a more pessimistic set of assumptions about the outcome scores of students who repeat a grade. Similarly, the results are also significant and positive when we drop all the students missing one or more prior achievement scores.

Our benchmark matching impact estimates likely underestimate KIPP's full impact on students during the time they are enrolled in KIPP, because our approach retains students in the KIPP treatment group even if they transfer from KIPP after only a year.[36] In Appendix D we show alternative results that estimate the impact of KIPP for the years students are actually enrolled at KIPP schools (that is, estimates that account for the fact that the treatment group in our main matching analysis includes some students no longer enrolled at KIPP in later years). We did not focus on these "enrolled student" estimates because they are likely to be biased for various reasons.[37] The true effect of KIPP schools on enrolled students is likely to be somewhat larger than our benchmark estimate and may be somewhat less than the estimates in Appendix D.

[36] Our analysis sample did not include students in the treatment group if they entered KIPP but left the school during their first year, before they appear as KIPP students in our data files. In most districts, the data reflect students' schools at the beginning of the school year. But in a smaller number of districts the data only identify students' schools at the time they take the state assessment; in these districts, students who leave KIPP schools prior to their first state assessment would not be included in the treatment group.

[37] These alternate estimates (in Appendix D) adjust the benchmark results by dividing the marginal KIPP impact in each year by the percentage of students in the treatment group who remained enrolled at KIPP in that year. This adjustment relies on the unlikely assumption that students experience no positive or negative KIPP effects after departing KIPP.

Key finding: The magnitude of KIPP's achievement impacts is substantial.

For the full matching sample of 41 KIPP schools, the average impact three years after enrollment is 0.36 standard deviations in math, which is equivalent to moving the KIPP students in our sample from the 44th percentile to the 58th percentile (Figure IV.1).[38] Another way of interpreting these impact estimates is to compare KIPP effect sizes to national norms regarding the amount of student academic growth that takes place during middle school (Bloom et al. 2008). Expressed this way, our impacts suggest that on average, KIPP middle schools produce approximately 11 months of extra learning growth in math after three years. For comparison, in study districts there is a gap of 0.90 standard deviations between the average math test scores of black students and white students; students eligible for reduced-price school meals have math scores that are an average of 0.77 standard deviations lower than other students.[39] In other words, the size of the math impact produced by KIPP schools after three years is equivalent to about 40 percent of the local black-white test score gap and 47 percent of the local achievement gap between higher and lower income students.

The average impact of KIPP after three years in reading (0.21 standard deviations) is somewhat smaller than that for math—equivalent to moving the KIPP students in our sample from the 46th to the 55th percentile. This is consistent with a variety of other studies that have found reading scores to be more difficult to move than math scores.[40] Compared to national norms, the estimated reading impact after three years represents approximately eight months of additional learning growth (Bloom et al. 2008). Black students in study districts have reading test scores that are 0.82 standard deviations lower than the scores of white students; students eligible for free or reduced-price school meals underperform other students by 0.72 standard deviations. Thus, after three years, the size of the KIPP impact in reading is equivalent to 26 percent of the local black-white disparity in reading scores, or 29 percent of the gap between higher and lower income students.

The impact estimates in science and social studies fall between those in math and reading. The estimated impact in science (0.33 standard deviations) is equivalent to moving KIPP students from the 36th to the 49th percentile, and the estimated impact in social studies (0.25 standard deviations) is equivalent to moving students from the 39th to the 49th percentile. Compared to national norms, KIPP's estimated impacts represent an accumulation over four years of an extra 14 months of learning growth in science and an extra 11 months of learning growth in social studies (Bloom et al. 2008). Expressed in terms of achievement gaps, KIPP's estimated impacts are equivalent 34 percent of the local black-white disparity in science and 28 percent of the disparity in social studies, or 39

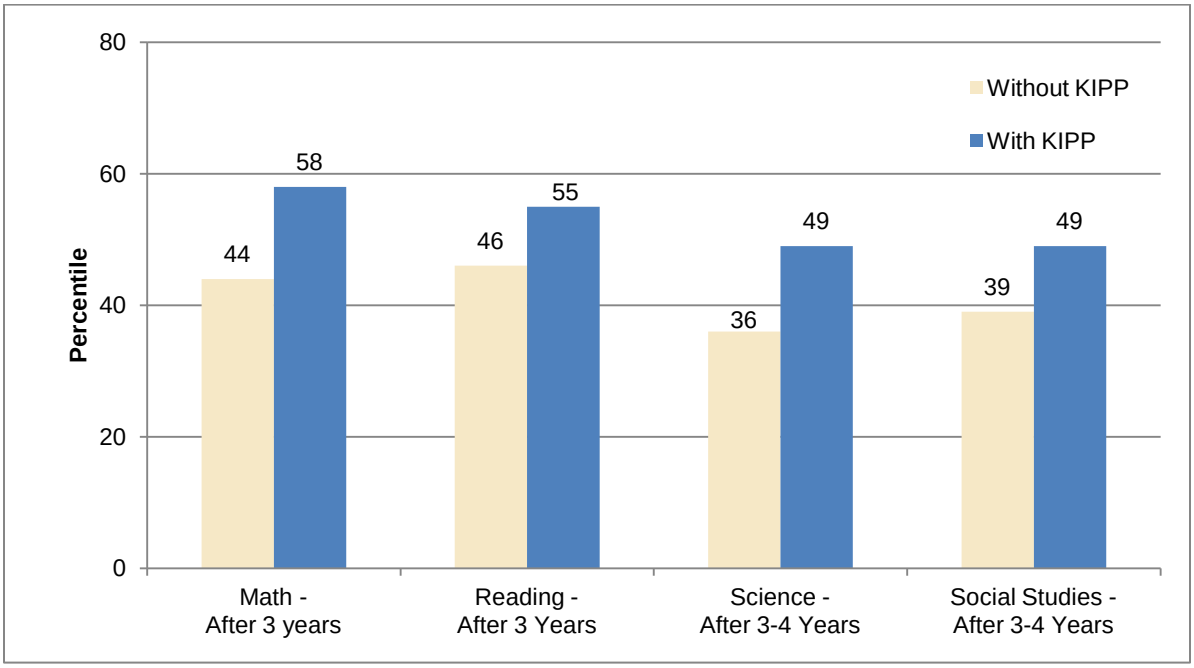
[38] On average, three years after enrolling in a KIPP school the students in our sample scored at the 58th percentile for their jurisdiction in math. Our impact estimate in math (0.36 standard deviations) implies that these students would have scored at the 44th percentile if they had never enrolled at KIPP. We performed an analogous calculation in reading, science, and social studies and show the results in Figure IV.1

[39] In math and reading, these benchmarks refer to the average difference in test scores in grades 6 and 7, for all students in the school districts and cohorts that are eligible for the study's matching analysis. In science and social studies, the benchmarks refer to the average difference in grade 8 test scores, in the school districts with data on those subjects.

[40] This pattern of larger effect sizes in math than in reading is consistent many experimental studies of educational interventions (Decker et al. 2004; Dobbie and Fryer 2009; Abdulkadiroglu et al. 2011; and Furgeson et al. 2012).

percent of the gap between higher and lower income students in science and 31 percent of the gap in social studies.[41]

Figure IV.1. KIPP Estimated Impacts on Student Achievement in Percentiles, by Subject



Note: For math and reading, the figure shows the impact of KIPP on the scores of tests taken three years after enrollment in a KIPP school; for science and social studies, the figure shows the impact on scores of tests taken three years after enrollment for some student cohorts and four years after enrollment for other student cohorts. The blue bar represents the mean percentile rank of KIPP students in the relevant analysis sample, relative to the local jurisdictions. The beige bar represents this observed mean rank minus the average KIPP impact estimate in each subject. In all four subjects, the difference in percentiles represents an impact that is statistically significant at the 0.05 level, two-tailed test.

These effect sizes are consistent with findings on high-performing charter schools in other studies. A lottery study of New York City charter schools estimated annual achievement impacts of 0.09 standard deviations in math and 0.06 standard deviations in reading (Hoxby et al. 2009). If these schools accumulate such impacts annually over three years, the effects would amount to 0.27 standard deviations in math and 0.18 standard deviations in reading—less than KIPP schools are producing in math and reading. KIPP impacts more closely resemble the results from studies of Boston charter schools, where these middle schools are estimated to produce annual achievement impacts of 0.18 in math and 0.09 in reading (Abdulkadiroglu et al. 2011); if these impacts were sustained over three years, Boston charters would somewhat outperform the KIPP average in math and reading. Evidence on the impacts of other charter-school management organizations (CMOs) suggests that KIPP is among the highest-performing charter network in the country. In a national quasi-experimental study of the impacts of 22 different CMOs, Furgeson et al. (2012) found that after three years, the average CMO had an impact of 0.15 in math and 0.05 in reading (neither effect

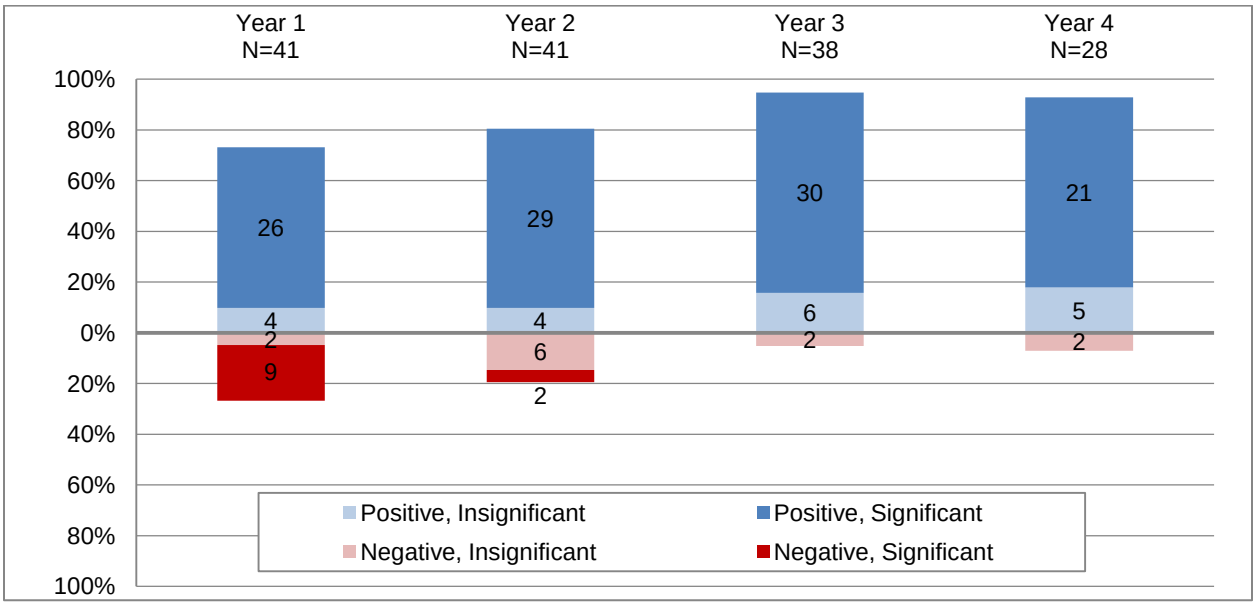
[41] Compared to white students in study districts, on average black students score 0.98 standard deviations lower in science and 0.87 standard deviations lower in social studies in grade 8; students eligible for free or reduced-price school meals underperform other students by 0.84 standard deviations in science and 0.77 standard deviations in social studies.

was statistically significant). The average impacts of KIPP schools are therefore much larger than those of most other CMOs in that study.

Key finding: Most KIPP middle schools have large positive impacts in multiple subjects.

One of KIPP's underlying aims is to produce consistently positive results across its schools. To shed light on whether KIPP is achieving this objective, we estimated impacts separately for each KIPP middle school in the sample.[42] Taken individually, KIPP schools show a consistent pattern of positive impact estimates in both reading and math, particularly in later outcome years. Figures IV.2 and IV.3 show the percentage of KIPP schools in each outcome year with math or reading impact estimates that are positive, negative, or statistically insignificant. In math after two years, 29 schools have significant positive impacts, two schools have significant negative impacts, and 10 schools have impacts that are not statistically significant. In reading after two years, 25 schools have significant positive impacts, two schools have significant negative impacts, and 14 schools have impacts that are not statistically significant. After three years and after four years, none of the schools remaining in the sample have significant negative impact estimates in either reading or math.

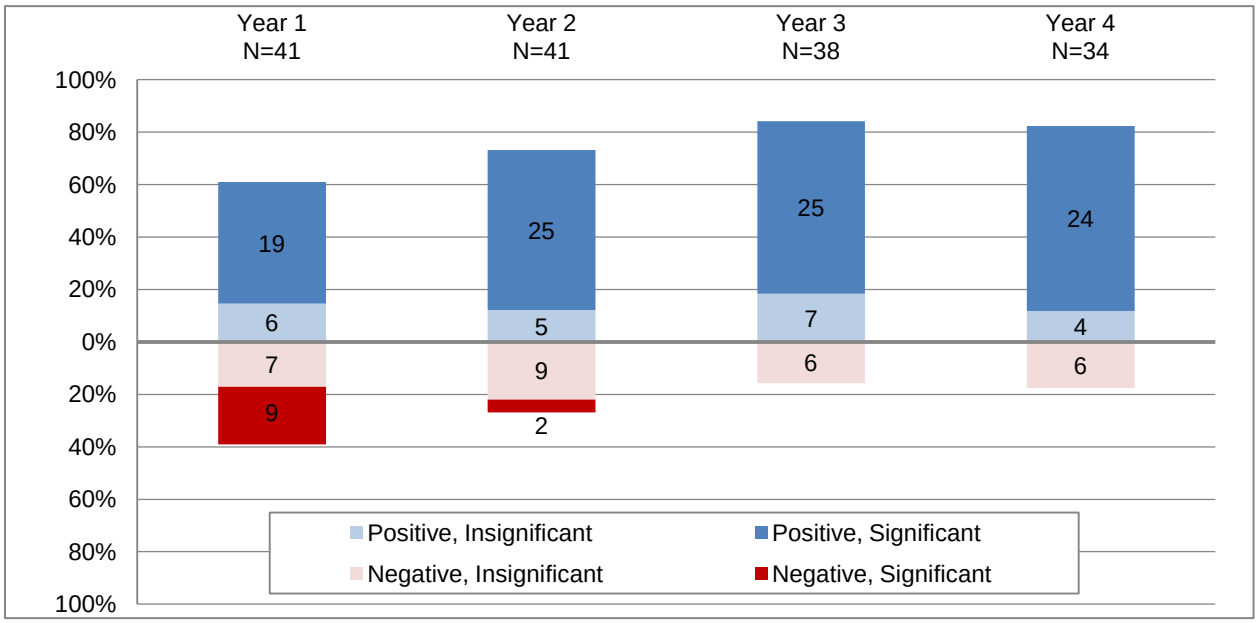
Figure IV.2. Percentage of KIPP Schools with Significant Effects in Math, by Year



Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact estimate is positive versus negative in a given year. Dark-blue and beige bars indicate results that are considered statistically significant at the 0.05 level, two-tailed test (where blue is positive and beige is negative).

[42] The sample size for each of the school-level impact estimates is largely determined by the number of student cohorts in the sample for each school. The total sample size for each of the schools ranges from 95 to 787 KIPP students; the standard error of the impact estimates for each school range from 0.02 to 0.15 standard deviations. In both math and reading, the median standard error for these school-specific impact estimates is approximately 0.05. To adjust for the different levels of precision in these estimates, we produced empirical Bayes shrinkage estimates of the school-specific impacts, following the approach described in Miller (1983). This adjustment moves less precise impact estimates closer to the average KIPP impact in each test subject and outcome year. For a more detailed description of this adjustment, see Appendix D.

Figure IV.3. Percentage of KIPP Schools with Significant Effects in Reading, by Year



Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact estimate is positive versus negative in a given year. Dark-blue and beige bars indicate results that are considered statistically significant at the 0.05 level, two-tailed test (where blue is positive and beige is negative).

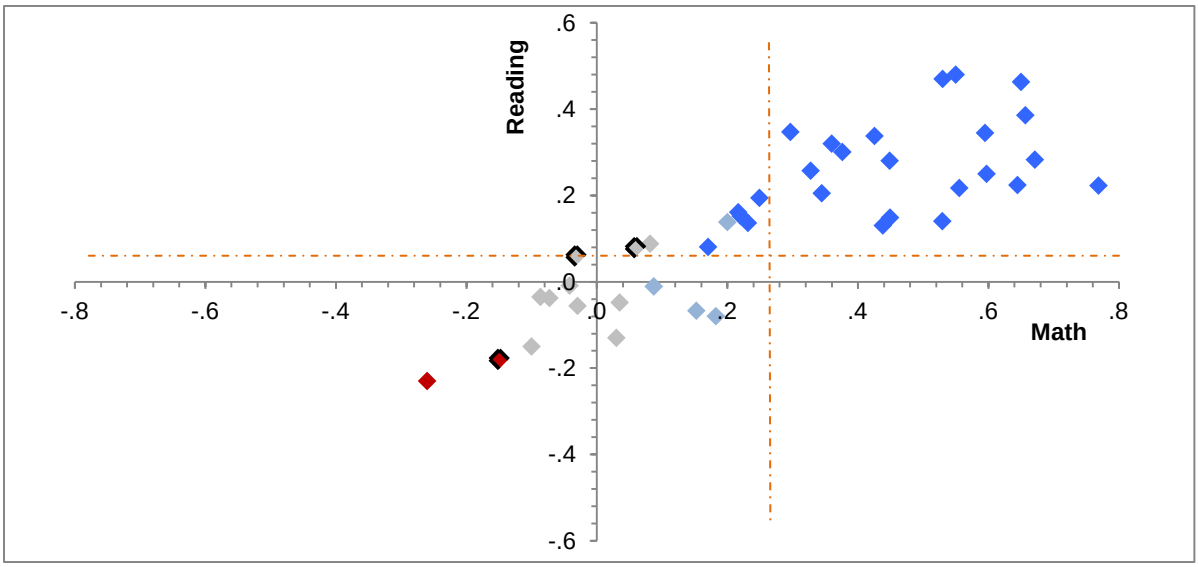
To further illustrate the magnitude of cumulative reading and math effects at KIPP schools, we also included two scatter plots that show estimated impacts in both subjects after two years (Figure IV.4) and after three years (Figure IV.5). In these figures, each diamond represents a different KIPP school, with color coding to indicate the sign and significance of each school's reading and math impact estimates. As shown in the figures, many individual KIPP schools produce impacts that are substantially different from the average KIPP impact estimates reported in Table IV.1. For example, after two years, we estimate that a third of KIPP schools are producing math impacts of 0.44 standard deviations or more and reading impacts of 0.22 standard deviations or more. On the other hand, the bottom third of KIPP schools are producing math impacts of 0.09 standard deviations or less (including eight schools with negative impact estimates) and reading impacts of 0.07 or less (including 11 schools with negative results), though most of the negative estimates are not statistically significant. A majority of KIPP schools in the matched comparison analysis (25 of 41) have positive and statistically significant impact estimates in both subjects.

After three years, the distribution of impact estimates becomes more positive. A third of KIPP schools produce three-year math impacts of 0.46 standard deviations or more and three-year reading impacts of 0.30 standard deviations or more. At this point, the bottom third of KIPP schools produce math impacts of 0.24 standard deviations or less (including two schools with negative impact estimates, all of which are statistically insignificant) and reading impacts of 0.12 or less (including six estimates that are negative but insignificant). Approximately two-thirds of the KIPP schools in our sample have positive and statistically significant year 3 impact estimates in both reading and math.

There are three schools in our sample that are no longer affiliated with KIPP (these schools are labeled with a black border in Figures IV.4 and IV.5). These schools show math and reading impact estimates that are either negative and significant or indistinguishable from zero after two years. In other words, the schools that lost their KIPP affiliations were producing lower-than-average impacts while they were part of the KIPP network.

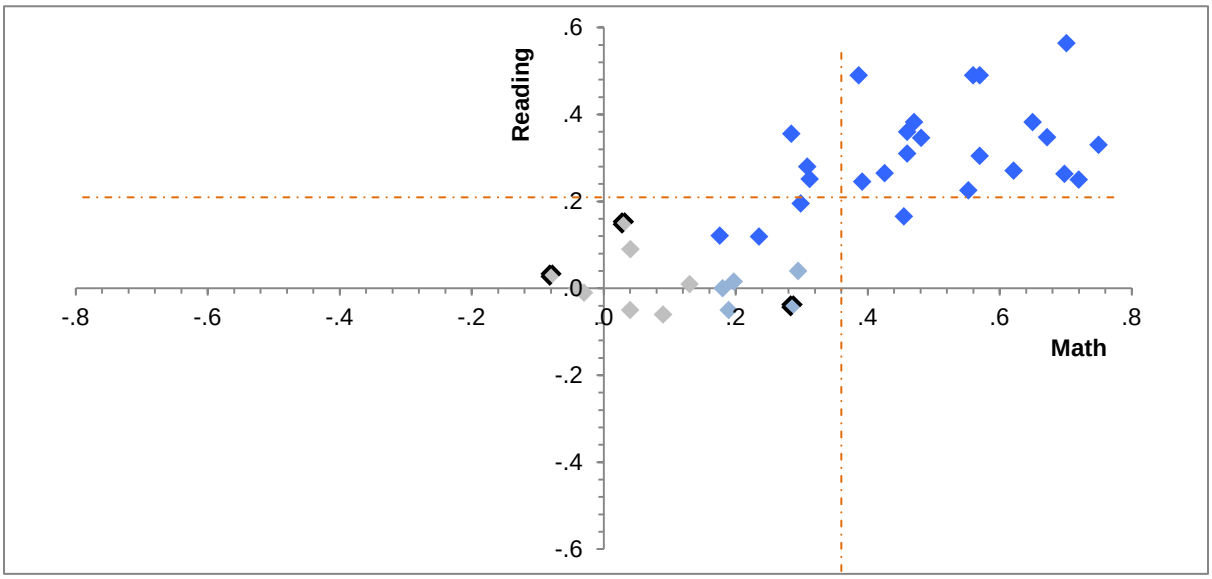
We find a similar pattern of school-level impact estimates in the sample of KIPP middle schools with data on science and social studies exams (Figure IV.6). In science, 18 schools have significant positive impacts and seven have insignificant impacts. In social studies, 12 schools have significant positive impacts and seven have insignificant impacts. No schools in the sample had significant negative impact estimates in either science or social studies.

Figure IV.4. Distribution of Reading and Math Impact Estimates After Two Years



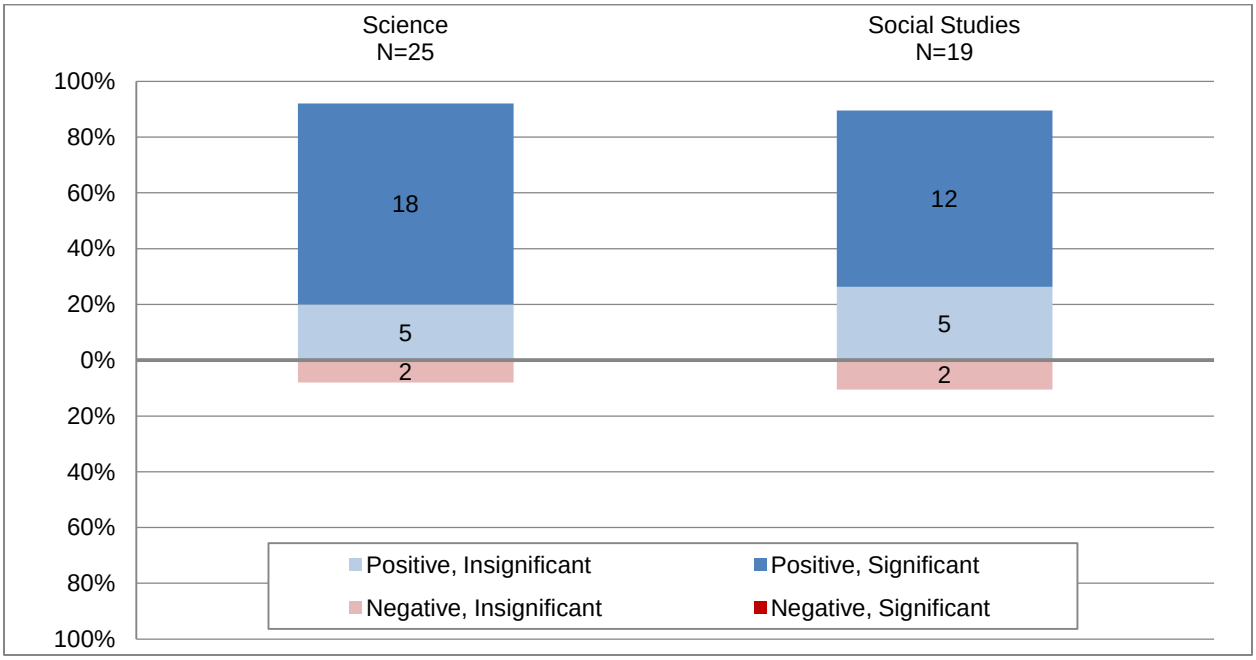
Note: Each diamond represents the math and reading impact estimate for one KIPP school. Diamonds with a black border represent schools that have lost their KIPP affiliation. Dark-blue and dark-red diamonds indicate that impacts in both subjects are statistically significant at the 0.05 level, two-tailed test (blue is positive, and red is negative). Light-blue and pink diamonds indicate that the impact in only one of the two test subjects is statistically significant. Grey indicates that both impacts are statistically indistinguishable from zero. The orange lines represent the average impacts across KIPP schools.

Figure IV.5. Distribution of Reading and Math Impact Estimates After Three Years



Note: Each diamond represents the math and reading impact estimate for one KIPP school. Diamonds with a black border represent schools that have lost their KIPP affiliation. Dark-blue and dark-red diamonds indicate that impacts in both subjects are statistically significant at the 0.05 level, two-tailed test (blue is positive, and red is negative). Light blue and pink diamonds indicate that the impact in only one of the two test subjects is statistically significant. Grey indicates that both impacts are statistically indistinguishable from zero. The orange lines represent the average impacts across KIPP schools.

Figure IV.6. Percentage of KIPP Schools with Significant Effects in Science and Social Studies



Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact estimate is positive versus negative in a given year. Dark-blue and dark-red colors indicate result that are considered statistically significant at the 0.05 level, two-tailed test (where blue is positive and red is negative).

For most student subgroups of interest, the average KIPP impact is not appreciably different from the overall average impact among all KIPP students. Using the matching approach, we also tested whether there are statistically significant differences in KIPP's impacts for students with different characteristics. Specifically, we measured the difference between KIPP's average impact on math and reading achievement among members of a given subgroup as well as those outside that subgroup. The average KIPP impact is statistically similar for nearly all the characteristics we tested. KIPP's math and reading impacts among males, black students, black males, Hispanic males, students with limited English proficiency, and special education students are not significantly different than KIPP's impacts on other types of students. But KIPP impacts tend to be significantly higher for Hispanics than non-Hispanics, although estimated impacts for each group in math and reading are positive and statistically significant in nearly all years.[43] Estimated impacts also are significantly higher for students with lower levels of prior reading achievement than for students who were higher-achieving at baseline; nevertheless, the impacts are positive and statistically significant in all years for both groups. A detailed discussion of these subgroup results can be found in Appendix D.

B. Lottery-Based Estimates of KIPP's Impacts on Student Achievement

For the subset of KIPP schools that met the participation criteria in the lottery analysis, we calculated the impact of KIPP schools by comparing the test scores of admission lottery winners to those of non-winners. As described in Chapter II, we produced two sets of lottery-based estimates of KIPP impacts: intent-to-treat (ITT) and treatment-on-the-treated (TOT). The ITT analysis yielded estimates of the effect of receiving a lottery-based offer of admission, while the TOT analysis yielded estimates of the effect of actually attending a KIPP school after receiving a lottery offer. As in the matching analysis, all test scores were standardized to allow for comparability across schools. The lottery analysis pools students across KIPP schools because the samples are not large enough to allow reliable lottery-based impact estimates for individual schools.

As discussed in Chapter II, lottery-based ITT analysis provides the estimates that have the greatest causal rigor, though they are limited to a subset of KIPP middle schools. This lottery analysis includes only 10 of the 41 KIPP middle schools in the sample for the matching analysis presented earlier. However, the average matching-based impact estimates for these 10 schools after two years (0.34 in math and 0.14 in reading) are similar to matching-based impacts for all 41-schools (0.27 in math and 0.14 in reading). Thus, the lottery analysis examines a set of KIPP schools that appears to be reasonably similar to the larger KIPP network.

Key finding: The average impact of an admissions offer to a KIPP middle school in the lottery sample is positive and statistically significant on state test scores in math and positive but not statistically significant in reading.

After one year, the average (ITT) impact of being offered admission to a KIPP middle school in the lottery sample is 0.13 standard deviations on math scores; this increases to 0.24 standard deviations in year 2 (Table IV.3). Both estimates are statistically significant. The impact on reading

[43] The exception is year 1 impacts in reading, when impacts are positive but not statistically significant both for Hispanics and non-Hispanics.

scores is positive but not statistically significant (0.02 standard deviations in the first year and 0.10 standard deviations in the second year).

The TOT impacts, measuring the impact of attending a KIPP middle school, are larger in magnitude than the ITT estimates and have the same pattern of statistical significance. Because the TOT estimates measure the impact of attending a KIPP middle school, these estimates are more directly comparable with the matching-based impact estimates presented in Section A above. In both math and reading, the magnitude of the estimated impact increases from year 1 to year 2; this is consistent with the pattern of results in the matching analysis. The TOT impact on math test scores rises from 0.19 in year 1 to 0.36 in year 2. For reading, the TOT impacts are 0.03 in year 1 and 0.15 in year 2. While the TOT impacts in reading are not statistically significant, the sample sizes in the lottery analysis (unlike the larger samples in the matching analysis) do not provide sufficient statistical power to detect impacts of this magnitude.

Table IV.3. Impact Estimates on State Assessments for Subset of Oversubscribed KIPP Schools

Outcome (z-scores)	Impact of Admission Offer (ITT)	Adjusted Impact of Attendance (TOT)
Math Achievement		
Year 1	0.11*	0.19*
	(0.05)	(0.08)
Year 2	0.22**	0.36**
	(0.06)	(0.10)
Reading Achievement		
Year 1	0.02	0.03
	(0.06)	(0.09)
Year 2	0.09	0.15
	(0.07)	(0.11)
Number of Schools with Valid Data	10	
Number of Students with Valid Data		
Year 1	536	
Year 2	441	

Note: All impacts in this table are based on regression models that pool all lottery schools and that control for baseline covariates. Standard errors are in parentheses. The ITT and TOT models are described in Appendix E.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

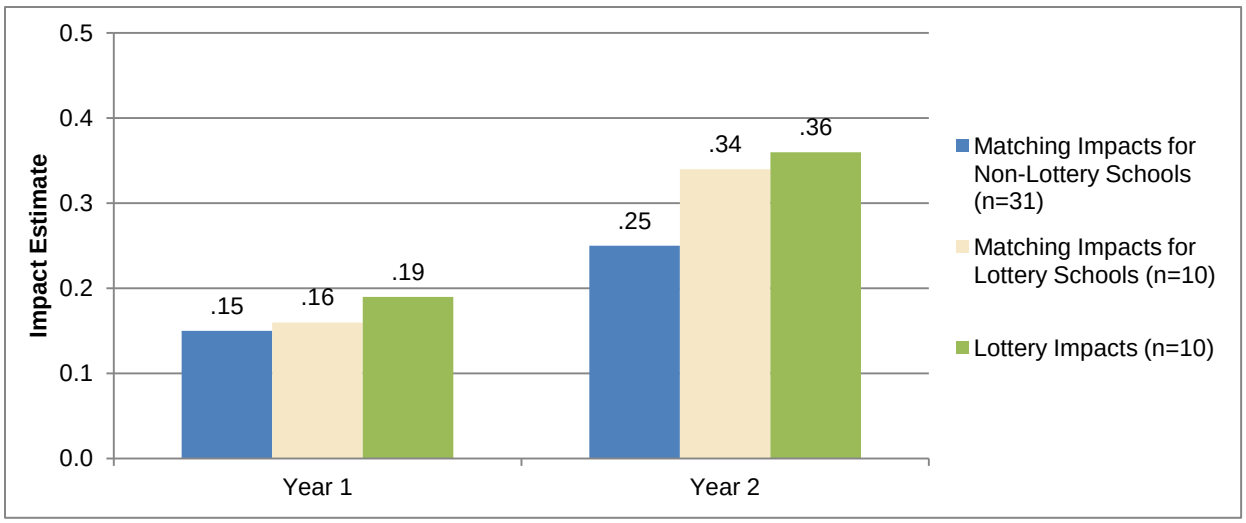
Key finding: For this sample of oversubscribed KIPP schools, lottery impact estimates are similar to impact estimates based on matching methods.

Our matching-based approach to estimating impacts for a large sample of KIPP schools could be biased if there are unobserved differences between treatment students and matched comparison students (such as student motivation or parent characteristics) that also affect academic results. To test whether such bias exists, we examined whether matching-based impact estimates replicated the

lottery-based impact estimates presented above. As mentioned above, one can compare the two sets of estimates in a variety of ways, from a simplistic comparison of the overall matching-based impact estimate with the overall lottery-based impact estimate to a careful comparison of the two sets of estimates where each is based on the same carefully constructed common sample of KIPP schools and students. While these approaches differ from one another, a consistent finding emerges that the matched comparison group analysis and lottery-based analysis produce very similar estimates of the impact of KIPP on student achievement.

Because the sample of schools and students in the matching analysis is much larger than the lottery analysis sample, a simple comparison of the overall impact estimates derived from the two approaches has the potential to be misleading. For example, the average matching-based impact estimates for the 10 schools in the lottery analysis tend to be slightly higher in mathematics than the matching-based impacts for the 31 schools that are not part of the lottery analysis (see Figures IV.7 and IV.8). A more careful way to compare the two approaches would be to examine estimates for the same KIPP schools. For the 10 schools in the lottery analysis, the TOT impact estimates correspond closely to matching-based impact estimates for the same set of schools. As shown below, analyzing effects of these 10 schools using matching (including in the sample all admitted KIPP students, regardless of whether they participated in a lottery) produces a set of impact estimates in math and reading that consistently fall within 0.05 standard deviations of the lottery-based impact estimates.[44] Because the lottery-based impacts are likely to be unbiased, this lends credibility to the matching methods we used in Section A for the larger sample of 41 KIPP schools.

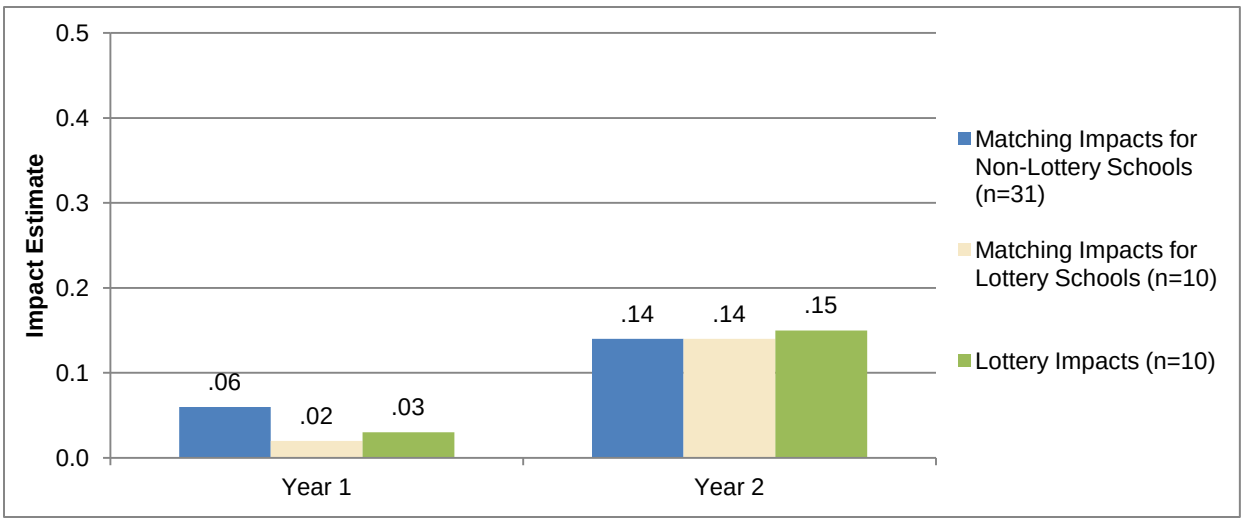
Figure IV.7. Comparison of Lottery Impact Estimates and Matching Impact Estimates in Math



Note: The green bars represent TOT impact estimates from the lottery analysis of 10 KIPP schools. The beige bars represent matching-based impact estimates for those same 10 schools. The blue bars represent matching-based impacts for the remaining 31 schools not in the lottery analysis; some of these 31 schools had admission lotteries that were not eligible for the lottery analysis.

[44] This matching analysis weights the 10 schools equally. Weighting the schools by lottery size (with school-level weights that correspond to those used in the lottery analysis) changes each of the average matching impacts by 0.02 standard deviations or less, and the estimates remain within 0.05 standard deviations of the lottery-based impacts.

Figure IV.8. Comparison of Lottery Impact Estimates and Matching Impact Estimates in Reading



Note: The green bars represent TOT impact estimates from the lottery analysis of 10 KIPP schools. The beige bars represent matching-based impact estimates for those same 10 schools. The blue bars represent matching-based impacts for the remaining 31 schools not in the lottery analysis; some of these 31 schools had admission lotteries that were not eligible for the lottery analysis.

But this simple comparison still includes a much larger sample of KIPP students in the matching analysis than the sample in the lottery analysis. In particular, the matching estimates include additional student cohorts that are not part of the study's lottery sample. In addition, within cohorts that are part of both the matching and lottery analysis samples, the lottery sample does not include some students in the matching analysis sample, such as students who were admitted outside of the lottery process because they had a sibling already attending the school. The most careful way to test for bias in the matching approach is to use the same sample of treatment students to obtain both matching and lottery-based impact estimates.

To carry out this comparison of estimates from the matching and lottery analyses, we first determined the sample of schools in which we could estimate impacts using both designs.[45] We then obtained (ITT) lottery impact estimates by comparing the state test scores of lottery winners to those of non-winners in these schools. Separately, we used the study's propensity score matching method to compare the achievement of the same group of lottery winners to a matched comparison group, selected from a pool of all other students in the same jurisdiction at each lottery school.[46] We

[45] The sample for the validation exercise is smaller than the sample in the full lottery analysis of state test scores for two reasons. First, two schools in the lottery analysis were excluded from the validation exercise because a large number of control students lost the admission lottery but were admitted to KIPP subsequently from a waitlist. In this circumstance, it is not possible for our matching approach to generate impacts that could be compared meaningfully to the lottery estimates—while the control group from the lottery included some students who attended KIPP, the comparison group in the matching design included only non-KIPP students. Second, the validation exercise required all students to have at least one baseline or pre-baseline state test score, which further reduced the sample.

[46] While the same treatment group is used in both analyses, the total sample size differs for the two approaches. Our matching procedure selects one unique comparison student for each treatment student (matching to the nearest neighbor), but the number of control students in the lottery analysis typically does not equal the number of treatment students at each school. Student weights were standardized by setting the sum of control or comparison student weights equal to the sum of treatment student weights at each school.

found no evidence of bias in the study's matching approach. For this subset of lottery schools, the study's matching method yields the same substantive conclusions as the lottery analysis: the two sets of estimates are not statistically distinguishable from each other, and the differences are small (matching estimates are 0.04 standard deviations lower than lottery estimates in math and 0.05 standard deviations higher in reading).

One caveat on this finding—due to the small samples used in this validation exercise, the statistical power of a test of the difference between the lottery and matching impact estimates is limited. In other words, the test is not likely to detect small differences in the estimates. But the finding is consistent with other studies that have successfully replicated lottery-based charter-school impact estimates using matching (Fortson et al. 2012; Furgeson et al. 2012). A detailed discussion of this validation exercise and its results are in Appendix F.

C. How Does KIPP Affect Students' Higher-Order Thinking Skills?

As discussed in Chapter II, we administered an additional test of math and reading (the TerraNova) to the lottery sample after two years of KIPP treatment. This test differs from state tests in several ways—in the TerraNova reading test, students' scores capture their higher-order thinking skills through the use of both selected and constructed (open-ended) response items. So if KIPP affects only students' basic skills, we would expect estimated impacts on TerraNova scores to be closer to zero than the estimated impacts on state test scores. In addition, the TerraNova is a low-stakes nationally-normed test, implying that students' performance has little or no consequences for students, teachers, or schools. Thus, if KIPP impacts were the result of the teachers in KIPP schools "teaching to the test," one would expect much smaller impacts on TerraNova scores. Finally, TerraNova scores allowed us to address two key analytical limitations of state test scores as measures of student achievement. First, state tests differ from state to state and so may capture different skills for different portions of the study sample; in contrast, the same TerraNova tests were administered to all students in the lottery sample. Second, state test scores are missing for certain sample members, including those who repeat a grade and those who leave the data catchment area or move to a private school. We attempted to administer the TerraNova in reading and math to all students, including those who were retained in grade or moved to a private school or different district, although not all sample members completed the exam.

Key finding: In the lottery sample, average KIPP impacts on a test with items designed to address higher-order thinking skills are similar to KIPP impacts on high-stakes state tests.

We found that impacts on the study-administered test are consistent with those on the state assessments. The impact of winning a KIPP lottery on the TerraNova math test (two years after KIPP entry), is 0.20 standard deviations and is statistically significant (Table IV.4).[47] The impact on the TerraNova reading test is 0.08 standard deviations but is not statistically significant. These estimates are similar in magnitude to the lottery impacts on state assessments presented in Table IV.3. As with the state assessments, the TOT impacts are larger than ITT but the pattern of statistical significance is the same.

[47] This effect size is calculated relative to the national norms for the TerraNova test.

Table IV.4. Impact Estimates on the TerraNova Test Administered in the Fall of the Third Follow-Up Year

Outcome (Z-Scores)	Impact of Admission Offer (ITT)	Adjusted Impact of Attendance (TOT)
Math achievement	0.20** (0.05)	0.35** (0.09)
Reading achievement	0.08 (0.07)	0.12 (0.11)
Number of schools with valid data	10	
Number of students with valid data	590	

Note: All impacts in this table are based on regression models that pool all lottery schools and that control for baseline covariates. Standard errors are in parentheses. The ITT and TOT models are described in Appendix E.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

V. IMPACTS ON STUDENT BEHAVIOR AND ATTITUDES

Recognizing that a student's long-term success depends on more than test scores, KIPP aims to improve student behavior and attitudes as well. To provide a richer picture of KIPP's effects, we complement our analysis of impacts on test score outcomes examined in Chapter IV with the first estimates of KIPP's impacts on key non-test outcomes. We used the lottery-based sample and design to estimate the impact of KIPP on a number of outcomes using student and parent surveys administered two years after the admissions lotteries.[48] These outcomes cover four domains:

1. Student engagement and effort in school
2. Educational aspirations and expectations
3. Student well-being and behavior
4. Satisfaction with and perceptions of school

Interpreting impacts on key aspects of behavior and attitudes is difficult when using multiple separate survey items, so in many cases we created indices that summarize students' or parents' responses on related data items. For example, the index for feelings about school represents a student's average response on the extent to which they agree with 10 statements about aspects of their school environment. Details on these outcomes are included in Appendix B. Paralleling the presentation of lottery-based impacts on test scores, we reported both ITT and TOT impact estimates for the sample of KIPP middle schools with valid lotteries.

A. How Does KIPP Affect Student Engagement and Effort in School?

The first set of survey outcomes covered student motivation and engagement. The outcomes include an index of student extracurricular activities, student and parent reports of whether the student has homework and how much time is spent daily on homework, and several indices that measure student engagement with school. Impacts on these outcomes could suggest a mechanism through which KIPP middle schools affect student academic achievement. For example, it may be the case that students' academic achievement is driven in part by the amount of academic work they do outside of school. Thus, we present estimates of the impact of KIPP on the amount of homework done by students. Alternatively, several outcomes reflect factors that may influence the extent to which students are motivated and feel it is within their power to do well in school, which may be captured by measures of students' self-control, academic self-concept, school engagement, and effort and persistence in school. The outcomes examined in this section could also be relevant to preparing students for longer-term success.

[48] These outcomes could not be measured for the larger sample used in the matched comparison design because there was no way to administer surveys to the comparison groups.

Key finding: KIPP schools produce positive and statistically significant impacts on the amount of homework done on an average night, as reported by students and parents, but have no impact on seven other measures of student engagement and effort.

Students offered admission to KIPP report that they spend 21 more minutes per night on homework than students who do not receive offers (Table V.1). Parents of lottery winners report 32 minutes more per night than non-winners. The adjusted TOT estimates suggest that students who actually enroll at KIPP spend an additional 35 to 53 minutes of homework per night compared to those not enrolled in KIPP, giving them an average amount of homework of more than two hours per night. KIPP students are in school for substantially more time each day and year, and their homework load further adds to the extra time they spend in formal schooling activities relative to students in non-KIPP schools. As discussed in Chapter III, KIPP schools have an average school day of 8.8 hours and an average school year of 191.5 days. Most non-KIPP schools have shorter school days and years.[49] We find no statistically significant impacts on other measures of student engagement, including student-reported extracurricular activities, school engagement, self control, student effort, and academic self-concept.

B. How Does KIPP Affect Educational Expectations and Aspirations?

The second domain involved student and parent reports of education goals and aspirations. The specific outcomes include whether the student and parent expect the student to graduate from high school on time, whether they aspire to college completion, whether they believe that the student is very likely to complete college, and the frequency of discussions about college. These outcomes are intended to reflect factors that may influence students' motivation to achieve long-term success in school.

Key finding: We found no statistically significant impacts of KIPP on students' educational aspirations.

There were no significant impacts on any of these measures (Table V.2). This may be because very high proportions of students and parents in both groups already have high educational goals and expectations. For example, 60 percent of lottery winners and 61 percent of non-winners believe they are very likely to complete college (Table V.2) and another 40 percent of lottery winners and 38 percent of non-winners believe they are likely to complete college (not shown in table), implying that all lottery winners and 99 percent of non-winners believe they are likely or very likely to complete college. The high levels of educational aspirations shown in this table suggest that the norm for the typical middle school student in our sample (in KIPP schools as well as other middle schools) is to aspire to, and expect to achieve, high levels of educational attainment such that these measures are likely not the best predictors of future academic success.

[49] U.S. public schools average 6.6 hours in a school day and 180 days in a school year, as reported in the 2007-08 Schools and Staffing Survey (Snyder and Dillow 2012).

Table V.1. Impacts on Student Motivation and Engagement

Outcome	Mean, Lottery Winners	Mean, Non-Winners	Impact of Admission Offer (ITT)	Adjusted Impact of Attendance (TOT)
Count of Extracurricular Activities (Mean)	2.95	2.84	0.11 (0.16)	0.18 (0.25)
Homework				
Student reports having homework on a typical night (proportion)	0.96	0.96	0.00 (0.02)	-0.01 (0.03)
Minutes spent on homework on typical night, student report (mean)	117.63	95.70	21.95** (8.5)	35.01** (12.8)
Minutes spent on homework on typical night, parent report (mean)	118.31	86.17	32.14** (4.6)	53.71** (7.0)
Parent says student typically completes homework (proportion)	0.94	0.93	0.01 (0.02)	0.02 (0.03)
Index of School Engagement (Mean)	3.64	3.64	0.00 (0.03)	0.01 (0.05)
Index of Self Control (Mean)	4.43	4.47	-0.04 (0.05)	-0.07 (0.09)
Index of Academic Self-Concept (Mean)	3.25	3.20	0.05 (0.03)	0.08 (0.05)
Index of Effort and Persistence in School (Mean)	3.46	3.51	-0.05 (0.03)	-0.07 (0.05)
Number of Schools with Valid Data	13			
Number of Students with Valid Data				
Student survey	754			
Parent survey	812			

Notes: All impacts in this table are based on regression models that pool all lottery schools and that control for baseline covariates. The means for non-winners are regression adjusted, controlling for the full set of baseline covariates; means for lottery winners are computed by adding the impact estimate to the mean for non-winners. The ITT and TOT models are described in Appendix E. Standard errors are shown in parentheses. Details on the outcome measures are provided in Appendix B.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

Table V.2. Impacts on Education Goals and Aspirations

Outcome	Mean, Lottery Winners	Mean, Non-Winners	Impact of Admission Offer (ITT)	Adjusted Impact of Attendance (TOT)
On-Time High School Graduation (Proportion)				
Student expects to graduate HS on time	0.97	0.96	0.01 (0.01)	0.01 (0.02)
Parent expects student to graduate HS on time	0.97	0.96	0.01 (0.01)	0.02 (0.02)
College Completion (Proportion)				
Student wishes to complete college	0.94	0.97	-0.02 (0.02)	-0.04 (0.03)
Parent wishes student to complete college	0.99	0.99	0.00 (0.01)	0.01 (0.01)
Student believes very likely to complete college	0.60	0.61	-0.01 (0.04)	-0.01 (0.07)
Parent believes student very likely to complete college	0.69	0.68	0.01 (0.04)	0.02 (0.07)
Discussions About College (Proportion)				
Student reports having discussions about college at school	0.79	0.77	0.02 (0.02)	0.04 (0.06)
Student reports having discussions about college at home	0.92	0.92	0.00 (0.01)	0.00 (0.04)
Parent report of having discussions about college	0.96	0.96	0.00 (0.02)	0.00 (0.03)
Number of Schools with Valid Data	13			
Number of Students with Valid Data				
Student survey	746			
Parent survey	838			

Notes: All impacts in this table were based on regression models that pool all lottery schools and control for baseline covariates. The means for non-winners are regression adjusted, controlling for the full set of baseline covariates; means for lottery winners are computed by adding the impact estimate to the mean for non-winners. The ITT and TOT models are described in Appendix E. Standard errors are shown in parentheses. Details on the outcome measures are provided in Appendix B.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

C. How Does KIPP Affect Student Well-Being and Behavior?

The next set of outcomes address student and parent reports of student behavior, both within and outside of school. Several of the outcomes relate to whether the student has been disciplined at school, which may be a measure of either the student's behavior or of discipline policies at the school. Other measures include indices that draw from multiple questions about peer pressure, how frequently students undertake specific good and bad behaviors, and the extent to which parents have concerns about bad behavior.

Key finding: We found two potentially negative impacts of KIPP schools on student behavior. Students offered admission to KIPP report that they are more likely to engage in undesirable behavior and to get into trouble in school. We found no impacts on eight other measures of student behavior.

The offer of admission to KIPP is estimated to lead to an increase in student-reported undesirable behavior by a statistically significant margin (Table V.3). The items included in this scale measure how frequently (often, sometimes or never) students argue with or lie to their parents, give their teachers a hard time, or lose their temper at home or school. Students are also less likely to indicate they never get into trouble at school, a difference that can reflect either differences in actual behavior between the two groups, or differences in the discipline policy or criteria for "getting into trouble" in the schools attended by the two groups. For both of these measures, the estimated impact of KIPP could reflect either true negative impacts on behavior or an effect on the likelihood that students will honestly report, or "own up to," negative behaviors.[50]

We measured behavior in a number of other ways, creating indices based on student and parent responses to questions about peer pressure, disciplinary incidents at school, illegal actions, parent concerns about the student's behavior, and measures of good behavior. We find no evidence that KIPP affected any of these measures of student behavior.

D. How Does KIPP Affect Satisfaction and Perceptions of School?

The final set of outcomes comprises student and parent perceptions of the school. The questions that formed the basis for these outcomes address student feelings about their teachers, other students, and the school disciplinary environment; parent perceptions of the school; and a measure of parental involvement.

Key finding: We found positive and statistically significant impacts on parent and student feelings about the student's school as well as parents' perceptions of the school's academic difficulty.

Table V.4 presents impact estimates for student- and parent-reported measures of satisfaction with the school. We find statistically significant impacts on the index of the student's feelings about school, the index of the parent's satisfaction with school, whether the parent gives an overall rating of excellent to the school, and the parent's perception of the academic difficulty of the school. Parents of lottery winners are significantly less likely to report that their child's school is "too easy" on a variety of dimensions, including homework and class materials.

[50] Evidence from bullying interventions suggests that impacts on student self-reports of behavior do not line up with impacts on parent- or school-reported behaviors for students who participated in school-wide interventions. One explanation is that the intervention increases student awareness of behavior issues and increases their reporting of negative behaviors even if no actual change in behavior occurs (see Smith et al. 2004; Swearer et al. 2010)

Table V.3. Impacts on Student Behavior

Outcome	Mean, Lottery Winners	Mean, Non-Winners	Impact of Admission Offer (ITT)	Adjusted Impact of Attendance (TOT)
Index of Peer Pressure for Bad Behaviors (Mean)	1.04	1.05	-0.01 (0.02)	-0.01 (0.03)
Negative Behaviors (Mean)				
Index of undesirable behavior[a]	1.73	1.63	0.09* (0.04)	0.15* (0.07)
Index of illegal action	1.03	1.02	0.01 (0.01)	0.02 (0.02)
Disciplinary Problems				
Parent reported any school disciplinary problems for student (proportion)	0.33	0.38	-0.04 (0.04)	-0.07 (0.07)
Index of parent-reported frequency of school disciplinary actions for student (mean)[a]	0.20	0.21	-0.01 (0.03)	-0.02 (0.02)
Positive Behaviors				
Student never gets in trouble at school (proportion)	0.41	0.54	-0.13** (0.04)	-0.21** (0.07)
Index of good behavior, student report (mean)[a]	2.32	2.31	0.01 (0.04)	0.01 (0.06)
Index of good behavior, parent report (mean)[a]	2.37	2.42	-0.05 (0.04)	-0.08 (0.07)
Index indicating well-adjusted student (mean)	3.34	3.44	-0.01 (0.04)	-0.02 (0.06)
Index of Parental Concerns About Student (Mean)	1.37	1.34	0.03 (0.05)	0.05 (0.09)
Number of Schools with Valid Data	13			
Number of Students with Valid Data				
Student survey	745			
Parent survey	836			

Notes: All impacts in this table are based on regression models that pool all lottery schools and that control for baseline covariates. The means for non-winners are regression adjusted, controlling for the full set of baseline covariates; means for lottery winners are computed by adding the impact estimate to the mean for non-winners. The ITT and TOT models are described in Appendix E. Standard errors are shown in parentheses. Details on the outcome measures are provided in Appendix B.[a] Index has an alpha smaller than 0.7, indicating low reliability.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

Table V.4. Impacts on School Experiences and Satisfaction

Outcome	Mean, Lottery Winners	Mean, Non-Winners	Impact of Admission Offer (ITT)	Adjusted Impact of Attendance (TOT)
Satisfaction				
Index of student's feelings about school (mean)	3.44	3.35	0.09* (0.03)	0.14* (0.05)
Student likes school a lot (proportion)	0.55	0.58	-0.03 (0.04)	-0.05 (0.07)
Index of parental satisfaction with school (mean)	3.28	3.17	0.11* (0.05)	0.18* (0.08)
Parent rates school as excellent (proportion)	0.55	0.40	0.15** (0.04)	0.24** (0.07)
Index of Student Perceptions of Schoolmates (mean)	2.84	2.79	0.05 (0.04)	0.09 (0.07)
Index of Student Perceptions of Teachers (mean)	3.54	3.49	0.06 (0.04)	0.10 (0.06)
School Discipline				
Index of school disciplinary environment (mean)	3.33	3.34	0.00 (0.04)	-0.01 (0.06)
Index of parental perceptions of problems in student's school (mean)	3.04	3.01	0.03 (0.09)	0.05 (0.14)
Index of Parental Involvement in Student's Education (Mean)[a]	2.79	2.74	0.06 (0.03)	0.10 (0.05)
Academic Difficulty, Parent Report (Mean)				
Index indicating school is too easy[a]	0.11	0.17	-0.06** (0.02)	-0.11** (0.04)
Index indicating school is too difficult[a]	0.08	0.07	0.01 (0.02)	0.02 (0.03)
Number of Schools with Valid Data	13			
Number of Students with Valid Data				
Student survey	754			
Parent survey	848			

Notes: All impacts in this table are based on regression models that pool all lottery schools and that control for baseline covariates. The means for non-winners are regression adjusted, controlling for the full set of baseline covariates; means for lottery winners are computed by adding the impact estimate to the mean for non-winners. The ITT and TOT models are described in Appendix E. Standard errors are shown in parentheses.[a] Index has an alpha smaller than 0.7, indicating low reliability. Details on the outcome measures are provided in Appendix B.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

VI. ANALYSIS OF FACTORS ASSOCIATED WITH KIPP IMPACTS

On average, KIPP middle schools have a positive impact on students' math and reading scores. Although the average impact is positive, individual KIPP middle schools vary in their impacts on student achievement and some schools are more successful than others. In this chapter, we explore the source of that variation—whether specific characteristics of individual KIPP middle schools are associated with their impacts on students.

A wide variety of school-level characteristics (factors) may be associated with the effectiveness of KIPP school relative to neighboring non-KIPP schools. We limited the number of factors we examined because increasing the number of factors raises the chances of spuriously (through random chance alone) finding a significant relationship between them and the estimated impacts of KIPP schools. Our analysis therefore focuses on a small set of factors meeting specific criteria. We limited the analysis to factors with substantial variation in values across KIPP schools in the sample—a factor that does not vary across schools could not possibly explain variation in school impacts. Among factors meeting this criteria, we then investigated those that met at least one of the following two conditions: (1) there is a theoretical or empirical reason to believe the factor might influence school effectiveness (for instance, the factor was found to be important in previous literature) or (2) the factor is within the control of the schools (for example, the amount of time in school or use of Saturday school).

We used two approaches to examine the relationship between the characteristics of KIPP schools and achievement impacts:

1. **Simple bivariate associations between individual factors and impacts.**

2. **Associations between individual factors and impacts while controlling for other factors.** We examined a multivariate model in which the relationships between school impact estimates and several factors potentially explaining these impacts were explored simultaneously. We included variables in the multivariate analysis only if they had a statistically significant relationship with impacts on either year 2 reading or math scores in the bivariate analysis. This allowed us to examine whether the significant bivariate associations persisted once we accounted for other school characteristics. We estimated three alternative versions of this multivariate model that differ only with respect to the factor included in the model to represent time in school.[51]

For both approaches, we relied on matched comparison (rather than lottery-based) impact estimates in order to maximize the number of schools that can be included and the precision of impact estimates for each school.

[51] The primary model included a factor measuring the overall average length of the school day. The alternative versions examined instructional time on core academics during the school day (model 2) and time on non-core activities during the school day (model 3). The core and non-core factors could not be included in the same specification because they are have a high negative correlation with one another (r = -0.85).

There is an important limitation of the bivariate analysis—since many factors are interrelated, we cannot tell from the simple correlations which factors are most likely to drive the results. Thus, we included the multivariate analysis to account for the potential interaction of any related factors. Regardless of what we find, however, our investigation is exploratory in that neither method allows strong causal inferences about what makes some KIPP schools more effective than others in improving student achievement. These analyses are correlational, and there is always the possibility that the causal factors are not included among our measures. Also, we may find some statistically significant relationships by chance. Therefore, the results of our analysis can suggest several hypotheses for further, more rigorous testing but cannot provide conclusive answers to questions about the reasons for particular KIPP schools' effectiveness.

A. Do KIPP Middle School Impacts Vary?

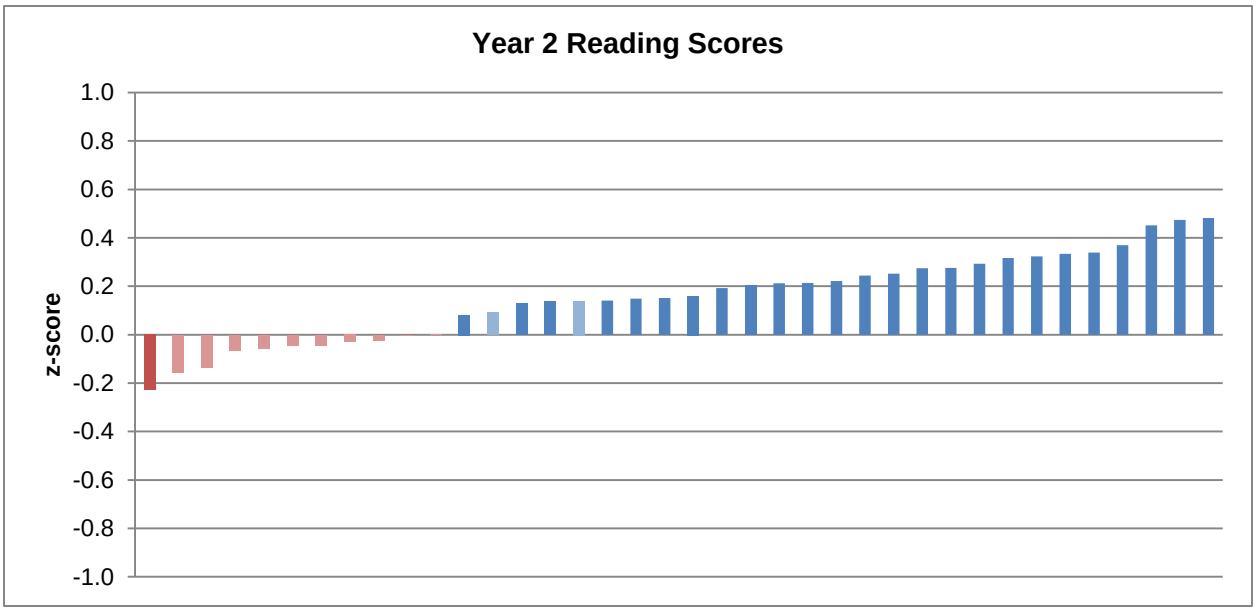
The first step in the analysis of factors related to impacts was to examine whether there was sufficient variation in the estimated impacts of KIPP middle schools to permit a useful analysis of characteristics explaining the variance. If all KIPP schools have similar impacts, there would be no differences to explore.[52] We focused on the estimated year 2 reading and math impacts of KIPP schools, since those reflect cumulative experiences of students in the largest number of schools.[53]

Figures VI.1 and VI.2 show the estimated variation in school-specific impacts on year 2 reading and math scores across the 38 currently-open KIPP schools included in the matching impact analysis. Impacts on year 2 standardized reading scores were estimated to range from -0.23 to 0.48, with a standard deviation of 0.18. One estimated reading impact was statistically significant and negative, and 25 were statistically significant and positive. Impacts on year 2 standardized math scores were estimated to range from -0.26 to 0.73, with a standard deviation of 0.26. One estimated math impact was statistically significant and negative, and 29 were statistically significant and positive. While we would expect some variation in impact estimates across schools due to chance, or random sampling variability, the observed variation is much larger than expected because of chance alone. A statistical test confirms that estimated KIPP matching impacts do vary significantly across schools.

[52] This analysis was based on school-specific impact estimates, adjusted to account for the different levels of precision in these estimates using an empirical Bayes shrinkage adjustment. For a more detailed description of this adjustment, see Appendix D.

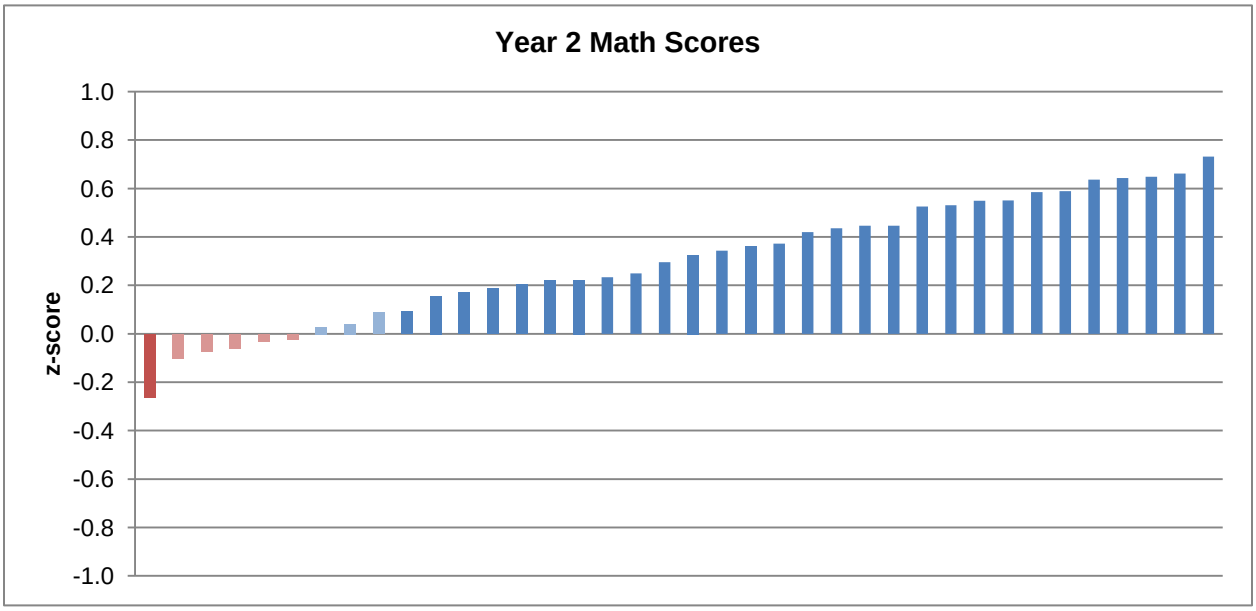
[53] It is likely that the results of this analysis would be similar if we had chosen impacts in a different year to be the focus since school-level impacts are highly correlated across years. The correlation between year 1 and year 2 impacts is .90 in reading and .87 in math, and the correlation between year 2 and year 3 impacts is 0.90 in reading and 0.89 in math.

Figure VI.1. Distribution of School-Level Impact Estimates in Reading



Note: The dark-red and dark-blue bars indicate differences from the district population that are statistically significant at the 0.05 level, two-tailed test.

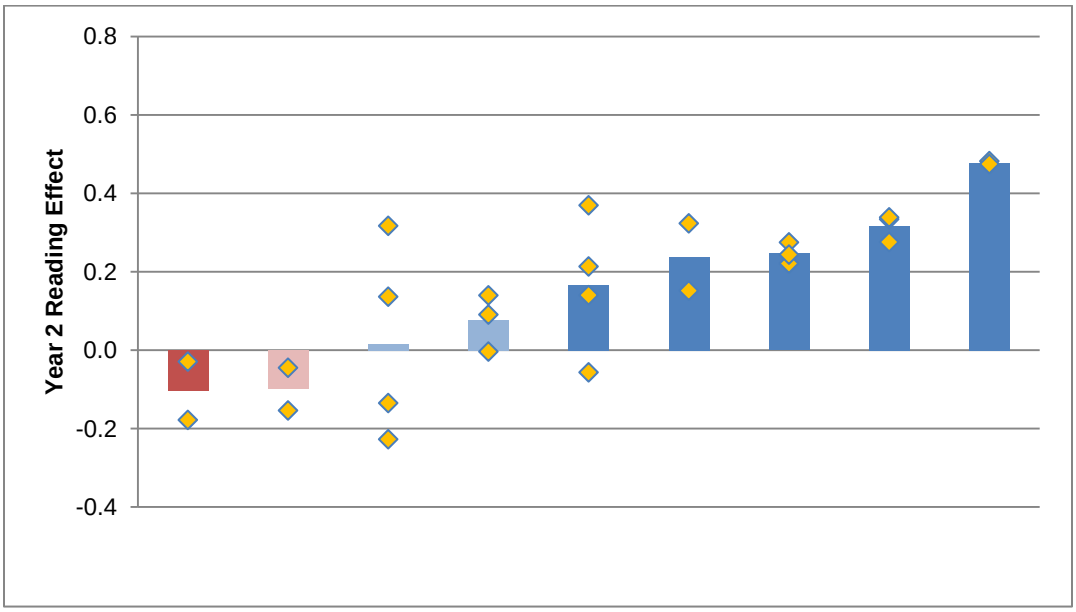
Figure VI.2. Distribution of School-Level Impact Estimates in Math



Note: The dark-red and dark-blue bars indicate differences from the district population that are statistically significant at the 0.05 level, two-tailed test.

Given the increasing role of KIPP regions in establishing local practices, identifying leaders, and providing support to local KIPP schools, we examined whether schools in the same region are producing similar impacts. Figures VI.3 and VI.4 show the variation in school-level reading and math impacts, by region.[54] Within regions, impacts are less varied. The intra-class correlation coefficient is 0.87 for year 2 math and 0.90 for year 2 reading; meaning that variation across regions explains 87 percent of the total variation in year 2 school-level impacts in math and 90 percent of the total variation in reading; the remaining 10 to 13 percent of variation is across schools within regions. The data suggest that differences in the characteristics of regions, common to KIPP schools within those regions, may help to explain much of the variation in the schools' effectiveness. We do not have sufficient sample size to fully explore the role KIPP regions play in this analysis, but these findings suggest a more thorough investigation of how region-specific characteristics influence impacts is needed.

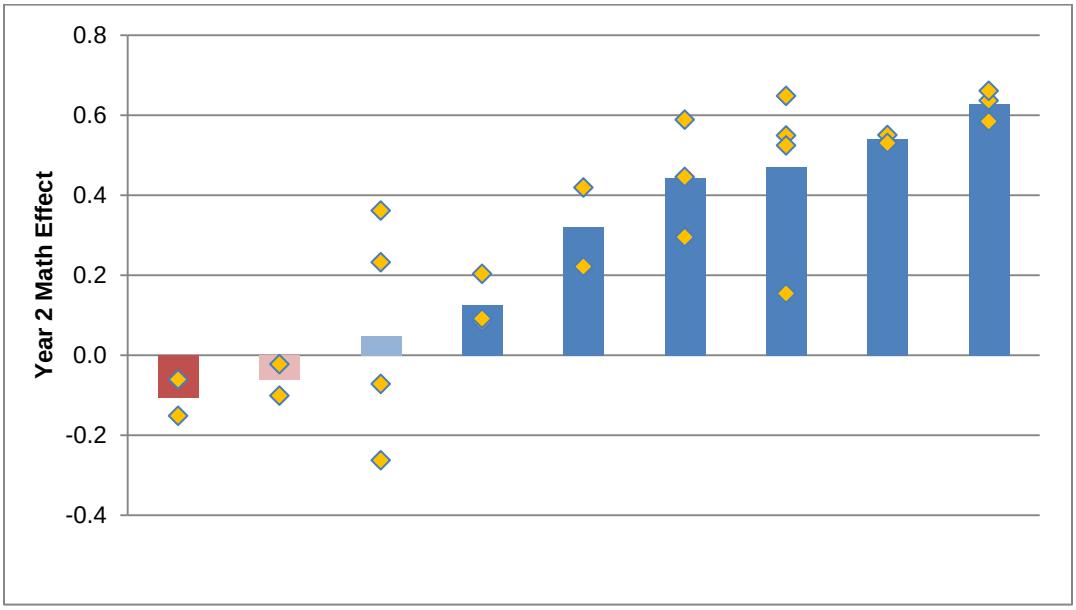
Figure VI.3. Distribution of School-Level Impact Estimates in Reading (by Region)



Note: The dark-red and dark-blue bars indicate differences from the district population that are statistically significant at the 0.05 level, two-tailed test. Each yellow diamond shows the impact estimate for one middle school managed by the KIPP Region whose average impact is represented by the associated vertical bar. Each region displays estimates for at least two schools and no more than four to prevent the identification of individual regions. Diamonds may overlap when estimates are very close to each other.

[54] Only regions where two or more middle schools have estimated matching impacts are included in the analysis.

Figure VI.4. Distribution of School-Level Impact Estimates in Math (by Region)



Note: The dark-red and dark-blue bars indicate differences from the district population that are statistically significant at the 0.05 level, two-tailed test. Each yellow diamond shows the impact estimate for one middle school managed by the KIPP Region whose average impact is represented by the associated vertical bar. Each region displays estimates for at least two schools and no more than four to prevent the identification of individual regions. Diamonds may overlap when estimates are very close to each other.

B. Factors of Interest

Next, we explored the relationship between school-level factors and estimated impacts. As described above, a wide variety of school-level factors may be associated with KIPP schools' effectiveness relative to neighboring non-KIPP schools. We focused our analysis on 14 factors of interest (Table VI.1), which were sorted into four categories of characteristics that might influence the effectiveness of KIPP schools:

- **Student characteristics.** We included these characteristics to address the concern that the impacts of a given school might be driven by the characteristics of the students attending the school rather than the practices of the school itself. Although theoretically KIPP schools could influence the distribution of student characteristics through focused or selective recruitment, these factors are largely outside the schools' control. Nevertheless, prior research has shown that charter schools serving lower-achieving students have higher impacts, for example, so we examined this and other student characteristics (Gleason et al. 2010).
- **Operational characteristics.** We also examined whether the logistics of the operation of the school—such as the amount of time students spend in school—are related to varying impacts.
- **School climate.** These factors are less straightforward to measure, but are designed to capture some of the more nuanced aspects of a school's culture and practices intended to build that culture, such as the systems used to manage student behavior.

Table VI.1. Factors Potentially Influencing Charter School Impacts

School Characteristic/Factor	Unit
Student Characteristics	
Baseline achievement [a]	Z-score (*KIPP/comparison difference*)
Student attrition percentage	Percentage point difference (*KIPP/comparison difference*)
Percentage of students identified for special education	Percentage point difference (*KIPP/comparison difference*)
Operational Factors	
Core class size	Typical class size in reading and math
Use of Saturday school	Number of Saturday school days in a typical month
Time in school	
Length of the school day	Hours per day
Instructional time in core subjects [b]	Hours per day
Time in non-core subjects	Hours per day
School Climate	
School-wide behavior plan index	Z-score
Index of principal time on problematic issues	Z-score
Parent involvement index	Z-score
Principal satisfaction index	Z-score
Staff Factors	
Teacher PD and mentoring indices	
For new teachers[c]	Z-score
For experienced teachers[c]	Z-score
Teacher turnover	Proportion of teachers who left the school during or following the 2010-2011 school year
Principal experience	Years of experience as principal (in any school)
Teacher experience	Proportion of teachers with more than four years of experience

[a] Baseline achievement is standardized within each jurisdiction, with a mean of 0 and a standard deviation of 1.
[b] Core subjects include math, English/language arts, science, and history
[c] Index has an alpha smaller than 0.7, indicating low reliability

- **Staff characteristics.** Finally, we examined the characteristics of the school staff, such as experience or training. Staff at KIPP schools have considerable autonomy to set the direction of the school and potentially influence its success.

The majority of these factors are defined by characteristics of the KIPP schools alone, but those within the student characteristics domain reflect the differences between KIPP schools and their district-wide comparison schools (noted as "KIPP/comparison difference" in the table). Some factors are measured as indices and reported as z-scores; these are constructed from responses to

survey items and scaled across the full sample of KIPP schools to have a mean of 0 and a standard deviation of 1 (more detail on how these indices are constructed is provided in Appendix B).

C. What School-Level Factors Are Related to Impacts?

In Table VI.2, we summarize the results of the bivariate analyses, examining the relationships between the characteristics of KIPP middle schools and matching impacts on year 2 reading and math achievement. In Table VI.3, we summarize the multivariate analysis, controlling for other factors that were significantly related to impacts in the bivariate analysis. In this analysis, we assess statistical significance at the 0.10 level, rather than the 0.05 level that we use elsewhere.[55] Overall, the selected factors explain relatively little of the variation in the estimated effectiveness of the KIPP middle schools. While a few factors show statistically significant bivariate correlations, few differences remain significant in multivariate models.

Key finding: We found limited evidence that KIPP schools with higher percentages of students identified for special education relative to their district counterparts have higher impacts in reading, but no evidence that other student characteristics are associated with impacts.

We found limited evidence that student characteristics are associated with estimated impacts. In the bivariate analysis, the proportion of students identified for special education is positively correlated with year 2 impacts in reading only. KIPP charter schools with a higher proportion of students identified for special education relative to the district had significantly more positive impacts on year 2 reading. A 1 percentage point increase in the proportion of students identified for special education relative to the district average is associated with a 0.02 standard deviation increase in the estimated impact of KIPP on reading achievement (p-value = 0.097). This finding counters a common criticism that KIPP achieves results by "creaming" higher-achieving and better behaved students and serving fewer students identified for special education. However, this relationship is no longer significant in any multivariate models controlling for other student, operational, climate, and staff factors. This might indicate that the relationship between the proportion of students identified for special education and KIPP's estimated impact in reading is driven by another factor in the model.

[55] We used a higher critical value for determining statistical significance in this analysis for two reasons. First, because the sample size of estimated school-level impacts is limited (with 38 school-level observations), the size of the true relationship between factors and impacts would have to be very large for the analysis to detect it as significant at the 0.05 level, and not quite as large to be able to detect it as significant at the 0.10 level. Second, because the analysis is exploratory, we were less concerned about concluding that a relationship is present when none exists in reality (type I error) than concluding that there is no relationship present when one exists in reality (type II error).

Table VI.2. Bivariate Relationships Between School Characteristics and KIPP School Impacts

School Characteristic/Factor	Year 2 Reading Score	Year 2 Math Score
Baseline Achievement	-0.16 (0.12)	-0.12 (0.15)
Student Attrition	0.00 (0.00)	0.00 (0.00)
Proportion of Students Identified for Special Education	0.02* (0.01)	0.01 (0.01)
Class Size	-0.01 (0.01)	0.00 (0.02)
Use of Saturday School	0.02 (0.04)	0.07 (0.05)
Time in School		
Length of school day	-0.13** (0.06)	-0.19** (0.08)
Instructional time in core subjects	0.06 (0.05)	0.12* (0.06)
Time in non-core subjects	-0.08** (0.03)	-0.13*** (0.04)
School-Wide Behavior Plan Index	0.06** (0.03)	0.08* (0.04)
Parent Involvement	0.02 (0.04)	0.04 (0.05)
Principal Satisfaction	0.03 (0.03)	0.04 (0.04)
Index of Principal Time on Problematic Issues	-0.07*** (0.02)	-0.07 (0.05)
Teacher PD and Mentoring		
For new teachers	-0.03 (0.02)	-0.04 (0.04)
For experienced teachers	0.00 (0.03)	-0.02 (0.04)
Teacher Turnover	-0.06 (0.23)	-0.14 (0.37)
Principal Experience	0.02** (0.01)	0.02* (0.01)
Teacher Experience	-0.05 (0.11)	0.05 (0.17)

Note: The estimates presented in the table are coefficient estimates from a regression of the estimated impact on a single variable—the school characteristic/factor shown in the row. Each row represents a separate regression estimate. Estimates are presented as z-scores. Robust standard errors are shown in the parentheses.

* Coefficient is statistically different from 0 at the 0.10 level, two-tailed test.
** Coefficient is statistically different from 0 at the 0.05 level, two-tailed test.
*** Coefficient is statistically different from 0 at the 0.01 level, two-tailed test.

Table VI.3. Multivariate Relationships Between School Characteristics and KIPP School Impacts

	Year 2 Reading Score			Year 2 Math Score		
School Characteristic/ Factor	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Proportion of Students Identified for Special Education	0.01 (0.01)	0.02 (0.01)	0.02 (0.01)	0.01 (0.02)	0.02 (0.02)	0.02 (0.02)
Time in School						
Length of school day	-0.11** (0.05)	n.a.	n.a.	-0.18*** (0.05)	n.a.	n.a.
Instructional time in core subjects	n.a.	0.08* (0.05)	n.a.	n.a.	0.15** (0.06)	n.a.
Time in non-core subjects	n.a.	n.a.	-0.09** (0.03)	n.a.	n.a.	-0.15*** (0.04)
Index of School-Wide Behavior Plan	0.04 (0.02)	0.05 (0.03)	0.05* (0.03)	0.07 (0.05)	0.10* (0.05)	0.10* (0.05)
Index of Principal Time on Problematic Issues	-0.04* (0.02)	-0.03 (0.02)	-0.02 (0.02)	-0.02 (0.04)	-0.01 (0.05)	0.01 (0.05)
Principal Experience	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Adjusted R^2	0.22	0.25	0.31	0.08	0.16	0.22
Sample Size Used in the Regression	36	35	34	36	35	34

Notes: Robust standard errors are shown in the parentheses. The multivariate regressions include only those school characteristics for which there is a statistically significant bivariate association with impacts on either year 2 reading scores or year 2 math scores. Model 1 is the primary multivariate model; additional specifications are included to explore the negative association between the length of the school day and estimated impacts in reading and math.

* Coefficient is statistically different from 0 at the 0.10 level, two-tailed test.
** Coefficient is statistically different from 0 at the 0.05 level, two-tailed test.
*** Coefficient is statistically different from 0 at the 0.01 level, two-tailed test.

n.a. = not applicable.

Key finding: We found some evidence that schools with particularly extended days have smaller impacts in reading and math, but the relationship seems to be driven by the amount of time students spend outside core academic subjects. We found no evidence that other operational characteristics are significantly related to impacts.

All KIPP schools have longer-than-normal school days (with an average KIPP school day of more than nine hours), so we cannot compare the effect of KIPP's extended school day to a school day of normal length in this analysis. Some KIPP schools have longer days than others, however, allowing us to examine whether KIPP schools with an especially long school day have larger or smaller impacts than KIPP schools whose school day is extended by a more modest amount. We found some evidence that, given an already-extended school day, the overall length of the school day is *negatively* related to impacts in reading and math. In the bivariate analysis, a one hour increase in the length of the school day is associated with a 0.13 standard deviation decrease in the impact of KIPP on reading achievement (p-value = 0.036) and a 0.19 standard deviation decrease in KIPP's impact on math achievement (p-value = 0.020). This negative relationship persists in the primary multivariate model where we control for other factors.

To further explore this relationship, the measure of the length of the school day was broken into two factors: instructional time in the core subjects of math, English or language arts, science, and history (measured in hours) and time in non-core subjects or activities (the remainder of the length of the school day). In this analysis, instructional time in core subjects is positively and significantly related to impacts in math and time in non-core subjects is negatively and significantly related to impacts in both math and reading. An hour increase in time in core subjects per day is associated with a 0.12 standard deviation increase in the impact of KIPP on math achievement (p-value = 0.062). In contrast, a one hour increase in time in non-core subjects is associated with a 0.13 standard deviation decrease in KIPP's impact on math achievement (p-value = 0.006) and a 0.08 standard deviation decrease in reading achievement (p-value = 0.017). This suggests that the negative relationship between the length of the school day and estimated impacts in reading and math is being driven by time in non-core instruction, which is positively correlated with the overall length of the school day. In contrast, the amount of time in core academic areas is not strongly correlated with the overall length of the school day. In other words, in schools spending more time in core areas, the total school day isn't necessarily longer.

We then ran two additional models replacing length of the school day with instructional time in core subjects (Model 2) or time in non-core subjects (Model 3). In the multivariate analysis, the positive relationships between instructional time in core subjects and impacts and the negative relationships between time in non-core subjects and impacts persisted, though they became slightly smaller in magnitude. This provides more evidence that the negative relationships between the length of the school day and impacts are driven by time in non-core subjects. However, these relationships are not necessarily causal. As with all of the factor analysis, unobserved aspects of KIPP schools or the counterfactual in comparison schools may be driving the results. In particular, the relationships between the length of the school day and impacts disappear when we account for regional fixed effects in an exploratory analysis. Thus, it is possible this factor is actually serving as a proxy for other, unmeasured regional differences. No other operational characteristic we examined was significantly associated with impacts in reading or math.

Key finding: We found some evidence that schools that emphasize school-wide approaches to managing behavior have more positive impacts in reading and math. No other school climate factors were significantly related to impacts.

In the bivariate analysis, a more comprehensive school-wide behavior system is positively correlated with impacts in reading and math. School-wide behavior is an index measuring the degree to which principals agree that: (1) behavior standards and discipline policies are established and enforced consistently across the entire school; (2) the school has a behavior code that includes positive rewards for students who consistently behave well; and (3) the school has a school-wide behavior code that includes negative sanctions for students who violate rules. A 1 standard deviation increase in a school's score on the index is associated with a 0.06 standard deviation increase in KIPP's estimated impact in reading and a 0.08 standard deviation increase in KIPP's estimated impact in math. As a result, we included this factor in a multivariate analysis, simultaneously controlling for principal time spent on problematic issues, proportion of students eligible for special education, time spent in core and non-core subjects, and principal experience. The associations between more positive scores on the school-wide behavior index and more positive estimated impacts in reading and math are slightly smaller and do not quite reach significance in the main multivariate model. However, in alternate specifications of the multivariate model, these

relationships tend to remain significant; in other words, the relationship persists even when controlling for other factors in the model.

Principal time on problematic issues is negatively related to impacts on reading in the bivariate analysis. The index was constructed based on the frequency with which principals spend an hour or more monthly resolving three problem issues: (1) complaints from parents, (2) conflicts among teachers; and (3) individual teacher complaints. A one standard deviation increase in the amount of time a principal spends dealing with problematic issues is associated with a .07 standard deviation decrease in the estimated impact of KIPP on reading achievement. This relationship decreases in magnitude, but remains marginally significant when we control for other factors in multivariate model 1. However, in alternate versions of that model, this relationship is no longer significant. Finally, we found no evidence that parent involvement or principal satisfaction are significantly related to impacts in reading or math.

Key finding: We found limited evidence that schools with more experienced principals have more positive impacts in reading and math. No other staff characteristics are significantly related to impacts.

Finally, principal experience is positively correlated with impacts in reading and math in the bivariate analysis. In this analysis, a one year increase in experience is associated with a 0.02 standard deviation increase in KIPP's impact on math achievement (p-value = 0.085) and a 0.02 standard deviation increase in reading achievement (p-value = 0.017). However, as with most relationships examined here, once we control for other factors, the magnitude of the correlation virtually disappears and is no longer significant. Further, we found no evidence that other staff characteristics, including professional development and mentoring for new and experienced teachers, teacher turnover, or teacher experience are significantly related to student achievement.

Key finding: The measured factors explain little of the variation in the estimated effectiveness of KIPP middle schools.

Few factors show statistically significant bivariate correlations with the estimated impacts of KIPP schools, and only length of the school day remains significant when we control for other characteristics simultaneously in our primary multivariate model. The comprehensive school-wide behavior factor remains positively associated with impacts in only some model specifications controlling for other factors, meaning that much of the variation in estimated impacts is not explained by the factors we examined in this chapter. There are four possible explanations for this result. First, given the small number of schools in our analysis, we may not have sufficient power to detect relationships between factors and estimated impacts. Second, the differences we observed in impacts might result from some other factor or set of factors not included in our data. For example, given the limited variation in impacts across schools within regions, it could be that region-specific characteristics or management practices explain much of the variation in estimated impacts. Richer and more extensive measures of school operations might help to identify impactful factors. Third, it could be that there is no single feature that explains the positive impacts of the most successful KIPP schools on its own. Rather, higher KIPP impacts may be driven by a *combination* of features, including some not captured in our data. Finally, it could be that characteristics of the schools attended by the comparison group, which we are unable to measure for most variables at this time, predominantly drive variation in impacts of KIPP schools. Future research, with more complete

data on the characteristics and climate of the comparison schools, could examine how differences in impacts relate to differences in the quality of comparison schools.

VII. CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

KIPP is a rapidly expanding network of public charter schools whose mission is to improve the education of low-income children. This report summarizes the results of an evaluation of KIPP using experimental and quasi-experimental methods to produce rigorous and comprehensive evidence on the effects of KIPP middle schools across the country. We estimated the effects of KIPP on student achievement based on state assessments for more than two-thirds of all KIPP middle schools. We also examined the effects of KIPP on student outcomes beyond state test scores, including student performance on a nationally norm-referenced test that includes measures of higher-order thinking skills, and survey-based measures of student attitudes and behavior.

Data on student characteristics provide little evidence that KIPP "creams" or selectively enrolls higher-performing students. Students entering KIPP are less likely to have received special education services or be classified as having limited English proficiency. On most identifiable characteristics, the students entering KIPP schools look much like those in their neighborhoods: low-achieving, low-income, and non-white. The typical KIPP student scored at the 45th percentile within the district in reading and math prior to entering KIPP, an achievement level that is also significantly lower than the average in their own elementary schools.

We also examined the rate at which students leave their middle school prior to completing the 8th grade. On average, students leave KIPP schools prior to middle school completion at about the same rate as students at other middle schools in the same districts. Many KIPP students take longer to get to high school, however—KIPP students are more likely than those at local district schools to repeat a grade.

Our impact estimates suggest four key results related to how KIPP affects student achievement:

Key finding 1: KIPP middle schools have positive and statistically significant impacts on student achievement across all four academic subjects examined, in each of the first four years after enrollment in a KIPP school, and for all student subgroups that were examined.

Key finding 2: The magnitude of KIPP's achievement impacts is substantial.

Key finding 3: The matched comparison group design produced estimates of KIPP's achievement impacts similar to those of the same impacts based on an experimental, lottery-based design.

Key finding 4: In the lottery sample, average KIPP impacts on a low-stakes test that included items assessing higher-order thinking skills are similar to impacts on high-stakes state tests.

In addition to potentially affecting student academic achievement, KIPP may also influence student behaviors and attitudes related to long-term academic success. For KIPP schools in the lottery sample, we use the experimental design to estimate impacts on various measures of student behavior and attitudes. Notable findings from this analysis include:

- Students enrolled at KIPP spend an additional 35 to 53 minutes on homework per night than they would in a non-KIPP school, resulting in an average of more than two hours per night of homework.
- KIPP has no statistically significant effect on a variety of measures of student attitudes that may be related to long-run academic success. For example, the estimated KIPP impacts on indices of student-reported self-control, academic self-concept, school engagement, effort/persistence in school, and educational aspirations are not statistically significant.
- KIPP has no statistically significant effect on several measures of student behavior, including self-reported illegal activities, an index of good behavior, and parent reports of behavior problems. However, KIPP has a negative estimated effect on a student-reported measure of undesirable behavior, with KIPP students more likely to report behaviors such as losing their temper, arguing or lying to their parents, or having conflicts with their teachers.
- Winning an admissions lottery to KIPP has a positive effect on students' and parents' satisfaction with school. In addition, the parents of KIPP students are less likely to report that their child's school is too easy.

The factors that drive the success of KIPP schools could not easily be determined in our analysis. Few of the school characteristics we examined are strongly correlated with the estimated impacts of the KIPP schools in the study sample. For example, class size, teacher experience, and professional development opportunities are not associated with impacts. The lack of significant correlations between these school characteristics and impacts may be explained, in part, by the limited sample size of schools for which impact estimates and school characteristics were available, affecting our ability to detect small to moderately-sized relationships.

Nonetheless, we identified two factors that are related to the strength of KIPP's impacts on student achievement. First, the size of KIPP's positive impact on student achievement is larger in schools where principals report a more comprehensive school-wide behavior system. Under these systems, schools have clearly defined and consistently enforced rewards for good behavior and consequences for negative behavior. Second, the length of the school day and how time is used are also significantly associated with impacts. All KIPP schools have longer-than-normal school days, but some have longer days than others. Overall, average impacts on student achievement are smaller in KIPP schools with a particularly extended school day. This counterintuitive relationship appears to be driven by the fact that, in these schools, the additional time tends to be spent in non-core academic activities. In contrast, average impacts are larger in KIPP schools in which relatively more time is spent on core academic activities.

Findings from this study raise three important questions related to KIPP schools and their effects on students that are worthy of further research. The first question concerns the factors driving KIPP's substantial positive achievement impacts—which reflect the difference in achievement among students at KIPP schools compared to those among similar students at non-KIPP schools. It would be useful to know what KIPP schools do differently from nearby non-KIPP schools that might drive these impacts. One way we tried to better understand possible mechanisms driving achievement effects was to estimate the KIPP impacts on other student outcomes that might in turn lead to improved achievement. For example, we found positive impacts on the amount of

time spent on homework, but few significant effects on measures of student behavior and attitudes. Another analysis measured the correlation between various characteristics of KIPP schools and their estimated impacts on student achievement. As described above, we found few characteristics strongly related to impacts. However, these analyses were limited in some respects. For example, our analysis of the correlations between KIPP school characteristics and impacts was based on a relatively small sample of schools. Thus, additional research aimed at learning more about the factors potentially driving positive KIPP achievement impacts, taking advantage of larger samples or better data on characteristics of both KIPP and non-KIPP schools, would be fruitful.

A second set of questions about KIPP that should be the focus of future research involves understanding whether the positive achievement effects can be maintained as the KIPP network grows. As additional schools join the network within each region, including more elementary schools and high schools, maintaining the key features of KIPP that lead to its positive impacts may become more challenging. The opening of new KIPP schools within a community, particularly at different levels, could result in a changing composition of students attending KIPP. In addition, KIPP may face challenges in recruiting enough qualified and effective teachers and principals as the total number of KIPP schools in a given geographic area increases. To shed light on whether the KIPP model is scalable, it will be useful to investigate whether the new KIPP schools have similar effects on student achievement as the ones in this study. This report begins to address this issue by capturing the earlier stages of KIPP expansion, although the question of whether impacts will remain positive as the KIPP network gets even larger remains open.

A third set of interesting future research questions surrounds the longer-term effects of KIPP. While student achievement is an important outcome to study, longer-term outcomes such as completing high school, entering and succeeding in college, and doing well in the labor market are more central to the KIPP mission. Research on the effects on these longer-term outcomes will be critically important for gaining a fuller picture of the KIPP network and its promise as model for serving disadvantaged students.

In future work evaluating the KIPP network's effort to "scale up," we will address each of these questions to some extent. We will calculate impacts for additional KIPP schools and generate separate impacts by school year, giving us a larger sample for analyzing factors that can be correlated to KIPP impacts and the opportunity to observe how the impacts of individual KIPP schools change over time. In addition, this work will enable us to estimate the effectiveness of newer KIPP schools, including elementary and high schools. Finally, as the network matures, researchers will be able to calculate longer-term impacts on students, assessing KIPP's progress towards its goals of seeing more students to and through college.

REFERENCES

Abdulkadiroglu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane and Parag A. Pathak. "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots." *Quarterly Journal of Economics*, vol. 126, no. 2, 2011, pp. 699–748.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. "Explaining Charter School Effectiveness." National Bureau of Economic Research Working Paper #17332. Cambridge, MA: August 2011.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, Christopher R. Walters, "Who Benefits from KIPP?" *Journal of Policy Analysis and Management*, vol. 31, no.4, 2012, pp. 837–860.

Battle, Danielle. "Characteristics of Public, Private, and Bureau of Indian Education Elementary and Secondary School Principals in the United States: Results from the 2007–08 Schools and Staffing Survey." Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, June 2009.

Bifulco, Robert. "Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison." *Journal of Policy Analysis and Management*, vol. 31, no. 3, 2012, pp. 729–751.

Bloom, Howard, Carolyn Hill, Alison Rebeck Black, and Mark Lipsey. "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions." October 2008. Available at [http://www.mdrc.org/sites/default/files/full_473.pdf]. Accessed February 18, 2013.

Cook, T., W. Shadish, and V. Wong. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management*, vol. 27, no. 4, 2008, pp. 724–750.

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

Dobbie, Will, and Roland G. Fryer, Jr. "Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics*, vol. 3, no. 3, 2011, pp. 158–187.

Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Ernst, and Philip Gleason. "Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, April 2012.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts." Cambridge, MA: Mathematica Policy Research, January 2012.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. "The Evaluation of Charter School Impacts." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2010.

Henig, Jeffrey R. "What Do We Know About the Outcomes at KIPP Schools?" East Lansing, MI: Great Lakes Center for Education Research and Practice, November 2008.

Hoxby, Caroline M., Sonali Murarka, and Jenny Kang. "How New York City's Charter Schools Affect Student Achievement: August 2009 Report." Second report in series. Cambridge, MA: New York City Charter Schools Evaluation Project, September 2009.

Keigher, Ashley. "Teacher Attrition and Mobility: Results from the 2008–09 Teacher Follow-Up Survey—First Look." Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, August 2010.

Morris, Carl N. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of American Statistical Association*, vol. 78, no. 381, 1983, pp. 47–55.

Nichols-Barrer, Ira, Christina Clark Tuttle, Brian P. Gill, and Philip Gleason. "Student Selection, Attrition, and Replacement in KIPP Middle Schools." Mathematica working paper series. Princeton, NJ: Mathematica Policy Research, September 2012.

Smith, J.D., B.H. Schneider, P.K. Smith, and K. Ananiadou. "The Effectiveness of Whole-School Antibullying Programs: A Synthesis of Evaluation Research." *School Psychology Review*, vol. 33, no. 4, 2004, pp. 547–560.

Snyder, Thomas D., and Sally A. Dillow. "Digest of Education Statistics 2011." Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, June 2012.

Swearer, Susan M., Dorothy L. Espelage, Tracy Vaillancourt, and Shelley Hymel. "What Can Be Done About School Bullying? Linking Research to Educational Practice." *Educational Researcher,* vol. 39, no.1, January 2010, pp. 38–47.

Tuttle, Christina Clark, Bing-ru Teh, Ira Nichols-Barrer, Brian P. Gill, and Philip Gleason. "Student Characteristics and Achievement in 22 KIPP Middle Schools." Washington, DC: Mathematica Policy Research, June 2010.

U.S. Department of Education. *What Works Clearinghouse Procedures and Standards Handbook, Version 2*. Washington, DC: U.S. Department of Education, December 2008.

Woodworth, Katrina R., Jane L. David, Roneeta Guha, Haiwen Wang, and Alejandra Lopez-Torkos. "San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement." Final Report. Menlo Park, CA: SRI International, 2008.

APPENDIX A

SAMPLE SELECTION AND BASELINE CHARACTERISTICS

In this appendix, we provide detail on how we defined school and student samples for both the propensity-score matching and lottery-based analyses. We also present evidence of baseline equivalence between the resulting treatment and comparison groups.

A. Cohorts Included in Matched Comparison Group Analysis Sample

Listed in Table A.1 below are the 46 KIPP schools from which we collected at least one year of student-level records and test score data. For each school, we show the number of 5th grade (entry) cohorts that have ever been served by the school through the 2010–11 school year, along with the number (and years) of cohorts included in our data. Overall, for these schools, we include 85 percent of all eligible cohorts in our analytic sample (252 of 296 possible cohorts at the 46 schools); this coverage increases to 96 percent if we exclude the cohorts of students served by the first two KIPP schools prior to the establishment of the network in 2001.

B. Propensity Score Matching Procedures

As described in Chapter III, our preferred quasi-experimental approach is a combination of identifying a matched comparison group of students who are similar to KIPP students and then applying an ordinary least squares (OLS) regression model to control for remaining differences. The matching process was performed separately for each of the 41 KIPP middle schools in the sample. This matching process, described in detail below, consisted of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of KIPP students in the sample; and (3) testing the balance of baseline characteristics between our KIPP sample and matched comparison group.

For the first step, we separated the students in each district-level data set into cohorts—grade-by-year groups for each typical KIPP entry grade (5th and 6th) in each year observed in the data. For each cohort of students at a given KIPP school, the pool of eligible comparison students was limited to those in the same district and grade as the KIPP students the year before they first enrolled in a KIPP middle school; comparison students were restricted to those never enrolled in KIPP at any time during middle school. We then performed an iterative propensity score estimation procedure on a combined data set of all cohorts. The dependent variable in this propensity score model is an indicator of whether the student enrolled in a KIPP school in either grade 5 or grade 6.[56] Covariates in the model were selected using an iterative process that identifies the baseline demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. (See Table A.2 for an exhaustive list of potential covariates for inclusion in each model.) At a minimum, we required the logistic model to include one year of baseline test scores in both math and reading. The other covariates were iteratively included and

[56] We did not distinguish between students who enrolled for part of middle school or for the entire duration of middle school. We also did not distinguish between students who enrolled in a single KIPP school and those who enrolled in multiple KIPP schools; before matching, all KIPP students in our data were grouped by the first recorded KIPP school they attended in our data.

Table A.1. Schools and Cohorts Included in QED Sample

State	KIPP School (Year Opened)	Total Number of KIPP Cohorts Through 2010-11	Number of KIPP Cohorts in Data (School Years)
AR	Delta College Prep (2002)	9	7 (2004-05 to 2010-11)
	* Blytheville College Prep (2010)	1	1 (2010-11)
CA	Bridge Academy (2002)	9	7 (2003-04 to 2009-10)
	Academy of Opportunity (2003)	8	6 (2004-05 to 2009-10)
	Adelante Prep (2003)	8	6 (2004-05 to 2009-10)
	Bayview Academy (2003)	8	8 (2003-04 to 2010-11)
	LA College Prep (2003)	8	4 (2006-07 to 2009-10)
	SF Bay Academy (2003)	8	8 (2003-04 to 2010-11)
CO	Sunshine Peak Academy (2002)	9	9 (2002-03 to 2010-11)
DC	DC KEY Academy (2001)	10	10 (2001-02 to 2010-11)
	DC AIM Academy (2005)	6	6 (2005-06 to 2010-11)
	DC WILL Academy (2006)	5	5 (2006-07 to 2010-11)
GA	Achieve Academy (2003)—*closed*	3	2 (2004-05 to 2005-06)
	WAYS Academy (2003)	8	7 (2004-05 to 2010-11)
	Strive Academy (2009)	2	2 (2009-10 to 2010-11)
	* Vision Academy (2010)	1	1 (2010-11)
IN	Indianapolis College Prep (2004)	7	7 (2004-05 to 2010-11)
	LEAD College Prep (2006)—*closed*	5	5 (2006-07 to 2010-11)
LA	Believe College Prep (2006)	5	4 (2006-07 to 2009-10)
	Central City Academy (2007)	4	3 (2007-08 to 2009-10)
MA	Academy Lynn (2004)	7	7 (2004-05 to 2010-11)
NC	Gaston College Prep (2001)	10	9 (2001-02 to 2009-10)
	Asheville Youth Academy (2002)—*closed*	4	4 (2002-03 to 2005-06)
	* Academy Charlotte (2007)	4	1 (2007-08)
NY	Academy NY (1995)	16	9 (2002-03 to 2010-11)
	STAR College Preparatory (2003)	8	8 (2003-04 to 2010-11)
	AMP Academy (2005)	6	6 (2005-06 to 2010-11)
	Infinity Charter (2005)	6	6 (2005-06 to 2010-11)
OK	Reach College Preparatory (2002)	9	9 (2002-03 to 2010-11)
	Tulsa College Prep (2005)	6	4 (2007-08 to 2010-11)
PA	Philadelphia Charter (2003)	8	6 (2005-06 to 2010-11)
	West Philadelphia Prep (2009)	2	2 (2009-10 to 2010-11)
TN	Memphis Collegiate Middle (2002)	9	9 (2002-03 to 2010-11)
TX	Academy Middle School (1995)	16	8 (2002-03 to 2009-10)
	3D Academy (2001)	10	9 (2002-03 to 2010-11)
	Austin College Preparatory (2002)	9	8 (2002-03 to 2009-10)
	Aspire Academy (2003)	8	7 (2003-04 to 2009-10)
	TRUTH Academy (2003)	8	7 (2003-04 to 2009-10)
	Liberation College Prep (2006)	5	5 (2006-07 to 2010-11)
	Spirit College Prep (2006)	5	5 (2006-07 to 2010-11)
	Polaris Academy for Boys (2007)	4	4 (2007-08 to 2010-11)
	Sharpstown College Prep (2007)	4	3 (2007-08 to 2009-10)
	Intrepid Prep (2008)	3	3 (2008-09 to 2010-11)
	* Arts & Letters (2009)	2	1 (2009-10)
	Voyage Academy for Girls (2009)	2	2 (2009-10 to 2010-11)
	* Camino Academy (2010)	1	1 (2010-11)

Note: Schools with only one cohort of available data (*) are included in the regression-based estimates only. Data was provided either by states or individual school districts. In each school year, all data files included the following student-level variables: school of enrollment; indicators for gender, race/ethnicity, and special education status; and test scores in reading and math. In CA, grade 8 math scores could not be analyzed because students do not all take the same test. Middle school test scores in science were provided in CA, CO, GA, IN, LA, MA, OK, NC, TN, and TX. Middle school test scores in history were provided in CA, GA, IN, LA, OK, TN, and TX. Most files also included students' free or reduced price lunch status (except in NY, PA, one district in OK, and two districts in CA) and limited English proficiency status (except in TX and one district in OK).

Table A.2. List of Potential Covariates for Inclusion in Propensity Score Model

Observed and imputed (when missing) math and reading baseline test scores from one year prior (always included)
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from one year prior
Observed and imputed (when missing) math and reading baseline test scores from two years prior
Observed (non-imputed) math and reading baseline test scores from two years prior
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from one or two years prior are imputed (see Appendix E)
Dummy variables indicating whether student repeated a grade one or two years prior
Demographic variables (gender, race/ethnicity, special education status, free or reduced price lunch status, and limited English proficiency status, where available)
Interactions of baseline test scores from one year prior and all available demographic variables
Interactions of gender and race/ethnicity variables
Interactions of special education status and race/ethnicity variables
Interactions of free and reduced price lunch status and race/ethnicity variables
Interactions of limited English proficiency status and race/ethnicity variables

tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

Next, we calculated propensity scores for KIPP entry. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (without replacement) of comparison group students to treatment group students, separately by cohort. In other words, for each KIPP student, we identified the non-KIPP district student whose propensity score was closest to that of the KIPP student. We then tested the balance of the KIPP group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, special education status, free and reduced price lunch status, and limited English proficiency status). For the matched comparison group sample associated with each KIPP school, we required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no more than one significant difference on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this covariate for the comparison group are not significantly different from the treatment group at the five percent level.[57] If the first round of

[57] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table A.3, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups.

matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that KIPP school, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[58] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

We also tested whether the study's impact estimates were sensitive to the procedure we used to match comparison group students to KIPP students. We found that the impact estimates were not dependent on whether matching was conducted with or without replacement, and the results also remained consisted with a caliper-based matching approach. For more details on these sensitivity tests, see Appendix D.

C. Baseline Equivalence: Matched Sample

As described above, we selected a single matched comparison group for each KIPP school in the analysis. However, the analysis sample we used to estimate impacts on the key test score outcomes varied from the original sample of KIPP and comparison group students to the extent that the test score outcomes may have been missing for individual sample members. In the following six tables, we provide evidence on the extent to which the baseline characteristics of KIPP and matched comparison group students were balanced for the analysis samples used in estimating impacts for each key test score outcome. For math and reading, the samples include all students with at least one math or reading score in each year after KIPP entry (outcome years 1 to 4). Separately, we also compare the baseline characteristics of the students in the science and history outcome samples.

The matching process included all 5th and 6th grade student cohorts with at least one year of outcome data. The analytic sample size decreases in subsequent outcome years for two main reasons: first, more recent student cohorts had fewer years of available outcome data than earlier cohorts, so fewer were included. Second, within a given cohort, we observed sample attrition at the student level as students transfer out of the jurisdiction or otherwise drop out of the dataset. As a result, impact estimates beyond the first year after KIPP entry do not include all treatment and matched comparison students measured in Table A.3. The science and history outcome samples are smaller for two reasons—fewer states administer these tests in middle school and the tests are usually (but not always) administered in grade 8, so more recent cohorts that could not have taken the tests are excluded. To investigate whether the treatment and comparison groups maintained baseline equivalence, the following tables repeat the comparison of baseline scores and demographic characteristics for the portion of the initial sample included in each year's impact estimate. The tables also include treatment and comparison sample sizes for each year, which demonstrate that the rate of analytic sample attrition in the treatment group did not differ substantially from the sample attrition rate in the matched comparison group.

As shown in Tables A.3–A.6, students included in the matched samples for the four years of math and reading outcomes maintained baseline equivalence in prior math and reading scores. That

[58] If balance was not achieved in the first round of matching for a given school, we adjusted the propensity score model by removing the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20).

is, in each of the four outcome years the mean baseline math and reading scores of KIPP students are not significantly different from those of matched comparison students. Similarly, for the science and history outcomes (Tables A.7 and A.8), the baseline reading scores of KIPP students are also not significantly different from comparison group scores. However, for the science and social studies outcomes there is a statistically significant difference in baseline math scores: the differences between KIPP and the comparison group (-0.05 standard deviations in the science sample and -0.08 standard deviations in the history sample) are small in magnitude, and suggest that in these two outcome samples the KIPP students slightly underperformed the matched comparison group during the baseline period.

Table A.3. Balance Between KIPP Students and Matched Comparison Students in Year One

Baseline Characteristic	KIPP	Comparison	Difference	Number with Valid Data
Math scores (mean z-score)	-0.135	-0.125	-0.010 (0.011)	31,832
Reading scores (mean z-score)	-0.100	-0.095	-0.006 (0.011)	31,832
Female	0.517	0.506	0.011 (0.006)	31,832
Black	0.664	0.664	0.000 (0.004)	31,832
Hispanic	0.300	0.297	0.003 (0.003)	31,832
Special education	0.083	0.085	-0.002 (0.003)	31,832
Limited English proficiency	0.099	0.102	-0.003 (0.003)	23,494
Free- or reduced-price lunch	0.774	0.774	0.000 (0.005)	29,746

Note: Standard errors reported in parentheses. Total sample includes 15,916 KIPP students and 15,916 matched comparison students. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.4. Balance Between KIPP Students and Matched Comparison Students in Year Two

Baseline Characteristic	KIPP	Comparison	Difference	Number with Valid Data
Math scores (mean z-score)	-0.112	-0.096	-0.016 (0.014)	22,819
Reading scores (mean z-score)	-0.078	-0.066	-0.012 (0.014)	22,819
Female	0.526	0.512	0.014 (0.008)	22,819
Black	0.669	0.662	0.007 (0.005)	22,819
Hispanic	0.297	0.299	-0.002 (0.004)	22,819
Special education	0.074	0.076	-0.002 (0.004)	22,819
Limited English proficiency	0.094	0.101	-0.007** (0.003)	17,165
Free- or reduced-price lunch	0.812	0.814	-0.002 (0.005)	20,704

Note: Standard errors reported in parentheses. Total sample includes 11,607 KIPP students and 11,212 matched comparison students. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.5. Balance Between KIPP Students and Matched Comparison Students in Year Three

Baseline Characteristic	KIPP	Comparison	Difference	Number with Valid Data
Math scores (mean z-score)	-0.120	-0.095	-0.026 (0.017)	16,218
Reading scores (mean z-score)	-0.082	-0.071	-0.012 (0.016)	16,218
Female	0.512	0.493	0.019* (0.010)	16,218
Black	0.649	0.639	0.011 (0.005)	16,218
Hispanic	0.316	0.318	-0.002 (0.005)	16,218
Special education	0.072	0.072	0.000 (0.004)	16,218
Limited English proficiency	0.101	0.111	-0.108** (0.004)	12,147
Free- or reduced-price lunch	0.780	0.780	0.000 (0.007)	15,299

Note: Standard errors reported in parentheses. Total sample includes 8,282 KIPP students and 7,936 matched comparison students. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.6. Balance Between KIPP Students and Matched Comparison Students in Year Four

Baseline Characteristic	KIPP	Comparison	Difference	Number with Valid Data
Math scores (mean z-score)	-0.220	-0.180	-0.041 (0.034)	8,262
Reading scores (mean z-score)	-0.163	-0.124	-0.039 (0.035)	8,262
Female	0.500	0.523	-0.024 (0.020)	8,262
Black	0.698	0.687	0.011 (0.013)	8,262
Hispanic	0.271	0.283	-0.013 (0.013)	8,262
Special education	0.074	0.084	-0.010 (0.015)	8,262
Limited English proficiency	0.065	0.082	-0.017 (0.012)	5,426
Free- or reduced-price lunch	0.768	0.772	-0.004 (0.013)	8,028

Note: Standard errors reported in parentheses. Total sample includes 4,299 KIPP students and 3,963 matched comparison students. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.7. Balance Between KIPP Students and Matched Comparison Students with Science Scores

Baseline Characteristic	KIPP	Comparison	Difference	Number with Valid Data
Math scores (mean z-score)	-0.152	-0.105	-0.047* (0.024)	8,699
Reading scores (mean z-score)	-0.111	-0.073	-0.038 (0.027)	8,699
Female	0.519	0.490	0.029 (0.016)	8,699
Black	0.630	0.617	0.013 (0.011)	8,699
Hispanic	0.324	0.322	0.002 (0.007)	8,699
Special education	0.069	0.057	0.011* (0.006)	8,699
Limited English proficiency	0.120	0.126	-0.006 (0.005)	5,971
Free- or reduced-price lunch	0.826	0.827	0.000 (0.010)	8,317

Note: Standard errors reported in parentheses. Total sample includes 4,386 KIPP students and 4,313 matched comparison students. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.8. Balance Between KIPP Students and Matched Comparison Students with Social Studies Scores

Baseline Characteristic	KIPP	Comparison	Difference	Number with Valid Data
Math scores (mean z-score)	-0.191	-0.108	-0.083** (0.029)	6,904
Reading scores (mean z-score)	-0.126	-0.065	-0.061 (0.033)	6,904
Female	0.502	0.485	0.017 (0.018)	6,904
Black	0.667	0.649	0.018 (0.011)	6,904
Hispanic	0.298	0.315	-0.018* (0.007)	6,904
Special education	0.059	0.062	-0.003 (0.006)	6,904
Limited English proficiency	0.110	0.119	-0.009 (0.007)	4,176
Free- or reduced-price lunch	0.829	0.831	-0.002 (0.010)	6,493

Note: Standard errors reported in parentheses. Total sample includes 3,385 KIPP students and 3,519 matched comparison students. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

We also tested for equivalence on demographic characteristics and did not find any large differences. The prevalence of specific demographic groups in the treatment and matched comparison outcome samples never differs by more than 2 percentage points. For the first outcome year in math and reading, there are no significant differences between the observed demographic characteristics of KIPP students and the matched comparison group. For the sample with outcomes in the second year, there is a small but statistically significant difference in the percentage of students with limited English proficiency (0.7 percentage points). In the year three sample, there is a small, significant difference in the percentage of female students (1.9 percentage points) and students with limited English proficiency (1.1 percentage points), but in the fourth outcome year none of these differences remain significant. For the science outcome sample, there is a small but significant difference in the percentage of special education students (1.1 percentage points), and in the history outcome sample there is a significant difference of 1.8 percentage points in the percentage of Hispanic students. However, the science and history samples maintained baseline equivalence on all other demographic characteristics tested.

D. Defining Treatment Status: Lottery Sample

Conceptually, the definition of treatment status in the lottery analysis is straightforward. Students in the sample are defined as treatment group members if they are offered admission to a KIPP school in the study on the basis of their draw in the school's admissions lottery. In practice, however, this definition is complicated by the fact that the schools make one set of admissions offers at the time of the lottery, placing other lottery participants on a randomly ordered waiting list. The schools then make additional admissions offers subsequent to the lottery by contacting students

on the waiting list as additional seats in the school become available (for example, as some lottery winners decline their admissions offers). In a few schools, this process of making subsequent admissions offers continues well into the school year. This process raised a question as to whether students offered admission from the waiting list, substantially after the beginning of the school year, receive the "full treatment."

The benchmark decision rule we used in determining treatment status involved selecting a single offer date that would give students the ability to attend KIPP for most of that school year. Offers made after this date were less likely to be considered by families who had made alternate schooling arrangements with which the students may have become comfortable. Since many schools stop making offers of admission after enrollment counts are submitted (typically in early October), the offer date we employed across all sites is October 15 (in 2008 for cohort 1; in 2009 for cohort 2).

We used a slightly different definition of the treatment and control groups in three schools in the sample (Lynn, WILL, and one of the lottery strata at Academy New York) that ultimately exhausted waitlists. In these schools, even though everyone in the lottery was eventually offered admission, there was a large difference in take-up rates—the rate at which students actually accepted the offer of admission and attended KIPP—between those who initially won an offer of admission at the time of the lottery and those who were offered admission later, off the waitlist. This allowed us to implement an alternative experiment (AE) wherein treatment status is defined based on the initial lottery outcome rather than post-lottery waitlist offers. In other words, treatment group students at these sites consist of students whose parents provided consent, were included in the lottery, and were offered admission at the time of the lottery on the basis of the lottery draw (that is, they were *initial* lottery winners). Control group students in AE sites consist of students whose parents provided consent, were included in the lottery, but were not offered admission at the time of the lottery on the basis of the lottery draw (that is, they were *initial* lottery non-winners).

Across all schools, the average probability of winning the lottery for all students was 45 percent; our treatment group also represents 45 percent of the consenting sample (535 out of 1,179 students). In Table A.9, we show the schools included in the lottery-based analysis, lottery details, and the number of treatment and control students from each school.

For the analysis, we created sample weights that accounted for the actual probability that each student was offered admission to a KIPP school (that is, the probability that they were assigned to the treatment group). In some schools, this probability was calculated in a straightforward manner by dividing the number of students in the treatment group by the total number of students in the sample at that school. In other schools, however, the sample weights took into account that not all lottery participants at a school had the same probability of being offered admission, either because of lottery stratification or due to sibling rules. Lottery stratification occurred when a school, in effect, held separate lotteries for separate groups of students. Sibling rules arose when a set of two or more siblings participated in the same lottery at a given KIPP school. In some cases, the school would automatically offer admission to one sibling if the other received a winning lottery draw, such that these simultaneous sibling applicants would be twice as likely to be admitted.

Table A.9. Lottery Detail for Schools Included in Lottery-Based Analysis

School	State	Cohort	Grade	Date	# Exempt	# Lottery Participants	Admission Probability	Consent Rate (%)	# Treatment	# Control	# Total	% Treatment Attended
Academy Lynn	MA	1	5	3/12/08	19	190	0.400	36.8	30	40	70	73.3
		2	5	3/17/09	21	214	0.341	44.4	33	62	95	81.8
Academy Houston	TX	1	5	3/12/08	65	139	0.284	68.9	16	35	51	50.0
		2	5	3/12/09	58	173	0.214	86.1	30	119	149	83.3
Academy New York	NY	2	5	4/7/09	16	101 [a]	0.789	79.1	70	15	85	30.4
Academy of Opportunity	CA	2	5	4/30/09	24	29 [a]	0.345	75.9	4	18	22	100.0
		2	6	4/30/09	0	90	0.300	67.8	13	48	61	53.8
Aspire	TX	2	6	5/28/09	19	32	0.313	65.6	9	12	21	100.0
Austin College Prep	TX	2	6	2/17/09	2	64	0.359	89.1	11	46	57	54.5
KEY Academy	DC	2	6	4/1/09	0	67	0.358	49.3	16	17	33	68.8
L.A. Prep	CA	2	6	4/30/09	6	122	0.115	89.3	13	96	109	69.2
South Fulton Academy	GA	1	6	3/4/08	8	75	0.520	88.0	33	33	66	39.4
		2	5	4/15/09	17	111	0.973	77.5	83	3	86	60.2
		2	6	4/15/09	9	118	0.576	79.7	56	38	94	48.2
Summit Academy	CA	1	5	5/30/08	74	76	0.579	72.4	30	25	55	90.0
		2	5	5/29/09	45	80 [a]	0.800	87.5	56	14	70	80.4
TRUTH Academy	TX	2	6	5/19/09	1	21	0.476	81.0	9	8	17	66.7
WAYS Academy	GA	2	6	4/16/09	5	23 [a]	0.826	91.3	17	4	21	88.2
WILL Academy	DC	2	6	4/1/09	0	23	0.348	73.9	6	11	17	100.0
Total					**389**	**1,683**	**0.450**	**70.1**	**535**	**644**	**1,179**	**63.9**

[a] The lottery at these schools were stratified, with different groups being randomized separately. The numbers here reflect only the single stratum at the school in which some students were lottery winners who were offered admission and some were lottery non-winners who were not. Other strata at this school were excluded because either all members of the stratum were lottery winners who were offered admission or all members were lottery non-winners who were not offered admission.

E. Baseline Equivalence: Lottery Sample

The following six tables compare baseline characteristics of the treatment and control samples for the lottery-based analysis. The baseline characteristics for the baseline sample of all students in our sample who participated in the lottery are presented in Table A.10. Next are baseline characteristics for each of the analytic samples (that is, the samples of students who had valid data for particular outcomes): year 1 state test scores (Table A.11), year 2 state test scores (Table A.12), study-administered test (Table A.13), parent survey (Table A.14), and student survey (Table A.15).

For our key baseline measures—reading and math state test scores at baseline and pre-baseline—we find no statistically significant differences between treatment and control groups in any samples. Across the other 19 characteristics measured, there are no significant differences between treatment and control groups for the state test score outcomes. For the other samples, just a few differences reach statistical significance; for example, students in the treatment group are significantly less likely than those in the control group to have mothers with less education than a high school diploma in the four non-state test score samples (baseline, study-administered test, and survey analysis samples).

Table A.10. Baseline Equivalence for Lottery Sample (Baseline Sample)

Baseline Characteristic	Treatment	Control	Difference	Number with Valid Data
Baseline reading score (z-score)	0.026	-0.029	0.055 (0.080)	610
Baseline math score (z-score)	0.048	-0.019	0.067 (0.087)	619
Pre-baseline reading score (z-score)	0.004	-0.106	0.110 (0.083)	553
Pre-baseline math score (z-score)	-0.027	-0.057	0.030 (0.089)	556
Student is female	0.521	0.486	0.035 (0.038)	1170
Age relative to cohort (in years)	0.047	0.008	0.039 (0.035)	1074
Student is Hispanic	0.551	0.557	-0.006 (0.026)	1094
Student is white	0.028	0.032	-0.004 (0.011)	1094
Student is black	0.376	0.333	0.043 (0.024)	1094
Student is other ethnicity	0.045	0.078	-0.033* (0.016)	1094
Student has IEP	0.106	0.130	-0.024 (0.026)	981
Student received free or reduced-price lunch	0.833	0.778	0.055 (0.029)	1020
Primary language at home is English	0.536	0.516	0.020 (0.030)	1067
Household has only one adult	0.266	0.236	0.030 (0.038)	962
Family income is less than $15k	0.191	0.202	-0.011 (0.029)	920
Family income is $15k to $25k	0.230	0.234	-0.004 (0.032)	920
Family income is $25k to $35k	0.213	0.200	0.013 (0.041)	920
Family income is $35k to $55k	0.219	0.188	0.031 (0.032)	920
Family income is greater than $55k	0.147	0.175	-0.028 (0.036)	920
Mother has less than HS education	0.203	0.267	-0.064* (0.027)	963
Mother has HS or GED education	0.299	0.240	0.059 (0.033)	963
Mother has some college education	0.203	0.255	-0.052 (0.038)	963
Mother has at least a college education	0.295	0.238	0.057 (0.039)	963

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data. The difference between lottery winners and non-winners is based on a regression of the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery non-winner mean is the unadjusted mean for lottery non-winners. The lottery winner mean is the sum of the lottery non-winner mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.11. Baseline Equivalence for Lottery Sample (Analytic Sample: Year 1 State Test Scores)

Baseline Characteristic	Treatment	Control	Difference	Number with Valid Data
Baseline reading score (z-score)	0.038	0.011	0.027 (0.086)	488
Baseline math score (z-score)	0.049	0.024	0.025 (0.092)	498
Pre-baseline reading score (z-score)	-0.019	-0.040	0.021 (0.090)	438
Pre-baseline math score (z-score)	0.009	0.010	-0.001 (0.096)	440
Student is female	0.520	0.500	0.020 (0.051)	536
Age relative to cohort (in years)	-0.004	-0.004	-0.000 (0.047)	503
Student is Hispanic	0.641	0.659	-0.018 (0.040)	522
Student is white	0.048	0.028	0.020 (0.018)	522
Student is black	0.283	0.254	0.029 (0.038)	522
Student is other ethnicity	0.028	0.059	-0.031 (0.019)	522
Student has IEP	0.061	0.095	-0.034 (0.036)	481
Student received free or reduced-price lunch	0.848	0.857	-0.009 (0.032)	500
Primary language at home is English	0.486	0.413	0.073 (0.049)	499
Household has only one adult	0.253	0.217	0.036 (0.045)	458
Family income is less than $15k	0.187	0.248	-0.061 (0.046)	445
Family income is $15k to $25k	0.223	0.270	-0.047 (0.052)	445
Family income is $25k to $35k	0.269	0.223	0.046 (0.048)	445
Family income is $35k to $55k	0.192	0.165	0.027 (0.042)	445
Family income is greater than $55k	0.128	0.094	0.034 (0.036)	445
Mother has less than HS education	0.250	0.321	-0.071 (0.044)	460
Mother has HS or GED education	0.348	0.300	0.048 (0.051)	460
Mother has some college education	0.195	0.220	-0.025 (0.042)	460
Mother has at least a college education	0.207	0.160	0.047 (0.045)	460

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data. The difference between lottery winners and non-winners is based on a regression of the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery non-winner mean is the unadjusted mean for lottery non-winners. The lottery winner mean is the sum of the lottery non-winner mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.12. Baseline Equivalence for Lottery Sample (Analytic Sample: Year 2 State Test Scores)

Baseline Characteristic	Treatment	Control	Difference	Number with Valid Data
Baseline reading score (z-score)	0.055	0.011	0.044 (0.091)	389
Baseline math score (z-score)	0.034	0.010	0.024 (0.085)	399
Pre-baseline reading score (z-score)	-0.044	-0.066	0.022 (0.097)	343
Pre-baseline math score (z-score)	-0.021	-0.028	0.007 (0.094)	345
Student is female	0.525	0.494	0.031 (0.054)	442
Age relative to cohort (in years)	0.047	0.044	0.003 (0.054)	410
Student is Hispanic	0.556	0.582	-0.026 (0.049)	427
Student is white	0.059	0.032	0.027 (0.021)	427
Student is black	0.342	0.321	0.021 (0.047)	427
Student is other ethnicity	0.042	0.064	-0.022 (0.024)	427
Student has IEP	0.050	0.100	-0.050 (0.043)	384
Student received free or reduced-price lunch	0.822	0.837	-0.015 (0.035)	405
Primary language at home is English	0.523	0.433	0.090 (0.052)	406
Household has only one adult	0.296	0.242	0.054 (0.055)	366
Family income is less than $15k	0.176	0.198	-0.022 (0.050)	353
Family income is $15k to $25k	0.217	0.266	-0.049 (0.057)	353
Family income is $25k to $35k	0.274	0.261	0.013 (0.050)	353
Family income is $35k to $55k	0.177	0.169	0.008 (0.042)	353
Family income is greater than $55k	0.157	0.106	0.051 (0.042)	353
Mother has less than HS education	0.255	0.301	-0.046 (0.046)	368
Mother has HS or GED education	0.300	0.296	0.004 (0.054)	368
Mother has some college education	0.200	0.208	-0.008 (0.046)	368
Mother has at least a college education	0.245	0.194	0.051 (0.053)	368

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data. The difference between lottery winners and non-winners is based on a regression of the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery non-winner mean is the unadjusted mean for lottery non-winners. The lottery winner mean is the sum of the lottery non-winner mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.13. Baseline Equivalence for Lottery Sample (Analytic Sample: TerraNova Test)

Baseline Characteristic	Treatment	Control	Difference	Number with Valid Data
Baseline reading score (z-score)	0.054	-0.048	0.102 (0.101)	356
Baseline math score (z-score)	0.110	0.050	0.060 (0.110)	361
Pre-baseline reading score (z-score)	0.073	-0.065	0.138 (0.108)	328
Pre-baseline math score (z-score)	0.059	-0.061	0.120 (0.111)	330
Student is female	0.563	0.481	0.082 (0.050)	590
Age relative to cohort (in years)	0.061	-0.047	0.108* (0.048)	552
Student is Hispanic	0.546	0.582	-0.036 (0.033)	579
Student is white	0.028	0.026	0.002 (0.013)	579
Student is black	0.393	0.341	0.052 (0.032)	579
Student is other ethnicity	0.033	0.051	-0.018 (0.014)	579
Student has IEP	0.106	0.128	-0.022 (0.034)	527
Student received free or reduced-price lunch	0.843	0.815	0.028 (0.036)	547
Primary language at home is English	0.561	0.484	0.077* (0.037)	573
Household has only one adult	0.257	0.227	0.030 (0.057)	520
Family income is less than $15k	0.185	0.222	-0.037 (0.039)	502
Family income is $15k to $25k	0.237	0.273	-0.036 (0.044)	502
Family income is $25k to $35k	0.191	0.196	-0.005 (0.057)	502
Family income is $35k to $55k	0.220	0.156	0.064 (0.037)	502
Family income is greater than $55k	0.167	0.153	0.014 (0.049)	502
Mother has less than HS education	0.184	0.292	-0.108** (0.036)	518
Mother has HS or GED education	0.304	0.246	0.058 (0.042)	518
Mother has some college education	0.162	0.210	-0.048 (0.051)	518
Mother has at least a college education	0.350	0.253	0.097 (0.050)	518

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data. The difference between lottery winners and non-winners is based on a regression of the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery non-winner mean is the unadjusted mean for lottery non-winners. The lottery winner mean is the sum of the lottery non-winner mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.14. Baseline Equivalence for Lottery Sample (Analytic Sample: Parent Survey)

Baseline Characteristic	Treatment	Control	Difference	Number with Valid Data
Baseline reading score (z-score)	0.018	-0.017	0.035 (0.094)	448
Baseline math score (z-score)	0.052	0.023	0.029 (0.100)	457
Pre-baseline reading score (z-score)	-0.007	-0.073	0.066 (0.095)	413
Pre-baseline math score (z-score)	-0.018	-0.008	-0.010 (0.103)	414
Student is female	0.547	0.464	0.083 (0.043)	848
Age relative to cohort (in years)	0.058	0.002	0.056 (0.040)	797
Student is Hispanic	0.539	0.569	-0.030 (0.030)	845
Student is white	0.032	0.033	-0.001 (0.013)	845
Student is black	0.383	0.326	0.058* (0.027)	845
Student is other ethnicity	0.046	0.073	-0.027 (0.018)	845
Student has IEP	0.112	0.132	-0.020 (0.030)	749
Student received free or reduced-price lunch	0.839	0.779	0.060 (0.033)	771
Primary language at home is English	0.539	0.496	0.043 (0.034)	845
Household has only one adult	0.249	0.212	0.037 (0.045)	740
Family income is less than $15k	0.186	0.204	-0.018 (0.029)	712
Family income is $15k to $25k	0.246	0.243	0.003 (0.037)	712
Family income is $25k to $35k	0.216	0.188	0.028 (0.048)	712
Family income is $35k to $55k	0.201	0.188	0.013 (0.034)	712
Family income is greater than $55k	0.151	0.177	-0.026 (0.043)	712
Mother has less than HS education	0.192	0.279	-0.087** (0.031)	742
Mother has HS or GED education	0.307	0.236	0.071* (0.035)	742
Mother has some college education	0.198	0.239	-0.041 (0.044)	742
Mother has at least a college education	0.304	0.247	0.057 (0.044)	742

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data. The difference between lottery winners and non-winners is based on a regression of the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery non-winner mean is the unadjusted mean for lottery non-winners. The lottery winner mean is the sum of the lottery non-winner mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

Table A.15. Baseline Equivalence for Lottery Sample (Analytic Sample: Student Survey)

Baseline Characteristic	Treatment	Control	Difference	Number with Valid Data
Baseline reading score (z-score)	0.016	-0.024	0.040 (0.096)	409
Baseline math score (z-score)	0.082	0.044	0.038 (0.107)	415
Pre-baseline reading score (z-score)	0.015	-0.044	0.059 (0.101)	376
Pre-baseline math score (z-score)	-0.001	0.009	-0.010 (0.110)	377
Student is female	0.567	0.480	0.087 (0.044)	756
Age relative to cohort (in years)	0.041	-0.003	0.044 (0.041)	713
Student is Hispanic	0.546	0.579	-0.033 (0.031)	756
Student is white	0.035	0.037	-0.002 (0.015)	756
Student is black	0.376	0.315	0.062* (0.028)	756
Student is other ethnicity	0.042	0.069	-0.027 (0.017)	756
Student has IEP	0.117	0.127	-0.010 (0.031)	674
Student received free or reduced-price lunch	0.844	0.784	0.060 (0.035)	695
Primary language at home is English	0.541	0.484	0.057 (0.035)	755
Household has only one adult	0.260	0.200	0.060 (0.047)	666
Family income is less than $15k	0.186	0.199	-0.013 (0.031)	643
Family income is $15k to $25k	0.246	0.252	-0.006 (0.039)	643
Family income is $25k to $35k	0.213	0.186	0.027 (0.051)	643
Family income is $35k to $55k	0.200	0.193	0.007 (0.036)	643
Family income is greater than $55k	0.156	0.171	-0.015 (0.045)	643
Mother has less than HS education	0.186	0.284	-0.098** (0.034)	670
Mother has HS or GED education	0.299	0.245	0.054 (0.037)	670
Mother has some college education	0.176	0.227	-0.051 (0.046)	670
Mother has at least a college education	0.340	0.245	0.095* (0.044)	670

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data. The difference between lottery winners and non-winners is based on a regression of the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery non-winner mean is the unadjusted mean for lottery non-winners. The lottery winner mean is the sum of the lottery non-winner mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

*Significantly different from zero at the 0.05 level, two-tailed test.
**Significantly different from zero at the 0.01 level, two-tailed test.

APPENDIX B

CONSTRUCTING SURVEY OUTCOMES

Several of the measures in Chapters III, V, and VI of this report are derived from survey items. Many of the outcomes are indices created by combining closely related survey items into a single measure, reducing measurement error, and capturing the breadth of a construct. The indices and their component items are listed in Table B.1 for principal survey items and Table B.2 for student and parent survey items. Both tables also include survey outcomes or measures that are not indices, but are derived from one or more survey items.

The process for creating the indices included a number of steps to maximize reliability and reduce dimensionality. We first identified all items from the surveys that were conceptually related to a specific construct. We used principal component analysis to confirm that the composite was unidimensional, excluding items not related to the underlying construct. For the indices used in the factors chapter, we then standardized the values for each item, such that the overall mean for all KIPP schools had a value of 0 and a standard deviation of 1.[59] We did not standardize the indices used for other chapters because the non-standardized versions were easier to meaningfully interpret. We then computed the standardized Cronbach's alpha, an estimate of the internal consistency or reliability of an index, and dropped indices with alpha values suggesting low reliability. [60]

[59] Before standardizing the indices, we checked to confirm there was sufficient variation in the responses to justify using them as a factor. In all cases, there was variation across respondents, but it was consolidated around a limited interval within the full range of possible values for the indices. As a result, we standardized the indices to better capture how the existing variation in the indices is associated with the variation in impacts.

[60] Conventionally, indices with alpha values greater than 0.7 are considered reliable. Following Gleason et al. (2010), we retained indices with alpha values somewhat lower than this threshold but indicate that these indices may have low levels of reliability. Indices with values of alpha below 0.7 are noted in the chapter tables.

Table B.1. Construction of Principal Survey Outcomes

Variable	Principal Survey Items Included	Scale/Definition
Total enrollment	A5. Around October 2010, how many students in grades K–12 were enrolled in this school?	Number of students
Enrollment per grade	A2. Which grades are offered in this school? Select all that apply.	Total enrollment divided by the number of grades reported
Enrolled students who withdrew	A8. Please indicate the number of students who left this school between October 1, 2012 and May 30, 2011. Please do not include students who graduated in your count.	Number of students who left divided by total enrollment
Enrolled students who enrolled mid-year	A9. Between October 1, 2010 and May 30, 2011, did any new students enroll in this school?	Number of new students enrolled divided by total enrollment (0 for schools reporting no new students)
	(Conditional on reporting "Yes" to the previous question):	
	A10. Please indicate the number of new students who have enrolled in this school.	
ELA/math class size	B3. What is the typical number of students in a language arts class at this school?	Mean number of students reported in language arts and math classes
	B5. What is the typical number of students in a math class at this school?	
Student-teacher ratio	C1. As of October 1, 2010, how many teachers and long-term substitute teachers held part-time and full-time positions at this school?	Total enrollment divided by number of full-time teachers
	(a) Part-time	
	(b) Full-time	
School day length in hours	A11. How long is the school day for students in this school?	Number of hours
Hours per day spent in core classes	B7. During a typical full week of school, approximately how many minutes do most 7th grade students spend on the following activities at this school?	Sum of hours spent in all four categories, converted to hours per day
	(a) English, reading, or language arts	
	(b) Arithmetic or math	
	(c) Social studies or history	
	(d) Science	
Hours per day spent outside of core classes	See items A11 and B7.	Difference between school day length and hours per day spent in core classes
School year length in days	A14. How many days are in the school year for students in this school? If applicable, please include Saturday school days in your count.	Number of days
School requires students to attend Saturday school	A12. Does this school require students to attend Saturday school?	Proportion answering "yes"
Number of days students attend Saturday school per month	(Conditional on reporting that the school requires students to attend Saturday school):	Number of Saturdays required (0 for schools reporting that Saturday school was not required)
	A13. In a typical month, how many Saturdays does this school require students to attend classes?	

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Average daily attendance	A15. For the 2010–2011 school year, what was the average daily attendance (ADA) at this school?	Percent
School serves as its own district	F1. Does your school serve as its own district or is it part of the local school district in which it is located?	Proportion answering that “school is its own district”
All core classes have students with mixed ability levels (ELA or math)	B2. Which of the following best describes this school’s approach to providing instruction in math and English/language arts to regular students? Do all, some, or none of the classes in core subjects have students assigned into classrooms of mixed ability levels?	Proportion answering “all classes have mixed ability levels” for both English/language arts and math
	(1) All classes have mixed ability levels	
	(2) Some classes have mixed ability levels	
	(3) No classes have mixed ability levels	
	(4) Not applicable, only one class per grade	
Students loop through multiple grades with teacher	B1. Is this school using the following methods to organize classes or student groups?	Proportion reporting “student groups that remain with the same teacher for two or more years (e.g., looping)”
	(a) Traditional grades or academic discipline-based departments	
	(b) Grades subdivided into small groups, such as “houses,” “families,” or “teams”	
	(c) Student groups that remain with the same teacher for two or more years (e.g., looping)	
	(d) Interdisciplinary teaching (when two or more teachers with different academic specializations collaborate to teach an interdisciplinary program to the same group of students)	
	(e) Paired or team teaching (when two or more teachers are in the same class at the same time and are jointly responsible)	
School uses interdisciplinary teaching	See item B1.	Proportion reporting ”interdisciplinary teaching”
School uses paired/team teaching	See item B1.	Proportion reporting “paired or team teaching”
Primary 7th grade math textbook is “no textbook”	B8. Please indicate which of the following textbooks are used in any 7th grade Math course at this school. Then, select the textbook that is used most often.	Proportion answering “no textbook” for the textbook used most often
Talented/gifted program for core subjects	B15. Which of the following types of enrichment programming are offered at your school?	Proportion indicating a “talented/gifted program for core subjects” is offered
	(a) Talented/gifted program for core subjects	
	(b) Talented/gifted program for other subjects	
	(c) Magnet program for the arts	
	(d) Magnet program for science and/or math	
	(e) Magnet program for general academics	
	(f) Other enrichment programming (specify)	

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Music and/or art program	B14. Which of the following programs or facilities are available to students in your school?	Proportion indicating “a music program” or “an arts program” is available
	(a) Special programs for non-English speakers	
	(b) Individual tutors	
	(c) A computer lab	
	(d) A library	
	(e) A gym	
	(f) A cafeteria	
	(g) Child counselors	
	(h) A nurse’s office	
	(i) A music program	
	(j) A physical education program	
	(k) An after-school program	
	(l) Breakfasts prepared at the school	
	(m) Lunches prepared at the school	
	(n) A before-school program	
	(o) An arts program	
Before- or after-school programming	See item B14.	Proportion indicating “a before-school program” or “an after school program” is available
Individual tutoring	See item B14.	Proportion indicating “individual tutors” are available
Limited English proficiency instruction for students	B9. Does this school have instruction specifically designed to address the needs of limited-English proficient students?	Proportion answering “yes”
Services for parents with limited-English skills (interpreters or translations of printed materials)	B13. Does this school provide the following services for any parents with limited-English skills?	Proportion answering “yes” to either
	(a) Interpreters for meetings or parent-teacher conferences	
	(b) Translations of printed materials, such as newsletters, school notices, or school signs	

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Index of use of school-wide behavior plan	E1. To what extent do you agree with the following statements? (a) Behavioral standards and discipline policies are established and enforced consistently across the entire school (b) Our school has a zero-tolerance policy for potentially dangerous behaviors (c) We have a school-wide behavior code that includes specific positive rewards for students who consistently behave well (d) We have a school-wide behavior code that includes specific negative sanctions for students who violate the rules (e) Instructional practices allow teachers to flexibly address the interests and needs of individual students (f) At this school, it is difficult to overcome the cultural barriers between teachers and parents (g) Staff at this school work hard to build trusting relationships with parents	Mean across items (a), (c), and (d) using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Percentage of enrolled students expelled from school	E7. During the 2010–2011 school year, how many students were expelled from this school, that is, removed or transferred for at least the remainder of the school year? If none, enter "0."	Number of students expelled divided by total enrollment
Percentage of enrolled students suspended out-of-school	E8. Does your school have policies or practices in place under which students can be suspended? (Conditional on reporting "Yes" to the previous question) E9. During the 2010–2011 school year, what was the total number of (a) in-school suspensions and (b) out-of-school suspensions, and the number of students who received each?	Number of students receiving out-of-school suspensions (substituting 0 for those who reported "no" to the first question) divided by total enrollment
Parents make participation commitments (e.g., interview, orientation session, commitment form)	A16. Before new students enroll in this school, which of the following are parents or students required to do? (1) Students must take an achievement test for diagnostic purposes (2) Students must meet some academic requirement, such as a minimum score on an achievement test or minimum GPA, or demonstrate special aptitude, skills, or talents (such as in music or the arts) (3) Parents must submit recommendations from adults who know the student (4) Parents are required to meet certain participation requirements, such as completing a personal interview, attending an orientation, or signing a commitment agreement (5) Students must sign an agreement describing their responsibilities (for example, related to their academic efforts or attendance in school) (6) Other (please specify)	Proportion answering "parents are required to meet certain participation requirements, such as completing a personal interview, attending an orientation, or signing a commitment agreement"
Students must sign a responsibilities agreement	See item A16.	Proportion reporting this activity

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Index of quality of parent/staff interaction	See item E1.	Mean across items (f) and (g) using the following scale [with scale reversed for item (f)]:
		Strongly disagree (1)
		Disagree (2)
		Agree (3)
		Strongly agree (4)
Index of amount of parent involvement in school activities [a]	E5. During the 2010–2011 school year, what proportion of parents participated in the following activities at your school?	Mean across items (a), (c), and (e) using the following scale:
	(a) Parents participated in instruction	None/not offered (1)
	(b) Parents attended parent/teacher conferences	Few (2)
	(c) Parents accompanied students on class trips	Some (3)
	(d) Parents attended open houses or back-to-school nights	All (4)
	(e) Parents attended education workshops	
School provides parents weekly or daily notes about their child's progress	E4. Does your school routinely provide any of the following to parents?	Proportion answering "weekly or daily notes about their child's progress"
	(a) Information about their child's grades halfway through the grading period	
	(b) Notification when their child is sent to the office for disruptive behavior	
	(c) Weekly or daily notes about their child's progress	
	(d) A newsletter about what's going on in their child's school or school system	
Number of full-time teachers	C1. As of October 1, 2010, how many teachers and long-term substitute teachers held part-time and full-time positions at this school?	Number of full-time teachers
Number of years as principal	D2. Prior to this school year, how many years did you serve as the principal of this school?	Sum of years as principal of this and other schools
	D3. Prior to this school year, how many years did you serve as the principal at any other school?	
Number of years of teaching experience before becoming a principal	D4. Before becoming a principal, how many years of elementary or secondary teaching experience did you have?	Number of years
Teachers with more than four years of experience	C4. As of October 1, 2010, how many teachers in your school had the following levels of elementary or secondary experience?	Sum of teachers with "four to nine years" and "ten or more years" divided by sum of the number teachers reported for all categories
	(a) No prior experience	
	(b) One year	
	(c) Two or three years	
	(d) Four to nine years	
	(e) Ten or more years	

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Teachers at school with full state certification	C2. As of October 1, 2010, of your school's instructional staff, how many had full state certification for the subjects and grade levels they taught in your school?	Number with full state certification divided by total number of full-time teachers
Principal time on work-related activities	D7. How many hours do you spend on all school-related activities during a typical full week at this school? Include hours spent during the school day, before and after school, and on the weekends.	Hours per week
Index of frequency of principal time on problematic issues	D9. How many times in a typical month do you spend at least an hour resolving…	Mean across items (e), (f), and (g) using the following scale:
	(a) Financial or payroll issues	Never (1)
	(b) Issues about facilities leases	1-2 times (2)
	(c) Issues about buildings or grounds maintenance	3-5 times (3)
	(d) Conflicts with school or district board	6 or more times (4)
	(e) Complaints from parents	
	(f) Conflicts among teachers	
	(g) Individual teacher complaints	
	(h) Union grievances	

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Index of principal satisfaction	D10. For the following statements, please select whether you strongly disagree, disagree, agree or strongly agree.	Mean across items (a), (l), (m), (n), and (o) using the following scale:
	(a) Overall, I am satisfied with this school	Strongly disagree (1)
	(b) I am proud to tell others that I work here	Disagree (2)
	(c) My day to day work makes good use of my strengths	Agree (3)
	(d) I have access to the tools I need to do my job	Strongly agree (4)
	(e) Someone at work has given me positive feedback in the past week	
	(f) I plan to work at this school for at least three more years	
	(g) My school's mission is important to me	
	(h) With hard work, all students at this school are capable of attending college	
	(i) I participate in professional development that helps me improve in my job	
	(j) I am continuously developing my leadership skills	
	(k) I have the support from my leadership team that I need to manage the demands of my role	
	(l) I find the length of the school day manageable	
	(m) The school keeps its best teachers and staff	
	(n) I am happy with our current staff retention/turnover level	
	(o) I am fairly compensated for my work	
Number of principals at the school in the past three years	D5. Including yourself, how many principals has this school had over the last three school years (since September 1, 2008)?	Number of principals
Teacher turnover	C10. Did the school dismiss any teachers for performance-related reasons during or following the 2010–2011 school year?	Sum of number of teachers who left for performance related and other reasons divided by number of full-time teachers (0 for schools reporting they did not dismiss or lose any teachers)
	(Conditional on reporting "Yes" to the previous question):	
	C10a. How many teachers were dismissed for performance-related reasons during or following the 2010–2011 school year?	
	C11. Did this school lose any teachers during or following the 2010–2011 school year for other reasons, including teachers moving to other schools, jobs in other organizations, or leaving the labor force?	
	(Conditional on reporting "Yes" to the previous question):	
	C11a. Enter the number of teachers that left for other reasons.	

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Principal reports difficulty obtaining suitable replacements is a barrier to dismissing poor-performing teachers	C9a. Do you consider any of the following factors to be significant barriers to dismissing poor-performing or incompetent teachers at this school?	Proportion reporting "difficulty in obtaining suitable replacements" as a factor
	(a) District personnel policies	
	(b) Length of time or amount of documentation required for termination process	
	(c) Teacher tenure	
	(d) Teacher associations or unions	
	(e) Difficulty in obtaining suitable replacements	
	(f) Resistance from parents	
Number of teacher vacancies on October 1, 2010	C12. As of October 1, 2010, were there teaching vacancies in this school, that is, teaching positions for which teachers were recruited and interviewed?	Number of vacancies divided by the number of full-time teachers (0 for schools reporting no vacancies)
	(Conditional on reporting "Yes" to the previous question):	
	C12a. Enter the number of vacancies.	
Applicants were not a good fit for school culture/goals	C13. In general, how easy or difficult is it to fill the vacancies in this school?	Proportion reporting "applicants were not a good fit for school culture/goals"
	(Conditional on reporting "somewhat difficult" or "very difficult" for the previous question):	
	C14. Thinking about the teacher vacancies that are difficult to fill, what are the reasons for the difficulty? (1) No applicants	
	(2) Applicants were not qualified	
	(3) Applicants were not a good fit for school culture/goals	
	(4) We made offers, but they were not accepted	
	(5) We were not able to offer a competitive compensation package	
	(6) Candidates had multiple other offers	
	(7) Vacancies were in a high need or shortage area	
	(8) Other (please specify)	
Applicants were not qualified	See item C14.	Proportion reporting "applicants were not qualified"
Vacancies were in a high-need or shortage area	See item C14.	Proportion reporting "vacancies were in a high-need or shortage area"
Midpoint of $ teacher salary range at school	C16. Currently, what is the range of yearly base salaries for full-time teachers at this school?	Average of minimum and maximum

Table B.1 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
School provides teacher incentive pay in "hard-to-staff" locations	C17. Does this school (or your district) currently use any pay incentives such as cash bonuses, salary increases, or different steps on the salary schedule to…	Proportion reporting "recruit or retain teachers in a less desirable location"
	(a) Reward teachers who have attained National Board for Professional Teaching Standards certification	
	(b) Reward excellence in teaching	
	(c) Recruit or retain teachers to teach in a less desirable location	
	(d) Recruit or retain teachers to teach in fields of shortage	
School provides teacher incentive pay in "hard-to-staff" subjects	See item C17.	Proportion reporting "recruit or retain teachers in fields of shortage"
School provides teacher incentive pay for excellence in teaching	See item C17.	Proportion answering "reward excellence in teaching"
Teachers covered by collective bargaining	C18. Are your school's teachers covered by a collective bargaining agreement?	Proportion answering "yes"
Index of intensity of new teacher coaching	C5. On average, how many times per year do new teachers experience the following at your school?	Mean across items (a) and (c) using the following scale:
	(a) Observed by a master teacher or someone else who coaches teachers	Not at all (1)
	(b) Observed by a principal, administrator, or someone else who monitors performance	Once (2)
	(c) Received feedback from someone who observed them teach	2–3 times (3)
	(d) Provided with diagnostic test results for individual students to help them determine which topics/skills to focus on	4–7 times (4)
	(e) Participated in content-related professional development	8 or more times (5)
	(f) Asked to submit lesson plans to master teacher, department chair, principal, or other administrator for review	
Index of intensity of experienced teacher coaching	C6. On average, how many times per year do experienced teachers experience the following at your school?	Mean across items (a), (b), (c), and (d) using the following scale:
	(a) Observed by a master teacher or someone else who coaches teachers	Not at all (1)
	(b) Observed by a principal, administrator, or someone else who monitors performance	Once (2)
	(c) Received feedback from someone who observed them teach	2–3 times (3)
	(d) Provided with diagnostic test results for individual students to help them determine which topics/skills to focus on	4–7 times (4)
	(e) Participated in content-related professional development	8 or more times (5)
	(f) Asked to submit lesson plans to master teacher, department chair, principal, or other administrator for review	

Note: The scales were re-oriented to be unidimensional in this table.

Table B.2. Construction of Student and Parent Survey Outcomes

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Count of extracurricular activities	C4. During the 2010–11 school year, please tell me whether you (participate/participated) in any of the following activities at school outside of your normal classes.	Sum across all items using the following scale:
	(a) Student government	Yes (1)
	(b) Band, orchestra, chorus, or choir	No (0)
	(c) School plays or musicals	
	(d) Organized sports or exercise	
	(e) A school yearbook, newspaper or magazine	
	(f) Community service activities	
	(g) Academic clubs, such as a math club, foreign language club, or an academic honor society, like the National Junior Honor Society	
	(h) Other types of clubs, for example, an arts or crafts club, computer club, drama club, or games club	
	(i) Tutoring	
	(j) Do you participate in any other activities at school that I have not already mentioned? (SPECIFY)	
Student reports having homework on a typical night	C5. On a typical school night, are you given homework? [student]	Proportion answering “yes”
Minutes spent on homework on typical night, student report	C6. On a typical school night, how much time do you spend doing your homework? [student]	Minutes reported
Minutes spent on homework on typical night, parent report	B2. On an average night, about how much time does/did that homework take? [parent]	
Parent says student typically completes homework	B3. How often does he/she complete all of the homework he/she is given? [parent]	Proportion answering “almost every day”

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of school engagement	C7. Now I'm going to read you some statements about school. For each, please tell me if you do each "almost always," "often," "sometimes," or "almost never." (a) Stick with a class assignment or task until it is done (b) Put in your best effort on class assignments, projects, and homework (c) Ask a teacher for help when you don't understand an assignment (d) Ask another student for help when you don't understand an assignment (e) Take part in class discussions or activities (f) Feel challenged in class (g) Receive recognition or praise for doing good school work (h) Learn from your mistakes at school (i) Complete class assignments, projects, and homework on time (j) Think of dropping out of school (k) Try to stay home from school	Mean across items (a), (b), (h), (i), (j), and (k) using the following scale (scale reversed for items j and k): Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of self control	C8. Last week in school/Thinking about a typical week during the 2010–11 school year, how many days did you do each of the following things? (a) Went to all of your classes prepared (b) Remained calm even when things happened that could upset you (c) Paid attention in all of your classes (d) Listened to other students speak without interrupting them (e) Were polite to adults and other students (f) Remembered and followed directions (g) Controlled your temper (h) Got to work right away rather than procrastinating	Mean number of days across all items

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of academic self-concept	B2. I'm going to read some more statements. For these, tell me how much you agree or disagree with each. For each statement I read, please tell me if you "strongly agree," "agree," "disagree," or "strongly disagree." (a) You like to work with other students (b) You learn things quickly in most school subjects (c) Because reading is fun, you wouldn't want to give it up (d) You are good at most school subjects (e) You learn most when you work with other students (f) English/Language Arts is one of your best subjects (g) You do your best work when you work with other students (h) Math is one of your best subjects (i) You like to help other people do well in group assignments (j) You do well in tests in most school subjects (k) It is helpful to put together everyone's ideas when you work on a project	Mean across all items using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of effort and persistence in school	B1. Now I'm going to read you some statements about your schoolwork. For each please tell me if these things apply to you "almost always," "often," "sometimes," or "almost never"? (a) You're certain you can understand even the most difficult material presented in textbooks or other written material (b) You can learn something really difficult when you want to (c) In school you work as hard as possible (d) You're certain you can understand even the most difficult material presented by the teacher (e) If you decide to not get any bad grades, you can really do it (f) In school, you keep working even if the material is difficult (g) You're certain you can do an excellent job on assignments and tests (h) You try to do your best to learn the knowledge and skills taught (i) You work hard in school so you can get a good job (j) If you want to learn something well, you can (k) You're certain you can master the material you are taught (l) If you don't understand something in your schoolwork, you try to find additional information to help you learn (m) You put forth your best effort in school	Mean across all items using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Student expects to graduate HS on time	D1. As things stand now, do you think you will… (1) Graduate from high school on time, in four years, (2) Graduate from high school late, taking more than four years, (3) Drop out of high school and complete the GED, or (4) Not finish high school or the GED	Proportion answering "graduate from high school on time, in four years"
Parent expects student to graduate HS on time	E1. As things stand now, do you think he/she will… (1) Graduate from high school on time, in four years, (2) Graduate from high school late, taking more than four years, (3) Drop out of high school and complete the GED, or (4) Not finish high school or the GED	Proportion answering "graduate from high school on time, in four years"
Student wishes to complete college	D2. How far would you like to get in school? (1) High school graduate or GED, (2) Vocational, trade, or business school, or (3) Graduate from college or a higher level of school after graduating college?	Proportion answering "graduate from college or a higher level of school"
Parent wishes student to complete college	E2. How far would you like her/him to go in school? (1) High school graduate or GED, (2) Vocational, trade, or business school, or (3) Graduate from college or a higher level of school after graduating college?	Proportion answering "graduate from college or a higher level of school"
Student believes very likely to complete college	D3. How likely is it that you will get this far in school? Would you say it is… (1) Very likely, (2) Likely, (3) Unlikely, or (4) Very unlikely?	Among those answering "graduate from college or a higher level of school," proportion answering "very likely"
Parent believes student very likely to complete college	E3. How likely is it that he/she will get this far in school? Would you say it is… (1) Very likely, (2) Likely, (3) Unlikely, or (4) Very unlikely?	Among those answering "graduate from college or a higher level of school," proportion answering "very likely"

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Student reports having discussions about college at school	D6. During the 2010–11 school year, how often (have/did) you (discussed/discuss) going to college with teachers or other school staff, such as a guidance counselor? Would you say…	Proportion answering “about once or twice” or “more than twice”
	(1) Never,	
	(2) About once or twice during the school year, or	
	(3) More than twice during the school year	
Student reports having discussions about college at home	D5. During the 2010–11 school year, how often (have/did) you (discussed/discuss) going to college with a parent or guardian? Would you say…	Proportion answering “about once or twice” or “more than twice”
	(1) Never,	
	(2) About once or twice during the school year, or	
	(3) More than twice during the school year	
Parent reports having discussions about college	E5. During the 2010–11 school year, how often did you discuss college with him/her? Would you say…	Proportion answering “about once or twice” or “more than twice”)
	(1) Never,	
	(2) About once or twice during the school year, or	
	(3) More than twice during the school year?	
Index of peer pressure for bad behaviors	E2. During the 2010–11 school year, (do/did) your friends pressure you to do any of the following things “often,” “sometimes, or “never”? Do your friends pressure you to:	Mean across all items using the following scale:
	(a) Skip class or school	Never (1)
	(b) Drink alcohol	Sometimes (2)
	(c) Smoke cigarettes	Often (3)
	(d) Use marijuana or other drugs	
	(e) Commit a crime or do something violent	

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of undesirable behavior	E1. During the 2010–11 school year, how often (do/did) you do each of the following things? (a) Argue with your parents or guardians (b) Smoke cigarettes (c) Lie to your parents or guardians (d) Take something from a store without paying for it (e) Give a teacher a hard time (f) Drink alcohol (g) Skip, or cut, classes during the school day (h) Skip, or cut, the entire school day (i) Use marijuana or other drugs (j) Get in trouble at school (k) Lose your temper at home or at school (l) Get arrested or held by police	Mean across items (a), (c), (e), and (k) using the following scale: Never (1) Sometimes (2) Often (3)
Index of illegal action	See item E1.	Mean across items (b), (d), (f), (g), and (i) using the following scale: Never (1) Sometimes (2) Often (3)
Parent reported any school disciplinary problems for student	F3. During the 2010–11 school year, how many times (has/was) (he/she) (been)… (1) Sent out of class for disciplinary reasons (2) Suspended from school (3) Expelled from school	Proportion answering a nonzero response for any of these items
Index of parent-reported frequency of school disciplinary actions for student	See item F3.	Mean across all items using the following scale: Not at all (0) 1-3 times (1) 4-6 times (2) 7-10 times (3) More than 10 times (4)
Student never gets in trouble at school	See item E1.	Proportion answering “never” to “get into trouble at school”

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of good behavior, student report	E3. During the 2010–11 school year, (do/did) you do each of the following things "often," "sometimes, or "never"?	Mean across all items using the following scale:
	(a) Help another student with school work	Never (1)
	(b) Help people in your local community, for example, help a neighbor or do volunteer work	Sometimes (2)
	(c) Read for fun	Often (3)
	(d) Go to the library outside of school	
	(e) Help your parents or guardians with chores	
Index of good behavior, parent report	F5. During the 2010–11 school year, how often did (he/she) do the following things? Would you say "often," "sometimes, or "never"?	Mean across all items using the following scale:
	(a) Help you with chores or other tasks	Never (1)
	(b) Help people in your local community, for example, help a neighbor or do volunteer work	Sometimes (2)
	(c) Read for fun	Often (3)
	(d) Go to the library outside of school	
Index indicating well-adjusted student	F1. Now I'm going to ask you some questions about (STUDENT'S NAME)'s behavior. For each of the following statements, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree."	Mean across all items using the following scale:
	(a) (He/She) gets along with others	Strongly disagree (1)
	(b) (He/She) likes school	Disagree (2)
	(c) (He/She) works hard at school	Agree (3)
	(d) (He/She) is self-confident	Strongly agree (4)
	(e) (He/She) is creative	
	(f) (He/She) is happy	
	(g) (He/She) respects adults	
Index of parental concerns about student	F2. For each of the following statements, please tell me if it is "not a problem," "a small problem," "a medium problem," or "a big problem" with (STUDENT NAME) in or out of school.	Mean across all items using the following scale:
	(a) Getting into trouble	Not a problem (1)
	(b) Smoking, drinking alcohol or using drugs	A small problem (2)
	(c) The friends (he/she) has chosen	A medium problem (3)
	(d) (His/Her) academic achievement	A big problem (4)

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of student's feelings about school	A1. I'm going to read you some statements on how you (feel/felt) about school (this school year/last school year, for the 2009–2010 school year). For each statement, please tell me if you "strongly agree," "agree," "disagree," or "strongly disagree."	Mean across all items using the following scale:
	(a) You have good friends at your school	Strongly disagree (1)
	(b) You are treated fairly at your school	Disagree (2)
	(c) You are happy to be at your school	Agree (3)
	(d) You feel like you are part of the community in your school	Strongly agree (4)
	(e) You feel safe in your school	
	(f) You are treated with respect at your school	
	(g) You know how you are doing in school	
	(h) You have the materials and equipment you need to do your school work right	
	(i) You get the chance to be independent at school	
	(j) You have opportunities to choose how you learn	
Student likes school a lot	A5. In general, how much do you like the school you (currently attend/attended last year)? Would you say…	Proportion answering "you like it a lot"
	You don't like it at all,	
	You think it is ok, or	
	You like it a lot?	
Index of parental satisfaction with school	D2. Please rate each of the following features of the school (he/she) (attends/attended) for the 2010–11 school year as "excellent," "good," "fair," or "poor."	Mean across all items using the following scale:
	(a) Facilities, like the library, cafeteria, or the gym	Poor (1)
	(b) Academics, the teachers and classes	Fair (2)
	(c) Safety	Good (3)
	(d) Discipline	Excellent (4)
Parent rates school as excellent	D3. Overall, would you rate the school (he/she) (currently attends/attended) for the 2010–11 school year as…	Proportion answering "excellent"
	Excellent,	
	Good,	
	Fair, or	
	Poor?	

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of student perceptions of schoolmates	A2. Please tell me how much you agree or disagree with these statements about the students in your classes (this/last) year at school.	Mean across all items using the following scale:
	(a) Students usually complete their homework	Strongly disagree (1)
	(b) Students get along well with the teachers	Disagree (2)
	(c) Students are interested in learning	Agree (3)
	(d) Students help one another	Strongly agree (4)
	(e) Students are well behaved	
Index of student perceptions of teachers	A3. These next statements are about your teachers (this/last) year at school. Please tell me whether you “strongly agree,” “agree,” “disagree,” or “strongly disagree” with each statement.	Mean across items (a), (b), (c), (d), (e), (g), and (h) using the following scale:
	(a) They are available for help	Strongly disagree (1)
	(b) They listen to what you have to say	Disagree (2)
	(c) They give corrections and suggestions for improvement	Agree (3)
	(d) They care about students	Strongly agree (4)
	(e) They encourage you to think about your future	
	(f) Their classes are challenging	
	(g) They make you feel like your school work is important	
	(h) You like your teachers	
Index of school disciplinary environment	A4. For each of the following statements about the rules (this/last) year at your school, please tell me whether you “strongly agree,” “agree,” “disagree,” or “strongly disagree.”	Mean across all items using the following scale:
	(a) Everyone knows what the school rules are	Strongly disagree (1)
	(b) The school rules are fair	Disagree (2)
	(c) The punishment for breaking school rules is the same no matter who you are	Agree (3)
	(d) If a school rule is broken, students know what the punishment will be	Strongly agree (4)
	(e) You follow the rules at school	

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index of parental perceptions of problems in student's school	D1. For each of the following issues, please tell me if you feel it (is/was) "not a problem," "a small problem," "a medium problem," or "a big problem"?	Mean across all items using the following scale:
	(a) Students destroying property	A big problem (1)
	(b) Students being late for school	A medium problem (2)
	(c) Students missing classes	A small problem (3)
	(d) Fighting	Not a problem (4)
	(e) Bullying	
	(f) Cheating	
	(g) Racial conflict	
	(h) Guns or other weapons at school	
	(i) Drugs or alcohol at school	
Index of parental involvement in student's education	C1. In a typical month during the school year, how often (do/did) you or another family member talk with (STUDENT NAME) about (his/her) experiences in school? Would you say	Mean across all items using the following scale:
	(1) Seldom or never	*For C1 and C2a:*
	(2) Once or twice a month,	Seldom or never (1)
	(3) Once or twice a week, or	Once or twice a month (2)
	(4) Almost every day	Once or twice a week (3)
	C2a. In a typical month during the school year, how often (do/did) you or another family member go over or help this child with (his/her) homework? Would you say…	Almost every day (4)
	(1) Seldom or never	*For C3:*
	(2) Once or twice a month,	Never (1)
	(3) Once or twice a week, or	Once or twice during the school year (2)
	(4) Almost every day	More than twice during the school year (3)
	C3. During the 2010–11 school year, how many times (do/did) you or another adult family member:	
	(a) Attend school activities?	
	(b) Contact the principal, teacher or other staff member at (his/her) school regarding (his/her) academic performance?	
	(c) Volunteer at (his/her) school?	

Table B.2 *(continued)*

Outcome/Category	Student and Parent Survey Items Included	Scale/Definition
Index indicating school is too easy	B4. Do you think the homework (is/was) too difficult, about right, or too easy for (STUDENT)? (1) Too difficult (2) About right (3) Too easy B5a. Do you think the material covered in (his/her) math class (is/was) too difficult, about right, or too easy for (STUDENT NAME)? (1) Too difficult (2) About right (3) Too easy B6a. Do you think the material covered in (his/her) English/language arts class (is/was) too difficult, about right, or too easy for (STUDENT NAME)? (1) Too difficult (2) About right (3) Too easy	Mean across all items using the following scale: Too easy (1) other responses (0)
Index indicating school is too difficult	See items B4, B5a, and B6a	Mean across all items using the following scale: Too difficult (1) other responses (0)

Note: The student survey was administered to cohort 1 in 2009-2010 and cohort 2 in 2010-2011. For students in cohort 1, the directions referring to school year 2010-2011 referred to school year 2009-2010.

APPENDIX C

SCHOOLS ATTENDED BY LOTTERY WINNERS AND LOTTERY NON-WINNERS

Since any impact measured in the lottery-based study should stem from differences in students' school experiences, it is important to understand the nature and extent of those differences. In this analysis, we explore the characteristics of schools attended by lottery winners and those attended by lottery non-winners in the spring of the second follow-up year (2009-10 for cohort 1, and 2010-11 for cohort 2). The school characteristics are weighted by the number of students in the sample attending the school. The data for these comparisons comes from the National Center for Education Statistics' Common Core of Data (CCD). We identified the school attended by students from parent surveys administered in year 2, supplemented with school records data from the same timeframe.

The schools attended by lottery winners—most commonly the KIPP school to which the student applied—differed significantly on a range of characteristics from those attended by non-winners (Table C.1). However, it is important to note that not all lottery winners actually attended a KIPP school in the lottery sample; conversely, some non-winners did attend KIPP schools in the lottery sample. By the spring of year 2, 63 percent of lottery winners were currently enrolled in KIPP schools, and 11 percent of non-winners were attending a KIPP lottery school (72 percent of lottery winners and 12 percent of non-winners had ever enrolled in a KIPP lottery school). Thus, any differences in outcomes when KIPP lottery winners are compared to non-winners may understate the effect of actually attending a KIPP school in the lottery sample; this also means that any differences in the characteristics of the schools attended by lottery winners and non-winners (described below) does not simply reflect differences between KIPP schools and nearby non-KIPP schools.

Lottery winners attended smaller schools than non-winners. Lottery winners attended schools with an average enrollment of 504 students, compared to 819 students in schools attended by non-winners. Enrollment per grade (calculated by dividing total enrollment at the school by the number of grades with students enrolled) was also smaller in lottery winners' schools (139 compared to 268), although student-teacher ratios were similar in both sets of schools.

Students at schools attended by lottery winners were less likely to be white or Hispanic; both sets of schools are attended by students who are predominantly minority and low-income. At schools attended by lottery winners, 35 percent of students are Hispanic, on average, and 6 percent are white. In contrast, students at schools attended by lottery non-winners are 50 percent Hispanic and 10 percent white. The proportion of students who were black is statistically similar in schools attended by lottery winners and non-winners. Both groups of schools have similar percentages of students receiving free or reduced price lunch.

There are many important differences between the schools attended by lottery winners and those attended by non-winners. Taken together, these findings suggest that students who were offered admission to a KIPP school in the lottery sample had different school experiences than those who entered the lotteries but did not receive an admission offer. These differences may help explain any impacts associated with being offered admission to a KIPP school.

Table C.1. Characteristics of Schools Attended by KIPP Lottery Winners and Non-Winners

	Schools Attended by Lottery Winners	Schools Attended by Lottery Non-Winners	Difference in Means	P-Value
Enrollment (Means)				
Total enrollment	503.8	819.4	-315.6	0.000**
Enrollment per grade	139.0	267.8	-128.8	0.000**
Student-teacher ratio	17.6	16.3	1.3	0.389
School Type (Percentages)				
Ever attended a KIPP School (lottery sample)	71.6	12.0	59.6	0.000**
School type in year 2				
KIPP school (lottery sample)	63.4	10.8	52.6	0.000**
KIPP school (not in lottery sample)	0.5	3.2	-2.7	0.222
Non-KIPP charter school	6.3	20.4	-14.1	0.001**
Traditional public school	28.4	61.8	-33.3	0.000**
Private school	1.4	3.8	-2.4	0.043*
Characteristics of Students at School (Mean Percentage)				
Hispanic	35.1	49.8	-14.7	0.036*
White	5.8	10.3	-4.5	0.010*
Black	45.3	32.0	13.3	0.141
Receive free- or reduced-price lunches	75.0	76.2	-1.2	0.704
School-Wide Title I (Percentage)	**96.5**	**91.3**	**5.3**	**0.071**
Number of Schools in Sample	**429**	**500**		

Note: NCES data reflects the 2009-2010 school year.

*Difference between KIPP lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
**Difference between KIPP lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

APPENDIX D

ANALYTIC METHODS FOR THE MATCHED COMPARISON GROUP ANALYSIS

A. Impact Model and Covariates

To make the analysis of state test scores comparable across states and districts, all raw test scores were converted to z-scores defined relative to the distribution of scores in each grade, year, subject, and jurisdiction. That is, for each jurisdiction associated with a given KIPP school, we calculated the difference between each student's raw score and the mean score in that grade, year, and subject, and then divided the difference by the standard deviation of raw scores in the jurisdiction in that grade, year, and subject. Thus, each z-score reflects the number of standard deviations above or below the mean for the relevant cohort and jurisdiction.[61]

As explained in Appendix A, the first step in our matching-based impact estimation approach was to obtain a matched comparison group with characteristics that resemble the study's sample of KIPP students. To obtain impact estimates using this matched sample, we estimated an ordinary least squares (OLS) regression model that considered all math and reading test score data from grades 5–8 to measure students' outcome test scores and incorporated baseline (4th grade) demographic controls including indicators for gender, race/ethnicity, free/reduced-price lunch status, special education status, grade retention in a baseline year, and limited English proficiency status; cohort (year by entry grade); outcome test grade level; and two years of baseline mathematics and reading test scores (3rd and 4th grade for cohorts entering KIPP in grade 5; 4th and 5th grade for cohorts entering KIPP in grade 6). See Table D.1 for a full list of these covariates. The basic form of the model for each school is defined in equation D1:

$$\text{(D1)}\ y_{it} = \alpha + X_i\beta + \delta_1 T1_{it} + \delta_2 T2_{it} + \delta_3 T3_{it} + \delta_4 T4_{it} + grade\ dummies + cohort\ dummies + \varepsilon_{it}$$

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and two years of baseline test scores) of student i; $T1_{it}$ through $T4_{it}$ are binary variables for treatment status in up to four years,[62] indicating whether student i had first enrolled at KIPP one, two, three, or four years previously, as of school year t. For example $T3_{it}$ would be equal to 1 for student i at time t if the student had first enrolled at KIPP at time (t-3), regardless of whether the student was still enrolled at KIPP at time t; otherwise, $T3$ would be equal to 0. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ_1, δ_2, δ_3, δ_4, and β are parameters or vectors of parameters to be estimated. As the estimated coefficient on the set of treatment indicators, δn represents the cumulative impact of n years of KIPP treatment. Robust standard errors were clustered at the student level since individual students could contribute up to four observations to the analysis sample.

We used the model to separately estimate the impact of each KIPP middle school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP middle schools in the

[61] By definition, the distribution of student z-scores has a mean of 0 and standard deviation of 1 for each subject (math, reading, science, and social studies) in each of the four outcome years examined in the matching analysis.

[62] Due to a combination of data availability and the year when the KIPP school opened, at three KIPP schools treatment students in the sample received no more than two years of KIPP treatment; at an additional four schools, students received no more than three years of treatment.

sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

Table D.1. List of Covariates Included in OLS Model

Included Covariate
Math baseline test score from 1 year prior
Math baseline test score from 2 years prior
Reading baseline test score from 1 year prior
Reading baseline test score from 2 years prior
Gender indicator variable
Set of race/ethnicity indicator variables
Special education status indicator variable
Free or reduced price lunch status indicator variable
Limited English proficiency status indicator variable
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from 1 and 2 years prior are imputed using method described in Appendix E, Section B
Dummy variables indicating whether student repeated grades in either of the two baseline years
Dummy variables for grades 5-8
Dummy variables for each student cohort in the sample

Note: Baseline test scores were imputed when missing. In some jurisdictions data was not available on special education status, free or reduced price lunch status, or limited English proficiency status. For more details on the data provided by each jurisdiction, see Appendix A.

We also investigated whether the study's impact estimates were robust to alternative specifications of the impact model in equation D1. We found that the overall impact estimates remain consistently positive and statistically significant regardless of whether or not pre-baseline test scores or opposite subject test scores were included in the model. The results also remain consistent when the model includes dummy variables representing the school each student attended in 4th grade. In addition, we tested whether the average impact estimates were robust to an alternative weighting approach that weights each school-specific impact estimate by the number of students in the sample (this alternative approach gives the greatest weight to the schools that were open for the longest period of time in our data). Results using these alternative weights are shown at the conclusion of this appendix in Tables D.6 and D.7 (see model 1a). As shown in the tables, results that use these alternative weights are very similar to average effect estimates that assign an equal weight to each school; under both approaches, impacts remain statistically significant and positive in reading and math for all outcome years, and the difference in impact estimates from the two weighting methods consistently falls within 0.05 standard deviations.

Finally, we tested whether the impact estimates may have been affected by details of the procedure we used to match comparison group students to KIPP students, which was based on nearest neighbor matching without replacement (see Appendix A). We found that the impact estimates were not dependent on whether matching was conducted with or without replacement, varying by no more than 0.02 standard deviation units. Separately, we also estimated impacts using a propensity score matching approach that used caliper matching—a procedure that constructs a

comparison group by identifying all comparison students with propensity scores that fall within a given range of the propensity score of each KIPP student.[63] Results using this alternative caliper matching approach are shown in Tables D.6 and D.7 (see model 1b). The caliper matching results are very similar to the average effect estimates based on nearest neighbor matching; under both approaches, impacts remain statistically significant and positive in both subjects for all outcome years, and the difference in impact estimates from the two matching methods consistently falls within 0.02 standard deviations.

B. Empirical Bayes Estimates of School-Level Impacts

In addition to estimating the average impact of KIPP in each test subject and outcome year, we also produced empirical Bayes shrinkage estimates of the impact of each KIPP school in the sample. We did this to estimate the distribution of impacts across KIPP schools in a way that would not be exaggerated by the sampling variability of individual schools' impact estimates. Because each school's impact is estimated with some sampling error, some schools' impact estimates will end up being larger than their true impacts and others will be smaller. As a result, the distribution of these impact estimates is likely to show more variation than the distribution of the schools' true impacts. In other words, the imprecision or "noise" in the impact estimates makes it more likely that we would overstate the magnitude of the differences between the highest-performing KIPP schools and the lowest-performing KIPP schools. The empirical Bayes shrinkage estimates are designed to produce a distribution of impact estimates that is closer to the distribution of true impacts.

As discussed in Chapter IV, the sample size for each school-level impact estimates can vary for multiple reasons. Most importantly, the sample size in each school is largely determined by the number of student cohorts in the sample (that is, the number of years the school has been in operation in our data). The total sample size for each included school ranges from 95 KIPP students to 787 KIPP students, meaning that we are able to estimate the impacts of some KIPP schools more precisely than others. The standard error of the unadjusted school-level impact estimates ranges from 0.15 to 0.02 standard deviations depending on the school's sample size, test subject, and outcome year.

To adjust for different levels of precision in the impact estimates, we produced empirical Bayes shrinkage estimates of the school-level impacts, following the approach described in Morris (1983). This adjustment corrects for statistical "noise" by shifting each school's impact estimate closer to the average KIPP impact in the relevant test subject and outcome year. Schools with less precise estimates receive the largest adjustments, and schools with more precise estimates receive smaller adjustments.

For each KIPP school, we began by calculating a set of "reliability weights" that correspond to the precision of the school's estimates, with a separate weight for each test subject and outcome

[63] For this sensitivity test, we conducted caliper matching with a radius of 0.001. In other words, each treatment student was matched to all comparison students (across all comparison cohorts) that had propensity scores falling within 0.001 of that treatment student's propensity score. Matching was conducted with replacement, and students in the comparison group were weighted according to the number of times they were matched to treatment students. The match rates for the 41 KIPP schools in the sample were between 83 percent and 100 percent; the average match rate was 96 percent. There were no statistically significant differences between the baseline math and reading scores of the treatment group (-0.13 and -0.10, respectively) and the resulting matched comparison group (-0.14 and -0.11).

year. For a given subject and outcome year, each school's reliability weight was defined by the following (simplified) equation:

(D2a) $W_k = \delta var \ / \ (\delta var + \sigma_k^2 \ )$

Where W_k is the reliability weight for school k, δvar represents the variance of unadjusted impact point-estimates across all KIPP schools in the sample (that is, the variance of the schools' set of treatment-indicator coefficients for the relevant subject and outcome year, as calculated in equation D1), and σ_k^2 represents the variance of the single impact estimate for school k. Thus, the reliability weight W_k would be closer to 1 for a school with an impact estimate that has a small standard error, and W_k would be closer to 0 for a school with an impact estimate that has a large standard error. Using these weights, we applied the following equation to produce the adjusted impact estimate for each school (again, the adjustments were calculated separately for each test subject and outcome year):

(D2b) $\delta_{k,EB} = \ W_k \delta_k + (1 - W_k)\delta mean$

Where $\delta_{k,EB}$ is the empirical Bayes impact estimate for school k, W_k is the reliability weight from equation D2a, δ_k is the unadjusted impact estimate for school k from equation D1, and $\delta mean$ is an equally-weighted average of the unadjusted impacts of all KIPP schools in the sample (that is, our estimate of the overall average KIPP impact for the relevant subject and outcome year). This equation "shrinks" each school's impact estimate by moving it closer to the average KIPP impact, with the greatest amount of shrinkage occurring for the schools with low reliability weights and a smaller amount of shrinkage occurring for the schools with high reliability weights.

In our sample, the shrinkage of school-level impact estimates ranged from a movement of 0.01 standard deviation units to a movement of 0.12 standard deviation units. Since all estimates moved toward the mean, in some cases the empirical Bayes impact estimate for a school was more positive than the original (unadjusted) impact estimate, and in others the empirical Bayes estimate was more negative than the original result. Because the sample sizes in our analysis tended to be large, the typical school's empirical Bayes impact estimate was close to the unadjusted estimate. Across our sample of schools and outcomes, the mean absolute value of the difference between the empirical Bayes estimate and the original estimate was 0.02 standard deviation units.

For more information on the exact procedures used to calculate the empirical Bayes shrinkage impact estimates and the standard errors of the adjusted school-level impacts, see Morris (1983).

C. Imputation for Missing Baseline Data and Retained Students

This section explains in greater detail how our analysis handled two types of missing data: (1) students missing data on one of their test scores either one year before a KIPP entry grade or two years before a KIPP entry grade; or (2) students who were retained in grade, and therefore are missing a test score on the outcome test(s) given to the remaining cohort.

1. Imputation for Missing Baseline Data

Our benchmark analyses used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores; imputation was completed separately by treatment status. This imputation process involved estimating the following model:

(D3a) $$Yp_math_{it} = \alpha + X_i\beta + \sum_r \varphi_r Yr_math_{it} + \sum_{q=3}^{8} \gamma_q Yq_reading_{it} + \varepsilon_{it}$$

(D3b) $$Yp_reading_{it} = \alpha + X_i\beta + \sum_r \varphi_r Yr_reading_{it} + \sum_{q=3}^{8} \gamma_q Yq_math_{it} + \varepsilon_{it}$$

where Yp_math_{it} is a single grade *p* math baseline test score for student *i* at time *t;* $Yp_reading_{it}$ is a single grade *p* reading baseline test score for student *i* at time *t;* X_i is a vector of demographic characteristics (gender, race/ethnicity, special education status, free or reduced price lunch status and limited English proficiency status, where available) of student *i*; Yr_math_{it} and $Yr_reading_{it}$ are all available grades 3–8, excluding grade *p*, math and reading baseline or outcome test scores for student *i* at time *t*; and Yq_math_{it} and $Yq_reading_{it}$ are all available grades 3–8 math and reading baseline or outcome test scores for student *i* at time *t*. Note that the treatment dummies are not part of the imputation model because imputation is performed separately for the treatment group and then the comparison group.

We first estimated equations (D3a) and (D3b) for baseline test scores one and two years prior to KIPP entry using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores (in other words, values of the right hand side variables in equations D3a and D3b) and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student. We only imputed missing baseline test scores for students who have at least one non-missing baseline test score in either math or reading.

Finally, to obtain the imputed baseline test scores used in our benchmark model, we added a stochastic component to the predicted values of Yp_mathit and Yp_readingit obtained from estimating equations (D3a) and (D3b) above. For each student, the stochastic component is randomly selected from the set of all residuals estimated in equations (D3a) and (D3b) for the full sample. The stochastic component is included to ensure that the variance of the imputed baseline test scores is the same as that of the observed values.

To test whether our results are sensitive to this imputation strategy, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores (see model 2, in Tables D.6 and D.7). The results for this smaller sample are nearly identical to our benchmark impact estimates: again, the KIPP impact in both subjects remains statistically significant and positive in all outcome years, and the magnitude of the impact estimates is nearly identical to the benchmark estimates as well.

2. Imputation for Students Repeating a Grade

As discussed in Chapter II, we also impute the math and reading state test scores of students who repeat a grade if they were retained in one of the study's four outcome years. For example, if a student in the treatment group entered KIPP in grade 5 and then repeated grade 6, they would still be in grade 6 (and would take the grade 6 state assessment) at the end of the third follow-up year. Members of their cohort who remained on track would have taken the grade 7 state assessment. Because the grade repeater's grade 6 assessment score would not be comparable to grade 7 scores, we treat this student's year 3 follow-up score as missing and impute its value. To do so, we use the following approach in the math and reading analyses: for each grade repeater, in the year of repetition and subsequent years, we impute the student's z-score on the cohort-appropriate (rather

than grade-appropriate) test by setting his or her score equal to the student's standardized score in the last year prior to grade repetition. In this example, we would use the standardized score of the grade repeater on the grade 6 assessment in the second follow-up year (the score from the first time the student took that assessment). In effect, this imputation procedure assumes students maintain the same percentile rank relative to their cohort in the year of grade retention and in all subsequent years. In other words, we assume that each retained student does neither better or worse in relative terms than before retention. If KIPP in fact has a positive impact on retained students, this would cause us to underestimate KIPP's impact. Conversely, if KIPP has a negative impact, this would cause us to overestimate the impact.

This imputation procedure was not possible for the matching-based analysis of science and social studies test scores—these are often administered only once during middle school (usually in grade 8). For these two subjects, the outcome scores for each student were drawn from the highest available middle school grade, regardless of whether students were retained in prior years.

To test the sensitivity of our results to the method used for retained students, we estimated KIPP impacts using several alternative approaches to analyzing the test scores of retained students. We considered several alternative approaches to addressing the relatively high rate of grade repetition in KIPP, especially compared to the prevalence of grade repetition in comparison schools. In model 3 shown in Tables D.6 and D.7, we present results from an alternative, more conservative approach handling grade repeaters. For all years following the year grade repetition was first observed for a given student, we assigned the test score of a student to the fifth percentile of the jurisdiction analysis sample in the grades they would have attended under a "normal" grade progression.[64] Using this conservative approach, the KIPP impact estimates remain positive and statistically significant in all four outcome years for both reading and math. However, as we might expect, the magnitude of each statistically significant positive impact is somewhat smaller than under our benchmark approach (the estimates are between 0.02 and 0.06 standard deviations smaller in both math and reading).

In addition, we also estimate the impacts of KIPP using the recorded test scores of grade repeaters in all years, without any adjustments (model 4). In other words, within each student cohort this analysis compares the scores of retained students taking one test in a given year to the scores of non-retained students taking a different test (one grade level higher) in that year. Using the observed scores of retained students in all years, KIPP's impact remains positive and statistically significant in both subjects for all four outcome years. As expected, the benchmark impact estimates (model 1) fall between the conservative estimates in model 3 and the estimates that use this non-imputed approach in model 4.

C. Testing for Selection Bias: "Impacts" Prior to KIPP Enrollment

As discussed in Chapter III, the principal threat to the internal validity of our matching approach is the question of whether our model is affected by unobserved factors related to impacts. The logic of our matched comparison group design involves using past test scores to control for important differences in the characteristics of students who apply to and enroll in KIPP and those

[64] On average, students who repeat a grade tend to have test scores that are higher than the fifth percentile in the year before they were retained. For example, in two large urban school districts in our sample the average prior scores of grade repeaters were respectively at the 23rd and 15th percentile in math and the 25th and 19th percentile in reading.

who do not. If the design works as intended, then any differences in the test scores of KIPP and non-KIPP students in the middle school years (after the former group enrolls in KIPP) can be attributed to the effect of KIPP. However, it is possible that differences in the unobserved characteristics of the two groups could contribute to the differences in their middle school test scores; that is, these unobserved differences may lead to selection bias in our estimate of the effect of KIPP.

In this specification test, we used students' fourth grade (pre-KIPP) test scores as an outcome, and estimate a model that controls for their prior (3rd grade) test scores along with other baseline student characteristics. Because this outcome is measured in 4th grade, before KIPP could possibly have had a causal effect on student achievement, any KIPP/non-KIPP differences in scores that we observe must have been caused by differences in the unobserved characteristics of the two groups. In other words, these KIPP/non-KIPP differences in 4th grade scores would be evidence of selection bias in a design that relies on prior test scores to account for key student characteristics affecting later student achievement. Unlike the other analyses discussed until now in this appendix, this exploratory analysis did not use a matched comparison group: instead, the sample of comparison students included all students in the jurisdiction associated with each KIPP school in the sample.[65] For the falsification analysis, we estimated results for the following equation:

$$\text{(D4)} \qquad y4_i = \alpha + X_i\beta + y3math_i + y3reading_i + y3repeat_i + \delta T_i + \varepsilon_i \,,$$

where $y3math_i$, $y3reading_i$ and $y4_i$ are respectively the 3rd grade mathematics and reading scores and 4th grade test scores (in either math or reading) for student i; α is the intercept term; X_i is a vector of demographic characteristics of student i; T_i indicates if student i ever enrolls in KIPP and $y3repeat_i$ indicates whether student i repeated third grade. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; and δ is the parameter of interest. Robust standard errors were clustered at the 4th grade school level.

Unless there is selection bias related to an unobserved factor associated both with KIPP attendance and grade 4 test scores, we should not find a significant KIPP effect in the year prior to enrollment. This is in fact what we observed: on average KIPP does not have any spurious impacts on baseline year test scores. As shown in Table D.2, the sample of all KIPP middle schools in our data files (n=46) does not have a statistically significant prior-year "effect" on KIPP students in reading or math, on average (with an average 4th grade "impact" estimate of 0.01 in math and 0.00 in reading). This suggests that our matching-based impact estimates are not meaningfully biased by factors that cannot be observed in our data, such as student motivation or parental characteristics. This is the case even though the falsification model specification described above could only include a single year of pretest scores (grade 3 scores) whereas our primary estimation model includes two years of pretest scores.

[65] Note, however, that our impact estimation methods are more sophisticated than the simple regression model in this test for selection bias. Among other differences, our impact estimates control for two years of prior test scores rather than one year, and the main estimates also use a sophisticated matching process to identify a comparison group rather than applying the simple regression approach used in this specification test.

Table D.2. Test for Selection Effects Prior to KIPP Enrollment

	Math	Reading
Falsification Impact	0.010 (0.008)	-0.002 (0.008)
Number of KIPP Schools	46	46

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

For most individual KIPP schools in the sample, we see no evidence of a KIPP "effect" prior to entry. In total, the falsification effect estimate was not statistically significant for 34 schools in reading and 33 schools in math. But the falsification results were significantly positive for 7 schools in reading and math, and significantly negative for 5 schools in reading and 6 schools in math. Thus, for some individual KIPP schools it may be important to control for more than just one year of prior scores to help ensure the KIPP group does not have important unobserved differences relative to the comparison group.[66]

D. KIPP impact estimates for student subgroups

In this section we present in detail the estimates derived to identify whether KIPP had differential impacts on particular subgroups of students. In general, our strategy to identify potential subgroup differences was to use interaction terms consisting of treatment indicators multiplied by subgroup variables. The coefficients on the interaction terms represent the marginal effect of KIPP for students in the specific subgroup above and beyond the average KIPP effect among other students. The statistical significance of the interaction term indicates whether the KIPP effect is different for the subgroup in question than for other KIPP students.

At the conclusion of this chapter, Tables D.3 and D.4 show whether there are statistically significant differences in a school's impact on math and reading achievement for students with different characteristics. In other words, the results described in the tables show whether there is a significant difference between KIPP's average impact among members of the listed subgroup and the impact among those who are not members of the subgroup. A positive and significant interaction indicates that KIPP's average impact is higher for the listed subgroup relative to all other KIPP students. Each subgroup analysis only included KIPP schools in which more than five percent of its students were part of the subgroup of interest. Thus, as shown in these two tables, the sample of included KIPP schools varies depending on the subgroup being examined. To calculate the average of subgroup effect estimates at these schools, all of the included KIPP schools were weighted equally. (The overall findings in Tables D.3 and D.4 remain the same if the schools are instead weighted by the percentage of KIPP students in each subgroup.)

We found evidence that KIPP impacts tend to be significantly higher for Hispanics than non-Hispanics and also higher for students with lower levels of prior reading achievement than for students who were higher achieving at baseline.[67] For both subgroups, the interaction effect estimate

[66] Tables showing these more detailed, school-level falsification test results are available from the authors upon request.

[67] For the baseline achievement interaction term, the baseline test score variable was interacted with treatment status; in this case, the negative coefficient indicates that KIPP impacts are highest for students who performed less well at baseline.

was statistically significant in at least three of the four middle school outcome years. In contrast, KIPP impacts do not differ in a majority of outcome years for students with any of the other characteristics we tested (limited English proficiency, special education, black students, or males). We also tested for race-gender interaction effects (not shown), and did not find consistent evidence that KIPP impacts differ for black males or Hispanic males.

Next, for the subgroup impacts that were statistically significant in more than one year, we examined the magnitudes of the effects. Specifically, we compared KIPP's impacts on Hispanics to impacts on non-Hispanics, and compared impacts on students with low baseline achievement to impacts on those with high baseline achievement. Table D.5 presents the results. For the sample of KIPP schools in the subgroup analysis, both Hispanics and non-Hispanics received consistently positive and statistically significant KIPP impacts, but the impacts among Hispanics are somewhat larger. For example, after two years Hispanic students received an average KIPP effect of 0.37 standard deviations in math and 0.19 standard deviations in reading; non-Hispanics received effects of 0.25 standard deviations in math and 0.09 standard deviations in reading. We also analyzed effects on KIPP students whose prior reading or math scores were half a standard deviation lower or higher than the average baseline score among all KIPP students in the sample.[68] While both groups experienced consistently positive and statistically significant impacts, KIPP students with lower prior achievement tended to receive larger effects in both subjects. The pattern is most consistent for reading test scores: after two years at KIPP, students with low baseline reading scores received an impact of 0.17 standard deviations on the reading exam; in contrast, students with high baseline scores received an impact of 0.12 standard deviations in reading after two years (impacts for both groups are statistically significant).

E. Alternative Model Specifications

Below, we describe two additional sets of results obtained from estimating KIPP impacts using alternative methods.

1. Effects on Students who Remain Enrolled at KIPP

Our benchmark approach includes any student in the treatment group who attended a KIPP school in grades 5 or 6, regardless of how many years he or she stayed enrolled subsequently. Because the sample of treatment students includes observations from those who were not enrolled at KIPP in some years, this approach likely underestimates the true impact of KIPP on students who actually attended in each year. Here we apply an alternative approach to explore the extent to which our benchmark estimates may be underestimated. Under this alternative approach, we calculate attrition adjusted estimates (AAE) by modifying the benchmark estimates in a way that accounts for the fact that not all treatment group students received the treatment (that is, attended KIPP) for the full follow-up period.[69]

[68] Specifically, we estimated the average KIPP impact associated with having a prior score that is 0.5 z-score units below or above the average baseline z-score for all KIPP students in reading or math.

[69] This approach is analogous to the "treatment on treated" Bloom adjustment used in the experimental analysis. We refrain from using the treatment on treated language in the context of the quasi-experimental analysis for the sake of accuracy, since all students in the treatment group attended KIPP for some length of time.

We obtained these AAE by re-estimating our benchmark model to obtain the marginal benchmark impact estimates (BE) of each additional year in a KIPP middle school. We then adjusted these marginal BE by dividing them by an adjustment factor, p, which is equal to the proportion of the treatment group currently enrolled at KIPP in year t.

(D4) $\textit{Marginal AAE} = \textit{Marginal BE}/p$

To obtain the alternative cumulative impact estimates for each of the four years, we used the following set of equations:

(D5a) *Cumulative AAE, year 1 = Marginal AAE, year 1*
(D5b) *Cumulative AAE, year 2 = Cumulative AAE, year 1 + Marginal AAE, year 2*
(D5c) *Cumulative AAE, year 3 = Cumulative AAE, year 2 + Marginal AAE, year 3*
(D5d) *Cumulative AAE, year 4 = Cumulative AAE, year 3 + Marginal AAE, year 4*

It should be noted that this procedure makes a strong assumption—that students who withdraw from KIPP schools experience no continuing effect of their prior enrollment at KIPP. If this is not true and KIPP does exert a continuing positive impact on these students' achievement, then the adjusted estimates will overestimate KIPP's full effect on students who remain enrolled. These attrition-adjusted estimates in math and reading are presented in Tables D.6 and D.7 (model 5). By definition, the number of statistically significant estimates does not change from our benchmark approach, but the magnitude of the impacts is between 0.03 and 0.06 standard deviations larger than our benchmark results (model 1) in later outcome years.

2. Districtwide Comparison Group

Our final set of alternative impact estimates present results that use the entire district as a comparison group. In other words, the comparison group is formed without propensity-score matching but the regression model in equation D1 is still used to control for baseline characteristics of KIPP students and comparison group students. For the 41 schools with matching-based impact estimates, using a district-wide comparison group produces impact estimates (model 6 in Tables D.6 and D.7) that are very similar to the benchmark results—in both reading and math, the impact estimates are positive and statistically significant in all four outcome years and the magnitude of each point-estimate is nearly identical to our benchmark results.[70]

Separately, in model 7 (also in Tables D.6 and D.7) we used this district-wide comparison group method to estimate impacts for five additional KIPP middle schools that could not be included in the matching estimates because we only received data for a single cohort of students. Because these five schools are newly opened, we could only estimate impacts after one year for this sample. For these five schools, KIPP's impact is statistically significant and positive in both reading and math—

[70] Although these district-wide comparison group estimates are very close to our matching results, there is a potential drawback to comparing KIPP students to all students in the relevant public school district. Under such an approach, the sample of comparison students may include individuals who are very different at baseline from the students who enroll in KIPP schools. OLS models adjust for these differences, but the adjustments depend on assumptions about the underlying relationship between each characteristic and the achievement results. Impact estimates that use a matched comparison group help to avoid relying on these assumptions, which is why our preferred matching-based impact estimates rely on propensity-score matching. This ensures the treatment and comparison groups share similar demographic characteristics and prior achievement trajectories.

the magnitude of the impact estimates is slightly lower in math but higher in reading than one-year impacts for the other 41 schools in our benchmark analysis.

Model 8 presents the district-wide comparison group analysis for the largest possible sample of schools: combining the 41 schools in our benchmark analysis with the 5 additional schools in model 7. As shown in Tables D.6 and D.7, our results for this larger 46-school sample are also very similar to the study's benchmark results.

Table D.3. Comparison of KIPP Effects on Subgroups to Effects on Other KIPP Students, Mathematics

	KIPP Impact on Student Subgroups, Compared to Other KIPP Students			
Subgroup	Year 1	Year 2	Year 3	Year 4
Black	Not Different [23]	Smaller [23]	Smaller [22]	Not Different [16]
Hispanic	Larger [21]	Larger [21]	Larger [18]	Larger [9]
Male	Not Different [39]	Smaller [39]	Not Different [37]	Not Different [27]
Special education	Not Different [30]	Not Different [27]	Not Different [20]	Not Different [13]
Limited English proficiency	Not Different [13]	Not Different [12]	Not Different [11]	Not Different [5]
Higher baseline math scores	Smaller [41]	Smaller [41]	Not Different [38]	Not Different [28]

Note: The number of KIPP schools included in the analysis is indicated in brackets. Table rows describe the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A "larger" label indicates that KIPP's average impact is higher for the examined subgroup by a statistically significant margin ($p < 0.05$). A "smaller" label indicates that the average impact is lower by a statistically significant margin for the examined subgroup. To analyze baseline scores, the baseline test score was interacted with treatment status: in this case, a "smaller" result signals that KIPP impacts are highest for students who performed less well at baseline.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

Table D.4. Comparison of KIPP Effects on Subgroups to Effects on Other KIPP Students, Reading

	KIPP Impact for Student Subgroups, Compared to other KIPP Students			
Subgroup	Year 1	Year 2	Year 3	Year 4
Black	Not Different [23]	Smaller [23]	Not Different [22]	Not Different [21]
Hispanic	Not Different [21]	Larger [21]	Larger [18]	Larger [15]
Male	Larger [39]	Not Different [39]	Not Different [37]	Not Different [33]
Special education	Not Different [31]	Not Different [28]	Not Different [21]	Not Different [16]
Limited English proficiency	Smaller [13]	Not Different [12]	Not Different [11]	Not Different [10]
Higher baseline reading scores	Smaller [41]	Smaller [41]	Smaller [38]	Smaller [34]

Note: The number of KIPP schools included in the analysis is indicated in brackets. Table rows describe the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A “larger” label indicates that KIPP's average impact is higher for the examined subgroup by a statistically significant margin ($p < 0.05$). A “smaller” label indicates that the average impact is lower by a statistically significant margin for the examined subgroup. To analyze baseline scores, the baseline test score was interacted with treatment status: in this case, a “smaller” result signals that KIPP impacts are highest for students who performed less well at baseline.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

Table D.5. KIPP Effects on Hispanics and Students with Low Prior Test Scores

Subgroup	Year 1	Year 2	Year 3	Year 4
	Average KIPP Effect on Mathematics Test Scores			
Hispanics	0.20**	0.37**	0.43**	0.34**
(standard error)	(0.04)	(0.04)	(0.04)	(0.06)
Non-Hispanics	0.14**	0.25**	0.29**	0.25**
(standard error)	(0.02)	(0.02)	(0.02)	(0.04)
Difference in Effects	0.06*	0.11**	0.14**	0.09*
(standard error)	(0.03)	(0.03)	(0.04)	(0.05)
	[21]	[21]	[18]	[9]
Low Prior Math	0.17**	0.30**	0.36**	0.32**
(standard error)	(0.01)	(0.01)	(0.02)	(0.02)
High Prior Math	0.14**	0.26**	0.35**	0.30**
(standard error)	(0.01)	(0.01)	(0.02)	(0.02)
Difference in Effects	0.03**	0.04**	0.01	0.02
(standard error)	(0.01)	(0.01)	(0.01)	(0.02)
	[41]	[41]	[38]	[28]
	Average KIPP Effect on Reading Test Scores			
Hispanics	0.07	0.19**	0.27**	0.32**
(standard error)	(0.04)	(0.03)	(0.04)	(0.05)
Non-Hispanics	0.01	0.09**	0.16**	0.14**
(standard error)	(0.02)	(0.02)	(0.02)	(0.03)
Difference in Effects	0.06	0.10**	0.11**	0.18**
(standard error)	(0.03)	(0.03)	(0.03)	(0.04)
	[21]	[21]	[18]	[15]
Low Prior Reading	0.07**	0.17**	0.23**	0.27**
(standard error)	(0.01)	(0.01)	(0.02)	(0.02)
High Prior Reading	0.04**	0.12**	0.19**	0.19**
(standard error)	(0.01)	(0.01)	(0.02)	(0.02)
Difference in Effects	0.04**	0.05**	0.04**	0.08**
(standard error)	(0.01)	(0.01)	(0.01)	(0.02)
	[41]	[41]	[38]	[34]

Note: The number of KIPP schools included in the analysis is indicated in brackets. Regressions were performed separately for each KIPP middle school in the sample. Reported effect sizes are an average of equally-weighted impact estimates from regressions of middle school math and reading z-scores on indicator variables for the number of years after a student's enrollment in a KIPP middle school and covariates. Subgroup-specific impacts were obtained by summing the main effect estimate with an interaction term between the subgroup variable and the KIPP treatment indicator. The effect estimates for students with "low" or "high" prior scores reflect the average KIPP impact on students with baseline scores that are (respectively) 0.5 z-score units below or above the mean baseline test score for all KIPP students. Regressions use robust standard errors and are clustered on student identifiers.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

Table D.6. Comparison of Benchmark Impact Model and Alternative Models, Mathematics

Model	Year 1	Year 2	Year 3	Year 4
1. Benchmark model, schools weighted equally	0.15** (0.01)	0.27** (0.01)	0.36** (0.01)	0.31** (0.02)
1a. Benchmark model, schools weighted by sample size	0.18** (0.01)	0.32** (0.01)	0.39** (0.01)	0.36** (0.02)
1b. Benchmark Model, caliper matching	0.15** (0.01)	0.27** (0.01)	0.36** (0.01)	0.32** (0.01)
	[41]	[41]	[38]	[28]
Alternative Approaches to Imputing Data				
2. Non-imputed baseline data	0.15** (0.01)	0.28** (0.01)	0.35** (0.02)	0.30** (0.02)
	[41]	[41]	[37]	[28]
3. Conservative approach to grade repeater scores	0.13** (0.01)	0.24** (0.01)	0.30** (0.01)	0.25** (0.02)
	[41]	[41]	[38]	[28]
4. Non-imputed grade repeater scores	0.16** (0.01)	0.31** (0.01)	0.40** (0.01)	0.32** (0.02)
	[41]	[41]	[38]	[28]
Adjusted Estimates Reflecting Impact of KIPP Attendance				
5. Attrition-adjusted estimates	0.15**	0.29**	0.41**	0.37**
	[41]	[41]	[38]	[28]
Districtwide Comparison Group Without Matching				
6. Benchmark KIPP sample	0.15** (0.01)	0.28** (0.01)	0.36** (0.01)	0.32** (0.01)
	[41]	[41]	[38]	[28]
7. New schools not in benchmark KIPP sample	0.12** (0.03)	NA	NA	NA
	[5]	NA	NA	NA
8. All KIPP students	0.16** (0.01)	0.29** (0.01)	0.36** (0.01)	0.32** (0.01)
	[46]	[42]	[39]	[28]

Note: The number of KIPP schools included in the analysis is indicated in brackets. Each row shows KIPP impact estimates under different analytical approaches and assumptions, with standard errors in parentheses. Models 1 through 5 use a matched comparison group; model 1b uses caliper matching, and all other matching estimates use the study's nearest-neighbor matching procedure. In model 2, after grade repetition students were assigned to the fifth percentile z-score for their cohort in each outcome year; model 3 uses the observed test scores of retained students; model 4 does not include imputed baseline test scores; model 5 adjusts the marginal yearly KIPP effect according to the number of early transfers from KIPP in the treatment sample, and derives cumulative impacts from these adjusted marginal effects (we do not show standard errors for these adjusted estimates, because the statistical significance of the result is derived from the benchmark analysis in model 1); model 6 includes all comparison students in local districts without matching, and models 7 and 8 use an unmatched approach to estimate impacts for an additional five newly opened KIPP middle schools.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

NA = not available

Table D.7. Comparison of Benchmark Impact Model and Alternative Models, Reading

Model	Year 1	Year 2	Year 3	Year 4
1. Benchmark model, schools weighted equally	0.05** (0.01)	0.14** (0.01)	0.21** (0.01)	0.22** (0.01)
1a. Benchmark model, schools weighted by sample size	0.05** (0.01)	0.16** (0.01)	0.21** (0.01)	0.23** (0.01)
1b. Benchmark Model, caliper matching	0.05** (0.01) [41]	0.14** (0.01) [41]	0.21** (0.01) [38]	0.22** (0.01) [34]
Alternative Approaches to Imputing Data				
2. Non-imputed baseline data	0.05** (0.01) [41]	0.14** (0.01) [41]	0.21** (0.02) [37]	0.22** (0.02) [34]
3. Conservative approach to grade repeater scores	0.03** (0.01) [41	0.10** (0.01) [41]	0.16** (0.01) [38]	0.16** (0.02) [34]
4. Non-imputed grade repeater scores	0.06** (0.01) [41]	0.17** (0.01) [41]	0.24** (0.01) [38]	0.25** (0.01) [34]
Adjusted Estimates Reflecting Impact of KIPP Attendance				
5. Attrition-adjusted estimates	0.05** [41]	0.15** [41]	0.24** [38]	0.25** [34]
Districtwide Comparison Group Without Matching				
6. Benchmark KIPP sample	0.05** (0.01) [41]	0.15** (0.01) [41]	0.21** (0.01) [38]	0.21** (0.01) [34]
7. New schools not in benchmark KIPP sample	0.12** (0.03) [5]	NA NA	NA NA	NA NA
8. All KIPP students	0.07** (0.01) [46]	0.15** (0.01) [42]	0.21** (0.01) [39]	0.21** (0.01) [34]

Note: The number of KIPP schools included in the analysis is indicated in brackets. Each row shows KIPP impact estimates under different analytical approaches and assumptions, with standard errors in parentheses. Models 1 through 5 use a matched comparison group; model 1b uses caliper matching, and all other matching estimates use the study's nearest-neighbor matching procedure. In model 2, after grade repetition students were assigned to the fifth percentile z-score for their cohort in each outcome year; model 3 uses the observed test scores of retained students; model 4 does not include imputed baseline test scores; model 5 adjusts the marginal yearly KIPP effect according to the number of early transfers from KIPP in the treatment sample, and derives cumulative impacts from these adjusted marginal effects (we do not show standard errors for these adjusted estimates, because the statistical significance of the result is derived from the benchmark analysis in model 1); model 6 includes all comparison students in local districts without matching, and models 7 and 8 use an unmatched approach to estimate impacts for an additional five newly opened KIPP middle schools.

* Statistically significant at the 0.05 level, two-tailed test.
** Statistically significant at the 0.01 level, two-tailed test.

NA = not available

APPENDIX E

ANALYTIC METHODS FOR LOTTERY-BASED ANALYSIS

This appendix presents additional detail about the analytic methods used in our lottery-based analysis of KIPP impacts. We describe the primary impact model, our approach to dealing with analytic issues, and the results of additional analyses that test the sensitivity of our impact estimates to alternative modeling assumptions.

A. Outcome Measures

Table E.1 presents summary statistics for outcome measures used in our lottery-based analysis. The outcomes are shown in the same order as they appear in the main text. For each measure, we show separately the mean, standard deviation and sample size for treatment and control groups. These statistics are unweighted, and therefore the treatment-control differences should not be interpreted as KIPP impacts.

B. Impact Model and Covariates

To obtain estimates of the impact of KIPP admissions for the subset of KIPP schools with lotteries we use the following model:

$$(1) \qquad y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * T + \gamma * X_i + \varepsilon_i$$

where i and k index students and schools, respectively, and y is the student-level outcome of interest. *SCHOOL* is a set of binary variables indicating the school that the student applied to, T is a binary treatment status variable indicating whether the student was offered admission to the school via the lottery, and X is a set of demographic and other controls. The βs represent site fixed effects, which capture differences in outcomes across sites that are not related to KIPP school attendance itself. These effects may capture variation across schools in the characteristics of KIPP applicants and/or the characteristics and performance of non-KIPP schools attended by control students. By including fixed effects in the model (as opposed to random effects), we acknowledge that KIPP schools were selected purposefully for the lottery-based analysis and that the results cannot be generalized beyond the study schools. The parameter δ represents the average impact of winning a KIPP middle school lottery; this is an intent-to-treat (ITT) estimate.

Our analysis includes student covariates to improve the precision of impact estimates, which include student baseline and pre-baseline test scores, student demographic characteristics, family income, and mother's education. The full set of covariates is presented in Appendix A, Table A.10. We also estimated models without covariates to test the sensitivity of our estimates to this modeling choice.

Table E.1. Unadjusted Means, Standard Deviations, Sample Sizes, and Reliability of Outcome Measures

Outcome	Lottery Winners			Lottery Non-Winners			Internal Consistency Reliabilty
	Unadj. Mean	Standard Deviation	Number of Observ.	Unadj. Mean	Standard Deviation	Number of Observ.	
State Assessments (Z-Score)							
Math achievement							
Year 1	0.10	0.86	202	0.01	0.86	334	n.a.
Year 2	0.30	0.80	181	0.07	0.90	260	n.a.
Reading achievement							
Year 1	0.03	0.83	202	0.05	0.80	333	n.a.
Year 2	0.17	0.84	181	0.10	0.90	260	n.a.
TerraNova Test Administered in the Fall of the Third Follow-Up Year (Z-Score)							
Math achievement	0.20	0.86	272	-0.06	1.00	317	n.a.
Reading achievement	0.10	0.95	272	0.00	1.01	318	n.a.
Student Motivation and Engagement							
Count of extracurricular activities	3.26	1.94	380	3.01	1.94	372	n.a.
Student reports having homework on a typical night (proportion)	0.97	0.18	380	0.96	0.19	372	n.a.
Minutes spent on homework on typical night, student report (mean)	111.80	70.57	365	94.11	64.15	357	n.a.
Minutes spent on homework on typical night, parent report (mean)	113.81	63.92	404	85.46	46.61	406	n.a.
Parent says student typically completes homework (proportion)	0.94	0.25	405	0.94	0.24	407	n.a.
Index of school engagement (mean)	3.66	0.37	379	3.64	0.40	372	0.633
Index of self control (mean)	4.42	0.66	378	4.44	0.66	372	0.841
Index of academic self-concept (mean)	3.26	0.37	380	3.20	0.37	373	0.747
Index of effort and persistence in school (mean)	3.47	0.43	380	3.50	0.42	374	0.839
Education Goals and Aspirations (Proportion)							
Student expects to graduate HS on time	0.96	0.19	375	0.96	0.19	371	n.a.
Parent expects student to graduate HS on time	0.97	0.17	418	0.95	0.21	420	n.a.
Student wishes to complete college	0.94	0.24	373	0.96	0.20	366	n.a.
Parent wishes student to complete college	0.99	0.11	418	0.99	0.12	424	n.a.
Student believes very likely to complete college	0.64	0.48	353	0.58	0.49	350	n.a.
Parent believes student very likely to complete college	0.72	0.45	408	0.65	0.48	418	n.a.
Student reports having discussions about college at school	0.78	0.40	376	0.79	0.40	368	n.a.
Student reports having discussions about college at home	0.92	0.28	375	0.92	0.27	368	n.a.
Parent reports having discussions about college	0.97	0.18	415	0.94	0.23	422	n.a.

Table E.1 *(continued)*

Outcome	Lottery Winners			Lottery Non-Winners			Internal Consistency Reliabilty
	Unadj. Mean	Standard Deviation	Number of Observ.	Unadj. Mean	Standard Deviation	Number of Observ.	
Student Behavior							
Index of peer pressure for bad behaviors (mean)	1.04	0.16	377	1.05	0.18	368	0.747
Index of undesirable behavior (mean)	2.31	0.45	377	2.33	0.47	368	0.591
Index of illegal action (mean)	2.97	0.12	377	2.97	0.14	368	0.577
Parent reported any school disciplinary problems for student (proportion)	0.36	0.48	410	0.34	0.48	413	n.a.
Index of parent-reported frequency of school disciplinary actions for student (mean)	0.22	0.42	415	0.19	0.34	420	0.699
Student never gets in trouble at school (proportion)	0.44	0.50	377	0.51	0.50	368	n.a.
Index of good behavior, student report (mean)	2.34	0.42	377	2.35	0.42	367	0.629
Index of good behavior, parent report (mean)	2.35	0.51	415	2.35	0.52	418	0.452
Index indicating well-adjusted student (mean)	3.45	0.46	416	3.45	0.44	420	0.854
Index of parental concerns about student (mean)	1.34	0.65	415	1.34	0.64	420	0.752
School Experiences and Satisfaction							
Index of student's feelings about school (mean)	3.42	0.40	380	3.33	0.42	374	0.848
Student likes school a lot (proportion)	0.52	0.50	380	0.56	0.50	373	n.a.
Index of parental satisfaction with school (mean)	3.26	0.61	418	3.16	0.69	424	0.832
Parent rates school as excellent (proportion)	0.52	0.50	417	0.40	0.49	423	n.a.
Index of student perceptions of schoolmates (mean)	2.85	0.49	380	2.82	0.47	373	0.781
Index of student perceptions of teachers (mean)	3.52	0.42	380	3.46	0.43	374	0.875
Index of school disciplinary environment (mean)	3.33	0.47	380	3.32	0.46	374	0.732
Index of parental perceptions of problems in student’s school (mean)	3.17	1.08	414	2.87	1.16	423	0.973
Index of parental involvement in student's education (mean)	2.82	0.44	419	2.69	0.47	429	0.563
Index indicating school is too easy (mean)	0.12	0.24	418	0.17	0.29	426	0.615
Index indicating school is too difficult (mean)	0.07	0.18	418	0.09	0.21	426	0.539

n.a. = not applicable

We also estimated a model that produces treatment-on-the-treated (TOT) estimates, which reflect the estimated impact of KIPP school attendance. To do so, we estimated an instrumental variables (IV) model in which the lottery outcome (treatment status) is an instrument for KIPP attendance. We used two-stage least squares to first estimate the effect of winning an admissions lottery on KIPP attendance (IV eqn 1), and in the second stage estimated the impact of KIPP attendance on outcomes (IV eqn 2). In effect, the TOT approach adjusts the ITT results to account for whether students actually attended a KIPP school.

(IV eqn 1) $$attendKIPP_i = \eta + \lambda * T + \upsilon_i$$

(IV eqn 2) $$y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * \widehat{attendKIPP}_i + \gamma * X_i + \varepsilon_i$$

Sensitivity to Inclusion of Baseline Covariates

Our primary model includes baseline covariates to improve the precision of our impact estimates. In addition, these covariates account for any differences between treatment and control group students in their baseline characteristics. The experimental design should ensure that there are no systematic differences in the baseline characteristics of treatment and control group students, but such differences may arise by chance. Table E.2 presents impacts from models that do and do not include baseline covariates. The first set of columns show the primary impact estimates and the second set shows estimates from models that do not make use of baseline covariates. Overall, the model without covariates produces findings that are qualitatively similar to our primary model that includes covariates. In particular, the magnitudes of the estimated impacts from the model without covariates tend to be similar to the magnitudes of the estimated impacts from the model with covariates. The levels of statistical significance from both models tend to be the same but occasionally differ. Out of 46 outcomes, the impacts across the two models varied in statistical significance in five cases. It is important to keep in mind, however, that the level of significance could vary because of the difference in the precision of this impact estimate, but not because of a difference in the magnitude of the estimated impact estimates.

C. Weighting

The impact model incorporates sample weights to account for the fact that not all students in the lottery have the same probability of being offered admission to the KIPP school (that is, being selected into the treatment group). Some students have a higher probability of being offered admission, either based on their inclusion in a particular stratum defined by a student characteristic or because they have a sibling in the lottery. If no sample weights were used and if these student characteristics were not otherwise accounted for in the impact model, then the characteristics of students in the treatment group and control group would differ on average, potentially leading to a bias in the impact estimate. For example, if KIPP schools tend to use sibling preference rules in their lotteries, then students with siblings will tend to be over-represented in the treatment group and students without siblings will be over-represented in the control group. If having siblings affects student performance directly or is correlated with some other student or family characteristic that is not accounted for, this could bias the impact estimate.

Table E.2. Sensitivity of Impact Estimates to Alternative Models

				Alternative Model					
	Primary Impact Model			Using Risk Set Binaries			With No Adjustment for Covariates		
Outcome	Effect Size	p-value		Effect Size	p-value		Effect Size	p-value	
Impacts on State Assessments (Z-Score)									
Math achievement									
Year 1	0.13	0.028	*	0.13	0.03	*	0.16	0.10	
Year 2	0.24	0.001	**	0.25	0.00	**	0.28	0.00	**
Reading achievement									
Year 1	0.02	0.775		0.02	0.80		0.03	0.80	
Year 2	0.10	0.217		0.10	0.20		0.16	0.09	
Impacts on the TerraNova Test Administered in the Fall of the Third Follow-Up Year (Z-Score)									
Math achievement	0.20	0.000	**	0.19	0.00	**	0.22	0.00	**
Reading achievement	0.08	0.246		0.04	0.52		0.09	0.33	
Impacts on Student Motivation and Engagement									
Count of extracurricular activities	0.06	0.480		0.06	0.45		0.21	0.03	*
Student reports having homework on a typical night (proportion)	-0.02	0.812		-0.01	0.89		-0.03	0.71	
Minutes spent on homework on typical night, student report (mean)	0.34	0.010	**	0.33	0.01	*	0.26	0.05	*
Minutes spent on homework on typical night, parent report (mean	0.69	0.000	**	0.69	0.00	**	0.62	0.00	**
Parent says student typically completes homework (proportion)	0.05	0.570		0.05	0.57		0.01	0.88	
Index of school engagement (mean)	0.01	0.881		0.01	0.91		0.05	0.51	
Index of self control (mean)	-0.06	0.448		-0.08	0.37		-0.07	0.38	
Index of academic self-concept (mean)	0.13	0.161		0.11	0.20		0.13	0.17	
Index of effort and persistence in school (mean)	-0.11	0.161		-0.11	0.16		-0.07	0.42	
Impacts on Education Goals and Aspirations (Proportion)									
Student expects to graduate HS on time	0.05	0.494		0.06	0.36		-0.03	0.68	
Parent expects student to graduate HS on time	0.05	0.464		0.05	0.51		0.02	0.71	
Student wishes to complete college	-0.12	0.137		-0.11	0.16		-0.14	0.08	
Parent wishes student to complete college	0.03	0.630		0.03	0.58		0.03	0.55	
Student believes very likely to complete college	-0.01	0.891		0.00	0.98		0.07	0.48	
Parent believes student very likely to complete college	0.03	0.753		0.03	0.73		0.06	0.53	
Student reports having discussions about college at school	0.06	0.568		0.07	0.47		0.11	0.33	
Student reports having discussions about college at home	-0.10	0.929		-0.01	0.95		-0.03	0.77	
Parent reports having discussions about college	0.00	0.977		0.00	0.97		0.08	0.30	

Table E.2 *(continued)*

Outcome	Primary Impact Model			Alternative Model: Using Risk Set Binaries			Alternative Model: With No Adjustment for Covariates		
	Effect Size	p-value		Effect Size	p-value		Effect Size	p-value	
Impacts on Student Behavior									
Index of peer pressure for bad behaviors (mean)	-0.03	0.704		-0.03	0.71		-0.02	0.77	
Index of undesirable behavior (mean)	-0.19	0.034	*	-0.19	0.03	*	-0.16	0.16	
Index of illegal action (mean)	-0.07	0.385		-0.07	0.41		-0.08	0.21	
Parent reported any school disciplinary problems for student (proportion)	-0.09	0.294		-0.08	0.36		-0.09	0.35	
Index of parent-reported frequency of school disciplinary actions for student (mean)	-0.04	0.661		-0.03	0.72		-0.01	0.95	
Student never gets in trouble at school (proportion)	-0.25	0.005	**	-0.26	0.00	**	-0.18	0.07	
Index of good behavior, student report (mean)	0.02	0.817		0.01	0.92		0.03	0.76	
Index of good behavior, parent report (mean)	-0.09	0.280		-0.09	0.29		-0.10	0.29	
Index indicating well-adjusted student (mean)	-0.03	0.695		-0.03	0.71		0.10	0.35	
Index of parental concerns about student (mean)	0.05	0.559		0.05	0.53		0.02	0.85	
Impacts on School Experiences and Satisfaction									
Index of student's feelings about school (mean)	0.21	0.012	*	0.19	0.02	*	0.19	0.02	*
Student likes school a lot (proportion)	-0.06	0.480		-0.07	0.47		-0.06	0.56	
Index of parental satisfaction with school (mean)	0.16	0.035	*	0.16	0.04	*	0.20	0.00	**
Parent rates school as excellent (proportion)	0.30	0.001	**	0.30	0.00	**	0.32	0.00	**
Index of student perceptions of schoolmates (mean)	0.11	0.232		0.09	0.32		0.14	0.12	
Index of student perceptions of teachers (mean)	0.13	0.147		0.13	0.16		0.09	0.36	
Index of school disciplinary environment (mean)	-0.01	0.908		-0.03	0.74		-0.03	0.79	
Index of parental perceptions of problems in student's school (mean)	0.02	0.751		0.02	0.79		0.04	0.62	
Index of parental involvement in student's education (mean)	0.12	0.074		0.13	0.05		0.15	0.04	*
Index indicating school is too easy (mean)	-0.21	0.005	**	-0.22	0.00	**	-0.16	0.05	*
Index indicating school is too difficult (mean)	0.04	0.555		0.05	0.51		0.04	0.54	

Notes: Effect size is calculated as the impact estimate divided by the standard deviation of the outcome for lottery non-winners.

* Difference between lottery winners and non-winners is statistically significant at the 0.05 level, two-tailed test.
** Difference between lottery winners and non-winners is statistically significant at the 0.01 level, two-tailed test.

The creation of the sample weights is based on the procedure used in Gleason et al. (2010). In the simple case, where all students interested in attending a particular KIPP school enter the lottery and no preferences are given for siblings or other characteristics, the sample weight for a given student is based upon the probability that he or she ended up in the experimental group (that is, treatment or control group). This probability is used in the calculation of each student's *base weight*. In particular, the base weight assigned to treatment group members is set to the inverse of the probability of being selected into the treatment group. The base weights for control group members are set to the inverse of the probability of being selected into the control group. We then normalize this weight to account for the fact that the sample will be representative of the set of all consenting lottery participants at that site. We set this normalization factor such that the weights of each experimental group sum to one-half of the total sample size within the site. Thus, the sum of all students' weights within a site will be equal to the overall sample size in that site (that is, the number of consenting lottery participants), with the sum of weights among treatments equal to that among controls.

In sites with sibling preference rules, the basic approach to calculating sample weights is the same as in the simple case above.[71] The difference, however, is in the calculation of the probability of admission. No longer can we simply use the number of students offered admission divided by the number of lottery participants. The exact probabilities of admission depend on the number of sets of siblings who participate in the lottery at the school as well as the number of students within each sibling set. With sibling preference rules, each sibling in the lottery has a higher probability of admissions than non-siblings, so the probabilities are adjusted to account for the number of siblings in each affected lottery.

An alternative to calculating sample weights that accounts for the probability of admission is to group students into "risk sets" that contain only students with the same probabilities of admission, and then control statistically for these risk sets in the impact models. In this approach, we only need to know which groups of students have the same chances of admission, and do not have to know the exact probability of admission for each student. As described below, we assessed the sensitivity of our main impact estimate models, which used sample weights, to estimating impacts using the risk set approach.

Testing Sensitivity to Sample Weighting

As described above, our main approach used sample weights to account for unequal probabilities of selection in the KIPP lotteries and to normalize treatment and control sample sizes within sites. We test the sensitivity of impacts to the alternate strategy of using a risk set approach to account for each lottery. To do so, we defined each separate lottery (typically one grade and cohort within a school) as a risk set and included these indicators as variables in the regression model; we did not use the sample weights. One site stratified their lottery by gender, so we included separate risk sets by gender. While several schools have sibling preferences, there are very few students in these preferred groups. We retained these students in their school-grade-cohort risk set and added an indicator variable for sibling status for students in schools that have sibling preferences. Table

[71] An example of sibling preference rules occurs when a school enters two siblings separately in an admissions lottery. If one of the two siblings is drawn as a lottery winner and offered admission to the school, the other sibling is pulled from the lottery pool and also offered admission.

E.2 presents our primary impacts and the impacts using the risk set approach. There are no appreciable changes to the estimates between these two specifications.

We also tested the sensitivity of our results to different normalization schemes for the sample weights. As described above, the base weights were normalized to equalize the contribution of the treatment and control groups within site, while the sum of weights in each site is the total sample size for that site. In addition, we tested three alternative normalizations that alter the relative weight of sites in the estimation—the treatment and control groups continue to contribute equally within site, but these normalizations affect the relative weight of different sites in the analysis. The first alternative normalizes the weights in each site to sum to the number of treatment observations in the site. The second normalizes the weights within each site to sum to the average sample size across sites, so each site contributes equally to the analysis. The third alternative normalizes the weights to sum to the school enrollment as reported in the CCD.

There were no substantive differences between the impacts estimated using the main weights and these alternatives for the outcomes on student achievement measured by state assessments or the TerraNova, and no differences for education goals and aspirations. There were some changes to the size or significance of impacts on some outcomes in the other domains. In particular, the impact estimates for several outcomes under student motivation and engagement (extracurricular activity index, parent reports that the student typically completes homework, and student report of academic self-concept) were larger by 0.12 to 0.30 effect size units when using alternative weights. In some cases the impacts become statistically significant although they were not statistically significant in the main results. The difference between the main sample weights and the alternative ones is how sites are weighted relative to one another. Thus, the change in estimated impacts implies that sites with larger impacts are being weighted more heavily when using the alternative normalizations. There are also changes to the estimated impacts for some outcomes under the domain of school experiences and satisfaction. The alternative weights result in larger impacts (by at least 0.13 effect size units) on whether a parent rates the school as excellent and the index of student perceptions of classmates. There are a number of other outcomes where the magnitude of the impact estimates does not change substantially, but the precision of the estimates does—these impacts are statistically significant in some models but not others.

D. Imputation

Our imputation procedure mirrors that used for the matching analysis described in Appendix D. One key difference is that for the lottery-based analysis we had additional data on participants gathered via the baseline survey and the study-administered test. In particular, we have baseline data on student age, household composition, language spoken at home, and parent reports of IEP status. We also included KIPP attendance, interactions of baseline test scores with KIPP attendance, and scores on the study-administered test to improve our predictions of missing baseline covariates. We used these variables in addition to baseline and pre-baseline state test scores and demographics from state records as covariates in the imputation to improve the prediction of missing values.

Grade Repeaters

In our main lottery-based analysis of state test scores, we excluded grade repeaters because they do not have the same grade progression as their peers and therefore do not have the same pretest-posttest relationship. This strategy is in contrast to the matching approach that "freezes" grade repeaters in the test score distribution. The other outcomes in the lottery-based analysis are not dependent on the grade level of students, so grade repeaters are included in these analyses and no adjustments were made. We tested the sensitivity of our impacts on state test scores to this approach by running the analysis using the alternate approach. When we retain grade repeaters in the sample but impute their outcome score, the impact estimates are unchanged.

APPENDIX F

VALIDATION OF MATCHING METHODS USING LOTTERY-BASED IMPACT ESTIMATES

This appendix summarizes results of the study's validation exercise. The analysis compares our lottery-based impact estimates to those based on the study's matched comparison group design at a subset of lottery sites where the data permit us to implement both approaches. This exercise allows us to determine whether there are likely to be sources of bias affecting the study's preferred matching-based impact estimates for a much larger group of KIPP middle schools (that is, the impact estimates for 41 middle schools presented in Chapter IV). As shown at the conclusion of this appendix in Table F.1, results from the validation exercise yield no evidence of bias in the study's matching-based impact estimates.

The validation exercise is designed to determine whether our matching approach—which could be biased if there are important unobserved differences between treatment students and matched comparison students associated with academic achievement—can replicate rigorous lottery-based impact estimates for an identical sample of treatment students (in this case, the intent to treat [ITT] impact estimates for lottery winners). In broad terms, our approach relied on the following steps. We first determined the sample of sites and students that could be used to estimate impacts using both the lottery-based and matching-based designs. We then obtained lottery-based impact estimates by comparing the first year follow-up achievement outcomes of lottery winners to those of lottery non-winners in these sites. Separately, we used the study's propensity score matching method to select a matched comparison group of non-KIPP students with similar baseline characteristics and prior achievement levels as the lottery-based treatment group. We generated matching-based impact estimates by comparing the achievement of the treatment group of lottery winners to this matched comparison group. In most respects, the lottery-based and matching-based impact estimation methods in the validation exercise were the same as those used for the full impact analyses described in Chapter II.

A. Sample of Schools and Students in the Validation Exercise

The validation sample used admission lottery data from the following eight KIPP schools: Academy Middle, Academy New York, Aspire, Austin College Prep, Key, Los Angeles Prep, Truth, and Ways. The sample for the validation exercise is smaller than the sample in the full lottery-based analysis of state test score outcomes for two reasons. First, there are two schools that appear in the full lottery-based analysis that could not be included in the validation exercise—these two schools were omitted because a substantial number of control students lost the admission lottery but were admitted to KIPP subsequently from a waitlist. In this circumstance, it is not possible for our propensity-score matching approach—which can only select comparison students who did not attend KIPP—to generate impacts that can be compared meaningfully to the lottery-based impact estimates. In particular, while the control group from the lottery analysis includes some students who attended KIPP, the comparison group in the matching design includes only non-KIPP students. Second, the validation exercise required all students to have at least one baseline or pre-baseline state test score, which further reduced the sample. With the administrative data obtained for the subset of students with baseline data, it was only possible to estimate state test score impacts (in math and reading) after one year.

B. Validation Methodology

The validation exercise was designed to ensure that any differences between the lottery-based impact and matching-based impact estimates would reflect sources of bias in the study's matching analysis. To do so, we used the exact same treatment group of 145 lottery winners in both the lottery-based and the matching-based analyses. To avoid conflating sample-driven differences with the potential bias we intend to measure, throughout this validation exercise we used precisely the

same treatment students in both the lottery-based and matching-based analyses. This created an "apples to apples" comparison of the two sets of impact estimates.[72]

For the lottery-based impact estimates, we used the exact same impact estimation approach as in our overall lottery-based estimates (for a detailed description, see Appendix E). This involved comparing the outcomes of lottery winners (n=145) to the outcomes of lottery non-winners (n=245) in a regression framework that controlled for two years of prior test scores and other student characteristics. Each treatment and control student received a weight corresponding to his or her probability of receiving a KIPP admission offer. These weights were then standardized by setting the sum of control student weights equal to the sum of treatment student weights at each site, to ensure that each lottery site was represented in the treatment group in the same proportion as that site was represented in the control group. The impact analysis pooled data across the eight lottery sites, weighting each site by the number of treatment students in the site's lottery sample. Weighting the sites in this way provides greater statistical power (because several of the lottery sites had small samples sizes, the approach we used gives less weight to the sites with less precise impact estimates, and more weight to the sites with impact estimates that have smaller standard errors).

For the matched comparison group estimates, we started with the same treatment group of 145 lottery winners and then performed propensity-score matching to select the comparison group. The matching-exercise for the validation did not use the same propensity scores estimated previously for the study's main analyses (because those earlier propensity scores estimated the probability of enrolling at a KIPP school, rather than the probability of entering a KIPP admission lottery). Instead, we estimated a new set of propensity scores based on validation sample's indicator for treatment status; we obtained these propensity scores using a logit model generated by a stepwise model-selection procedure (using a p-value cutoff of 0.20), as described in Appendix A. We estimated propensity scores for the treatment group and all comparison students located in the same jurisdictions as the lottery sites in the validation sample. We then performed nearest neighbor matching (without replacement) of comparison group students to treatment students: each comparison student was matched within the same jurisdiction and cohort as the relevant treatment student. There were no statistically significant differences between the baseline test scores or the baseline demographic characteristics of the treatment group and this matched comparison group.[73]

Using this matched sample, impact estimates for the validation exercise were obtained with the same regression model used in the main matching analysis described in Appendix D (this model controlled for two years of prior test scores and student demographic characteristics). However, the main analysis and validation analysis differ with respect to the weights assigned to the students and sites in the sample. In the main matching analysis, all treatment and comparison students were weighted equally, and each KIPP school received an equal weight in the aggregate KIPP impact estimate. For the validation exercise, we needed to ensure that we weighted the treatment students

[72] To understand why consistent samples are necessary, consider a hypothetical comparison of estimates that include all available KIPP students in the matching analysis, regardless of whether they were admitted to KIPP through a random lottery or included in the lottery-based analysis. Under such an approach, there are two possible reasons why the matching-based impact estimates could differ from lottery-based impact estimates: either the matching methods were biased or the treatment students in the matching-based analysis received a different KIPP effect, on average, than the smaller group of treatment students in the lottery-based analysis.

[73] The difference between the baseline scores of this treatment group and matched comparison group were 0.03 and 0.05 standard deviations in math and reading, respectively. Detailed baseline equivalence results are available from the authors upon request.

and lottery sites in the same way the sample was weighted for the lottery-based analysis. Thus, both the lottery-based and matching-based analyses standardized student weights such that (1) treatment students were weighted according to the probability of receiving an admission offer through the relevant lottery; (2) the sum of comparison or control student weights was set equal to the sum of treatment student weights at each site; and (3) each lottery site was assigned a weight corresponding to the total number of treatment students at that site.

To determine whether there was a statistically meaningful difference between the lottery-based and matching-based impact estimates for the validation sample, we calculated "bootstrapped" standard errors. This method re-estimates the lottery-based and matching-based impacts repeatedly, using a different configuration of the sample of lottery winners in each estimation round. We implemented 1,000 estimation rounds. In each round, we randomly sampled the lottery-based treatment and control groups in the validation exercise (with replacement), restandardized the student weights to ensure the sum of treatment student weights equaled the sum of comparison student weights at each site, and re-estimated the lottery-based and matching-based impact estimates using the new sample.[74] Calculating the standard deviation of the differences between the lottery-based and matching-based estimates allowed us to determine whether the differences between our primary validation estimates were statistically significant.[75]

C. Results of the Validation Exercise

The differences between the lottery-based and matching-based impact estimates are not statistically significant. As shown in Table F.1, the matching approach produces estimates that are 0.04 standard deviations lower than the lottery approach in math and 0.05 standard deviations higher in reading. These differences are not statistically significant, and may be too small to represent an educationally meaningful difference.[76] Additionally, the fact that the difference is negative in math but positive in reading suggests that these differences are not systematically positive or negative. However, an important caveat to this finding is the fact that due to the small samples used in the validation exercise, the statistical power of the test of the difference between the two methods is limited. The analysis is not able to detect differences smaller than approximately 0.19 standard deviations in the impact estimates.

In summary, we believe this validation exercise yields no evidence that the study's main matching-based impact estimates have a substantial bias.

To explore the sensitivity of these findings to the details of our non-experimental comparison group methods, we also compared the lottery-based impact estimates for this validation sample to impact estimates that use a much larger district-wide comparison group instead of a matched

[74] The sample size of lottery winners and non-winners was the same in each bootstrapping estimation round; a given student's observation may be used multiple times in the same impact estimate (that is, the treatment and control groups were both sampled with replacement). During each round, for the matching analysis we generated a new set of propensity-scores and selected a new matched comparison group using nearest-neighbor matching without replacement. The propensity scores in these bootstrapping rounds were generated using a consistent set of logit-model covariates (the list of covariates was selected using the original sample of all treatment students in the validation sample).

[75] Specifically, the standard deviation of the differences between matching-based and lottery-based impacts in these estimation rounds represents the standard error of the difference between our original validation impact estimates.

[76] For a student at the 45th percentile for his or her district, these differences would represent changes of approximately two percentile-points.

comparison group.[77] As shown in Table F.1, results that use a district-wide comparison group without matching produced impact estimates that are not significantly different from the study's lottery-based estimates.[78] The results do not provide any evidence that a district-wide comparison group approach is biased. This finding is consistent with the sensitivity tests presented in Appendix D, which showed that our preferred impact estimates using a matched comparison group are very similar to results without matching.

Table F.1. Comparison Between Lottery-Based Impact Estimates and Non-Lottery Impact Estimates

	Lottery-Based Analysis: Experimental Benchmark	Propensity Score Matching Analysis	Alternative Analysis Without Matching
Impacts on Math Test Scores			
Estimated impact (Standard error)	0.12 (0.07)	0.08 (0.10)	0.12** (0.04)
Difference from lottery-based analysis (Standard error)		-0.04 (0.10)	-0.001 (0.08)
Treatment sample size	145	145	145
Control or comparison sample size	245	145	199,899
Impacts on Reading Test Scores			
Estimated impact (Standard error)	-0.01 (0.10)	0.04 (0.10)	0.06 (0.06)
Difference from lottery-based analysis (Standard error)		0.05 (0.13)	0.07 (0.12)
Treatment sample size	145	145	145
Control or comparison sample size	245	145	200,544

Note: All analyses use the same treatment group. The matching analysis uses propensity-score matching followed by OLS, and the alternative analysis uses the same OLS impact model as the matching analysis, but without identifying a matched comparison group. The significance of the difference between lottery-based and non-lottery impact estimates was calculated using bootstrapped standard errors, shown in parentheses.

* Significantly different from zero at the 0.05 level, two-tailed test.
** Significantly different from zero at the 0.01 level, two-tailed test.

[77] The district-wide comparison group analysis used the exact same OLS impact estimation model used in the matching analysis. The only difference between the methods is the composition of the comparison group.

[78] Due to the much larger comparison group sample size in this alternative analysis, the estimates have more statistical power than both the main matching-based analysis and the lottery-based analysis. Using a district-wide comparison group, the impact estimate in math (0.12) is statistically significant, even though the similar lottery-based point-estimate for KIPP's impact (also 0.12 in math) is not statistically significant. Under both approaches, the impact estimate in reading is not statistically significant.

MATHEMATICA
Policy Research

www.mathematica-mpr.com

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

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

Mathematica® is a registered trademark of Mathematica Policy Research

MATHEMATICA Policy Research

REPORT

FINAL REPORT

Understanding the Effect of KIPP as it Scales: Volume I, Impacts on Achievement and Other Outcomes

Final Report of KIPP's *Investing in Innovation* Grant Evaluation

September 17, 2015

Christina Clark Tuttle
Philip Gleason
Virginia Knechtel
Ira Nichols-Barrer
Kevin Booker
Gregory Chojnacki
Thomas Coen
Lisbeth Goble

Submitted to:
KIPP Foundation
135 Main Street, Suite 1700
San Francisco, CA 94105
Project Officer: Danielle Eisenberg

Submitted by:
Mathematica Policy Research
1100 1st Street NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Project Director: Christina Clark Tuttle
Reference Number: 06826.750

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

ACKNOWLEDGMENTS

This study was made possible by the efforts of many individuals and organizations over several years. We sincerely appreciate the willingness of students, their parents, and KIPP school principals and other administrative staff to participate in the evaluation. We would also like to acknowledge the staff of the state and district departments of education who generously made their data available to our team and who provided valuable assistance and guidance.

This report would not have been possible without contributions from many other people at Mathematica. Primary data collection efforts were supported by Nikkilyn Morrison, Amanda Skaff, and Kim Mook, along with staff at Mathematica's Princeton Survey Operations Center, led by Hugo Andrade, Shelley Jacobs, and Michael Smith, and Information Services staff, led by Zuling Deng. Margaret Sullivan was responsible for securing participation from KIPP schools. Mary Anne Anderson led the efforts to collect administrative data from school districts and states. Maureen Higgins, Kevin Collins, and Mark Timms provided assistance with project management; Kevin also assisted with programming for the high school impact estimates. Clare Wolfendale led the team of three programmers who cleaned the study's administrative records from states and school districts with support from Tess Kerwin, Abigail Mosca, and Brynn Hagen. Juha Sohlberg and Emma Kopa led the review of the programming code. Brian Gill provided critical technical review and feedback. Felita Buckner led the production of the report, which was edited by Amanda Bernhardt and Betty Teller.

Finally, the study and this report benefited greatly from input at various stages from Danielle Eisenberg, Rachel Goldklang, Theresa Chen, Ann Wang, Jonathan Cowan, Freddy Gonzalez, Lisa Nouri, Shannon Stockdale, Kelly Wright, Rebecca Raber, Zoe Fenson, and Steve Mancini at the KIPP Foundation.

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

CONTENTS

EXECUTIVE SUMMARY xiii

I BACKGROUND 1

A. The KIPP network of schools 1

B. Findings from prior research 2

C. The Investing in Innovation grant program 4

D. Research questions 5

II STUDY DESIGN 7

A. Elementary schools 10

1. Sample of schools and students 11

2. Data and methods 12

B. Middle schools 13

1. Sample of schools and students 14

2. Data and methods 16

C. High schools 17

1. Sample of schools and students 18

2. Data and methods 20

III KIPP'S IMPACTS ON ACHIEVEMENT AND GRADUATION 23

A. Elementary school achievement 23

B. Middle school achievement 26

C. Patterns of KIPP middle school impacts over time 29

D. High school achievement 35

1. Why do results differ for different analyses of KIPP high school impacts? 40

IV KIPP'S IMPACTS ON STUDENT BEHAVIOR AND ATTITUDES 43

REFERENCES 51

APPENDIX A: LIST OF KIPP SCHOOLS IN NETWORK A.1

APPENDIX B: DETAIL ON SURVEY OUTCOMES B.1

APPENDIX C: CUMMULATIVE MIDDLE AND HIGH SCHOOL RESULTS C.1

APPENDIX D: DETAILED ANALYTIC METHODS: ELEMENTARY SCHOOL (LOTTERY-BASED ANALYSES) D.1

APPENDIX E: DETAILED ANALYTIC METHODS: MIDDLE SCHOOL (LOTTERY-BASED ANALYSES) E.1

APPENDIX F: DETAILED ANALYTIC METHODS: MIDDLE SCHOOL (MATCHED-STUDENT ANLAYSES)..... F.1

APPENDIX G: DETAILED ANALYTIC METHODS: HIGH SCHOOL (MATCHED-STUDENT ANALYSES)..... G.1

APPENDIX H: DETAILED ANALYTIC METHODS: HIGH SCHOOL (MATCHED-SCHOOL ANLAYSES).....H.1

APPENDIX I: DETAILED TABLES FOR WHAT WORKS CLEARINGHOUSE REVIEW..... I.1

TABLES

I.1 Summary of KIPP's project goals 5
II.1 Overview of study designs, by grade level 8
II.2 Study schools 9
III.1 Impacts of offer of admission to KIPP elementary school 24
III.2 Middle school impacts on state tests in math and reading 27
III.3 Middle school impacts on state tests in science and social studies 28
III.4 Impacts of KIPP middle schools on students two years after enrolling, by calendar year 30
III.5 Impacts of KIPP high school on achievement for new entrants 36
III.6 Impacts of KIPP high schools on achievement for continuing students 39
IV.1 Summary of significant effects, by domain 44
A.1 KIPP elementary schools (2014-2015) A.3
A.2 KIPP middle schools (2014-2015) A.5
A.3 KIPP High Schools (2014-2015) A.9
B.2 Elementary school parent survey baseline equivalence B,4
B.3 Middle school student survey baseline equivalence B.5
B.4 Middle school parent survey baseline equivalence B.6
B.5 High school student survey baseline equivalence B.7
B.6 Estimated impact of offer of admission to KIPP elementary school B.9
B.7 Estimated impact of offer of admission to a KIPP middle school B.10
B.8 Estimated impact of opportunity to attend KIPP high school B.12
B.9 Description of indices: elementary school parent survey B.17
B.10 Description of indices: middle school parent and student surveys B.20
B.11 Description of indices: high school student survey B.28
C.1 Cumulative middle and high school analysis sample C.4
C.2 Cumulative impact of KIPP middle and high school on high school achievement and graduation C.6
C.3 Baseline equivalence and impact estimates on sample with non-imputed baseline data (cumulative impact of KIPP middle and high School) C.10
D.1 Elementary school student sample sizes (lottery-based analysis) D.3
D.2 Type of schools attended by KIPP elementary school applicants D.4

D.3 Baseline equivalence for elementary school samples D.6

D.4 ITT and CACE estimates of KIPP elementary school impact D.11

D.5 ITT Estimates from alternative analysis approaches D.12

E.1 Middle school student sample sizes (lottery-based analysis) E.3

E.2 Baseline equivalence for middle school baseline sample E.5

E.3 Baseline equivalence for middle school analysis sample, year 1 ..... E.6

E.4 Baseline equivalence for middle school analysis sample, year 2 ..... E.7

E.5 Baseline equivalence for middle school analysis sample, year 3 ..... E.8

E.6 List of covariates included in impact model E.10

E.7 ITT and CACE estimates of KIPP middle school impact E.13

E.8 ITT estimates using alternative methods E.14

F.1 Middle school matched-student analytic sample F.4

F.2 Baseline equivalence for matched middle school analysis (analytic sample: year 1 reading and math outcomes) F.6

F.3 Baseline equivalence for matched middle school analysis (analytic sample: year 2 reading and math outcomes) F.7

F.4 Baseline equivalence for matched middle school analysis (analytic sample: year 3 reading and math outcomes) F.8

F.5 Baseline equivalence for matched middle school analysis (analytic sample: year 4 reading and math outcomes) F.9

F.6 Baseline equivalence for matched middle school analysis (analytic sample: social studies and science) F.10

F.7 List of potential covariates for inclusion in the propensity score estimation model F.14

F.8 List of covariates included in OLS model F.16

F.9 Baseline equivalence and impact estimates on sample with non-imputed baseline data (matched-student middle school impact analysis) F.18

G.1 High school matched-student analytic sample G.4

G.2 Baseline equivalence for the matched-student high school impact analysis, by outcome sample G.6

G.3 High school test outcomes and data source by school G.9

G.4 List of potential covariates for inclusion in the propensity score estimation model G.12

G.5 List of covariates included in OLS model G.14

G.6 Baseline equivalence and impact estimates on sample with non-imputed baseline data (matched-student high school impact analysis) G.16

H.1 High school matched-school achievement analysis, combined sample H.4

H.2 Baseline equivalence for matched high school achievement analysis (combined sample) H.6

H.3 Baseline equivalence for matched high school achievement analysis (same middle school, adjacent cohorts sample) H.6

H.4 Baseline equivalence for matched high school achievement analysis (same cohort, matched middle schools sample) H.7

H.5 List of potential covariates for inclusion in the propensity score estimation model H.10

H.6 List of covariates included in OLS model H.12

I.1 Elementary school sample sizes (full randomized sample and analytic samples) I.3

I.2 Impact of offer of admission to KIPP elementary school (ITT) I.4

I.3 Middle school sample sizes (full randomized sample and analytic samples) I.5

I.4 Impact of offer of admission to KIPP middle school (ITT) I.6

I.5 Baseline equivalence for lottery-based middle school impacts (sample with non-imputed data) I.7

I.6 Impact of KIPP middle schools (lottery-based estimates for sample with non-imputed data) I.8

I.7 Baseline equivalence for matched-student middle school impacts (sample with non-imputed data) I.9

I.8 Impact of KIPP middle schools (matched-student estimates for sample with non-imputed data) I.10

I.9 Baseline equivalence for matched-student middle school impacts (sample with non-imputed data) for new KIPP middle schools I.11

I.10 Impact of new KIPP middle schools (matched-student estimates for sample with non-imputed data) I.11

I.11 Baseline equivalence for matched-student high school impacts (sample with non-imputed baseline data) I.12

I.12 Impact of KIPP high schools for new entrants in grade 9 (matched-student estimates for sample with non-imputed baseline data) I.13

I.13 Baseline equivalence for matched-student cumulative middle and high school impacts (sample with non-imputed baseline data) I.14

I.14 Cumulative impact of KIPP middle and high schools (matched-student estimates for sample with non-imputed baseline data) I.15

I.15 Baseline equivalence for matched-school impacts of opportunity to attend a KIPP high school (adjacent cohorts sample) I.16

I.16 Impact of opportunity to attend a KIPP high school (adjacent cohort estimates) I.17

I.17 Baseline equivalence for matched-school impact of opportunity to attend a KIPP high school (matched middle schools sample) I.18

I.18 Impact of opportunity to attend a KIPP high school (matched-middle school estimates)..... I.18

FIGURES

I.1 States in the KIPP network 1

I.2 Number of KIPP schools and students, by year 4

II.1 Overview of study designs, by grades served 7

II.2 Elementary schools in study sample (2014–2015) 11

II.3 Characteristics of KIPP elementary school applicants 12

II.4 Middle schools in study sample (2014–2015) 14

II.5 Characteristics of KIPP middle school applicants 15

II.6 High schools in study sample (2014–2015) 19

II.8 Characteristics of continuing students and new entrants to KIPP high schools 20

III.1 Distribution of elementary school reading and math impact estimates after three years 25

III.2 Distribution of middle school reading and math impact estimates after two years 29

III.3 Change in the size of the KIPP network and middle school impacts over time 31

III.4 Impacts of KIPP middle schools on students two years after enrolling, by year opened 32

III.5 Differences in achievement outcomes within first two years of operation 33

III.6 Achievement impacts of KIPP middle schools over time, two years after enrollment, by school (schools opened by fall 2005) 34

III.7 Impact of KIPP high school on graduation among new entrants 37

III.8 Distribution of KIPP high school impacts on new entrants in ELA and math 38

III.9 High school impacts for continuing students, by cohort examined 40

IV.2 KIPP impacts on time spent on homework, by grade level 45

IV.3 KIPP middle school impacts on behavior 46

IV.4 KIPP impacts on parental satisfaction 47

IV.5 KIPP high school impacts on assistance for college planning 48

IV.6 KIPP high school impacts on advanced coursetaking 49

C.1 Distribution of cumulative middle and high school impacts in ELA and math C.7

C.2 Average test scores by grade level, comparing the MS-HS analytic sample to the analytic sample of new grade 9 entrants to KIPP C.8

G.1 Characteristics of KIPP versus non-KIPP high school students G.5

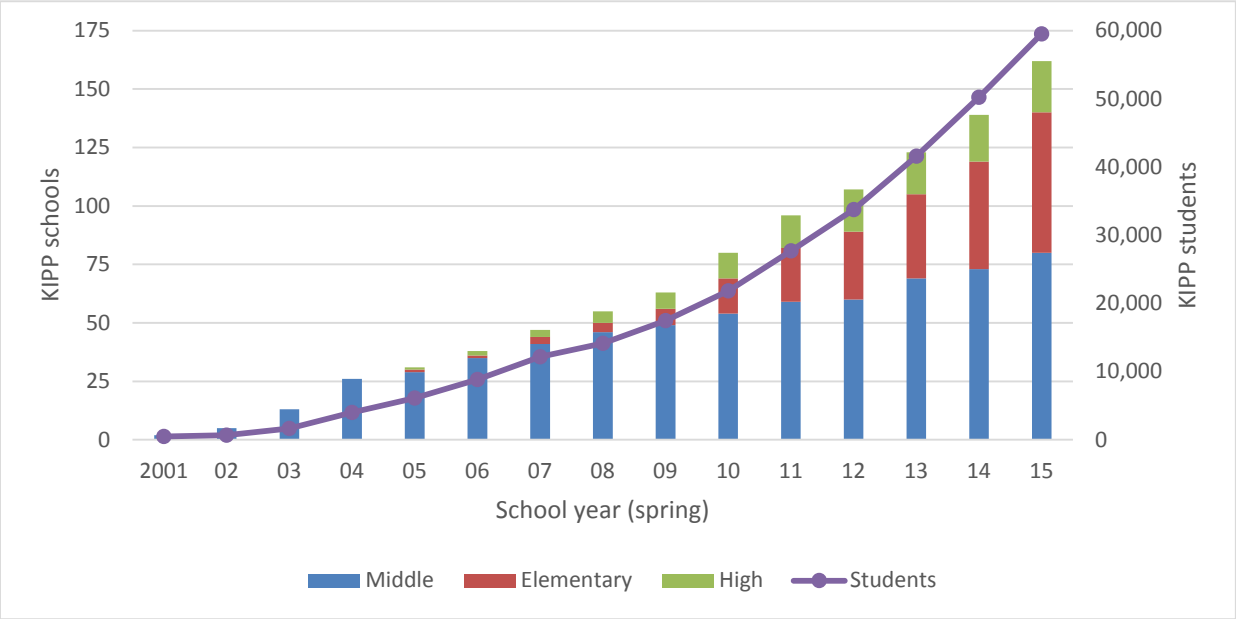
H.1 Type of high schools attended by KIPP middle school students H.5

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

EXECUTIVE SUMMARY

KIPP is a national network of public charter schools whose stated mission is to help underserved students enroll in and graduate from college. KIPP began exclusively as a middle school program in 1994, but began expanding into the elementary and high school levels in 2004. By 2009–2010, KIPP was educating students in grades prekindergarten through 12, and as of 2014–2015 the network included 162 elementary, middle, and high schools serving 59,495 students (Figure ES.1). Prior studies (see Tuttle et al. 2013) have consistently found that attending a KIPP middle school positively affects student achievement, but few have addressed longer-term outcomes and no rigorous research exists on impacts of KIPP schools at levels other than middle school.

Figure ES.1. Number of KIPP schools and students, by year

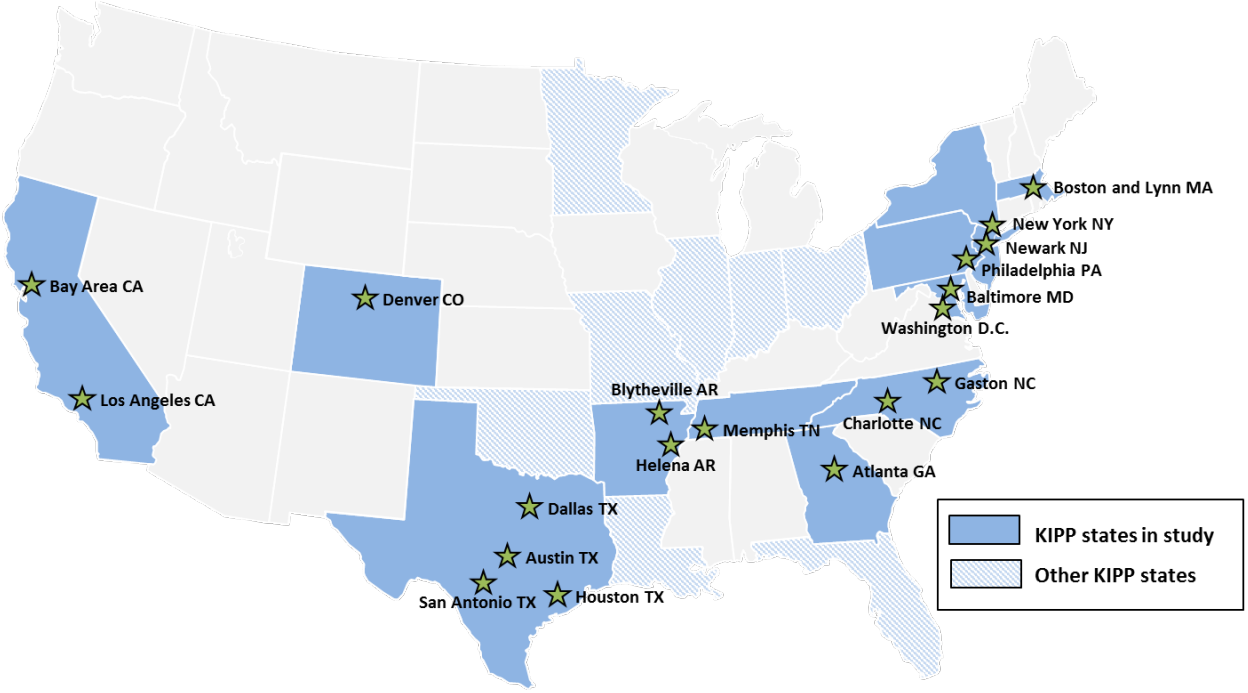


Source: KIPP Foundation.

As the KIPP network continues to grow, it faces the challenge of building a pipeline of leaders to effectively serve more students and schools. In 2010, the KIPP Foundation was awarded a five-year, $50 million Investing in Innovation (i3) scale-up grant by the U.S. Department of Education. The foundation used the i3 grant to scale up its network with the aim of sustaining KIPP's positive impacts—specifically by bolstering its leadership pipeline—while doubling the number of students served from 27,000 to over 55,000 by 2014–2015. The KIPP Foundation contracted with Mathematica Policy Research to conduct an independent evaluation of its success in improving student outcomes on a larger scale under the i3 scale-up grant. This study builds on two prior reports published by Mathematica (Tuttle et al. 2010, Tuttle et al. 2013), and is the first rigorous research to examine the impacts of KIPP schools at all three grade levels.

The key evaluation objective is to measure the impact of KIPP on student outcomes as the network scales up the number of schools, students, and grades served. To do this, we use a combination of lottery-based and quasi-experimental designs in a set of 8 elementary, 43 middle, and 18 high schools in 20 cities (Figure ES.2), employing the most rigorous study designs possible at each school level. Under different designs and samples, we measure KIPP's impacts on outcomes up to four years after students enter a KIPP school. The analysis uses data from study-administered student achievement tests; state assessments in math, English/language arts (ELA), science, and social studies; and student and parent surveys.

Figure ES.2. Location of KIPP schools in the study



Network-wide, KIPP schools have positive, statistically significant, and educationally meaningful impacts on student achievement, particularly at the elementary and middle school grades. We find that KIPP elementary schools have positive impacts on students' reading and math achievement. KIPP middle schools, meanwhile, have maintained a pattern of positive and significant impacts on reading and math over the last decade, even as the network has grown rapidly. While average impacts across the middle schools in the network declined somewhat since 2007, they stabilized during the i3 scale-up period. Moreover, the KIPP middle schools that have opened most recently—during the i3 scale-up period beginning in fall 2011—are producing positive impacts that are generally similar to those produced by older KIPP middle schools when they were in their first years of operation.

KIPP high schools have positive, statistically significant, and educationally meaningful impacts on achievement for new entrants to the network. For students continuing from KIPP middle schools, the marginal impacts of having the option to attend a KIPP high school were not statistically significant, on average (in comparison to students who did not have the option to attend a KIPP high school and instead attended a mix of other non-KIPP charter, private, and traditional public high schools). Among these continuing students, KIPP high schools have

positive impacts on several aspects of college preparation, including more discussions about college, increased likelihood of applying to college, and more advanced coursetaking.

Across grade levels, we generally find no impacts of KIPP schools on measures of students' motivation, engagement, educational aspirations, or behavior, but positive impacts on the satisfaction of parents with their child's school.

We describe these findings in more detail below.

What are the impacts of KIPP elementary schools on student achievement?

To measure impacts of KIPP elementary schools, we use a research design that uses school admissions lotteries as randomized experiments. This type of randomized design is the "gold standard" for research measuring the impacts of schools on student achievement. Students offered admission via the lottery are included in the treatment group; those not offered admission through the lottery (and enroll at other charter, private, or traditional public preschools or elementary schools) are included in the control group. This design ensures that treatment and control group students are similar at baseline (pre-KIPP) in terms of demographics and academic preparation as well as key factors such as motivation and parental support.

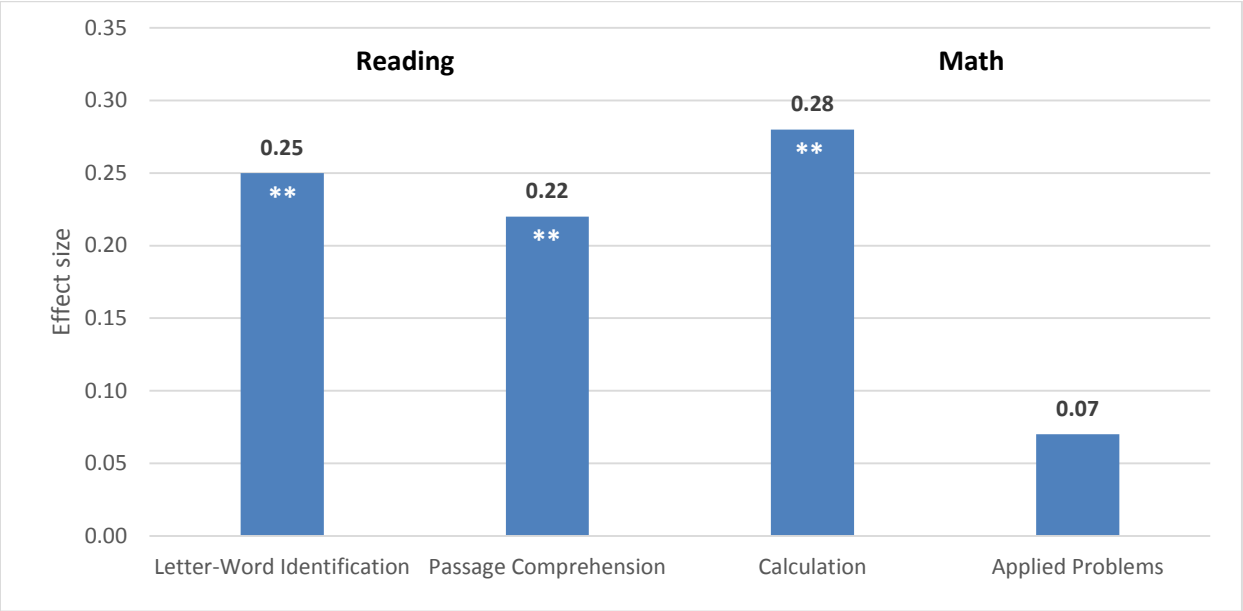
We used admissions lotteries for the 2011–2012 school year to measure the impacts of eight KIPP elementary schools. These schools make up 28 percent of KIPP elementary schools in operation in 2011–2012. To measure elementary school students' academic achievement in reading and math, we administered the Woodcock-Johnson III (WJ-III) assessment in the spring of the third follow-up year after the lottery, when most students who applied to pre-kindergarten at age 3 (PK3) were in kindergarten, and most who applied to kindergarten were in grade 2.

KIPP elementary schools have positive, statistically significant, and educationally meaningful impacts on three of four measures of students' reading and mathematics skills. On tests administered three years after entry, being offered admission to a KIPP elementary school leads to an increase of 0.25 standard deviation units on the Letter-Word Identification test and 0.22 on the Passage Comprehension test in reading (Figure ES.3). These impacts are equivalent to boosting a student's Letter-Word Identification score from the 78th percentile (the percentile corresponding to the control group students' mean score) to the 84th percentile, and boosting the Passage Comprehension score from the 48th to the 57th percentile. In math, being offered admission to a KIPP elementary school has a positive and statistically significant impact on students' Calculation score of 0.28, equivalent to an increase from the 58th to the 68th percentile. The impact on the Applied Problems score is smaller and not statistically significant.

What are the impacts of KIPP middle schools on student achievement?

We use two different approaches for measuring the impacts of KIPP middle school on student outcomes: the lottery-based design described above in sufficiently oversubscribed KIPP middle schools and a matched-student design in a broader set of KIPP middle schools. In the matched-student design, we identify a treatment group of students who enter KIPP middle schools in grade 5 or 6 and use propensity-score matching to define a comparison group of students—not attending KIPP—who most closely "match" the treatment group in terms of demographic characteristics and baseline test scores. This approach has been previously

Figure ES.3. KIPP elementary school achievement impacts



Notes: Model: Lottery-based design. Outcome: Woodcock-Johnson III Test. Sample size: eight schools; 654 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

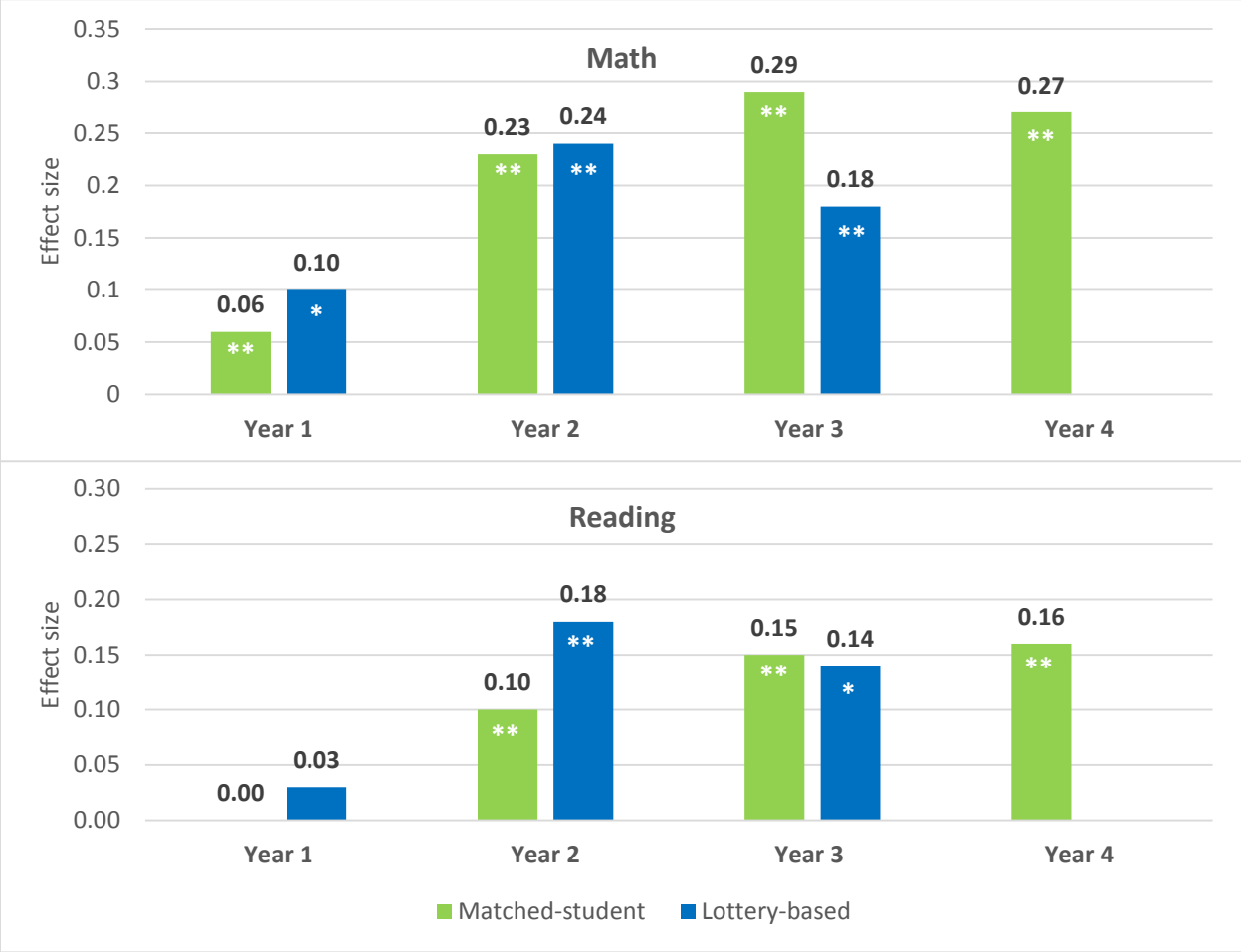
validated using lottery-based results (Tuttle et al. 2013; Fortson et al. 2015) and allows us to include as many KIPP middle schools as possible in our sample. In both designs, the comparison group comprises students attending other charter or traditional public schools.

Across the lottery-based and matched-student designs, our middle school sample includes 43 schools (53 percent of all KIPP middle schools in operation in 2014–2015), including 7 of 21 new KIPP middle schools that opened during the scale-up period. For both designs, we measured academic achievement scores on statewide assessments drawn from state- or district-provided administrative records. Students' scores were standardized (converted into z-scores) using statewide means and standard deviations, so scores represent students' achievement level relative to the typical student in the state at their grade level. We collected test score outcomes corresponding to the first three years after the lottery for the lottery-based sample, and the first four years after the treatment group entered KIPP for the matched-student sample.

Consistent with prior research, KIPP middle schools have positive, statistically significant, and educationally meaningful impacts in math, reading, science, and social studies. Based on both study designs, KIPP middle schools have positive and statistically significant impacts on students' state test scores in both math and reading, by the second year after students are admitted (Figure ES.4). For example, the lottery-based design suggests that being admitted to a KIPP middle school leads to an increase in students' average math score of 0.24 student standard deviation units after two years, equivalent to a student moving from the 40th to the 50th percentile in the state. The two-year reading impact of 0.18 is equivalent to a student moving from the 37th to the 44th percentile. The impact estimates from the matched-student design are similar for a larger sample of 37 schools, suggesting that KIPP middle schools lead to an increase in average math scores of 0.23 standard deviations and reading scores of 0.10

standard deviations. The matched-student design also suggests that, on average, KIPP middle schools have a positive and statistically significant impact of 0.25 standard deviations in both science and social studies (not shown in the figure), equivalent to moving the average student from the 48th percentile to the 58th percentile in science and from the 51st to the 61st percentile in social studies.

Figure ES.4. KIPP middle school achievement impacts



Notes: Model: Lottery-based and matched-student designs. Outcome: State test scores. Sample size: 15 schools, 608 students (lottery-based); 37 schools, 36,798 students (matched-student). Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

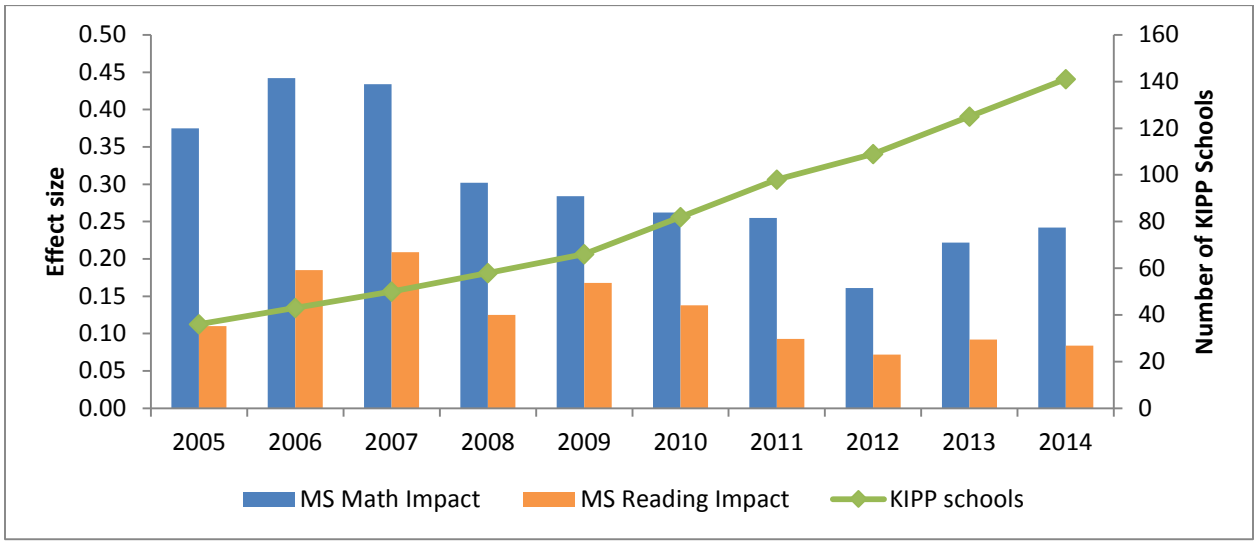
Are the previous positive impacts in KIPP middle schools maintained within the larger network as it scales?

Across the KIPP network, the average impacts of middle schools were positive and statistically significant throughout the 10-year period for which we have data, though higher in earlier years than recent years. To examine changes in the effectiveness of the network over time, we focus on trends in the impacts of KIPP middle schools, since the network has always included middle schools, and added elementary and high schools only in recent years. We calculate the average impact for each KIPP middle school in each school year, using the

results for students two years after KIPP entry. KIPP middle schools have positive and statistically significant impacts in both math and reading for all years from 2005 to 2014. Impacts were largest in 2007 and earlier, especially in math, ranging from 0.38 to 0.50 standard deviations, compared with 0.16 to 0.30 between 2008 and 2014. In 2013 and 2014, when these two-year impacts fully reflect the performance of KIPP schools during the scale-up period, math impacts are 0.22 and 0.24, respectively.

Several factors may explain the trends in KIPP middle school impacts, including changes in the number and composition of schools in the sample, the relative performance of newer versus older schools, and changes over time in the effectiveness of existing KIPP schools as the network has expanded. Overall, KIPP's student achievement impacts decreased during a time of high growth in the network, although they rebounded somewhat during the i3 scale-up period (Figure ES.5).

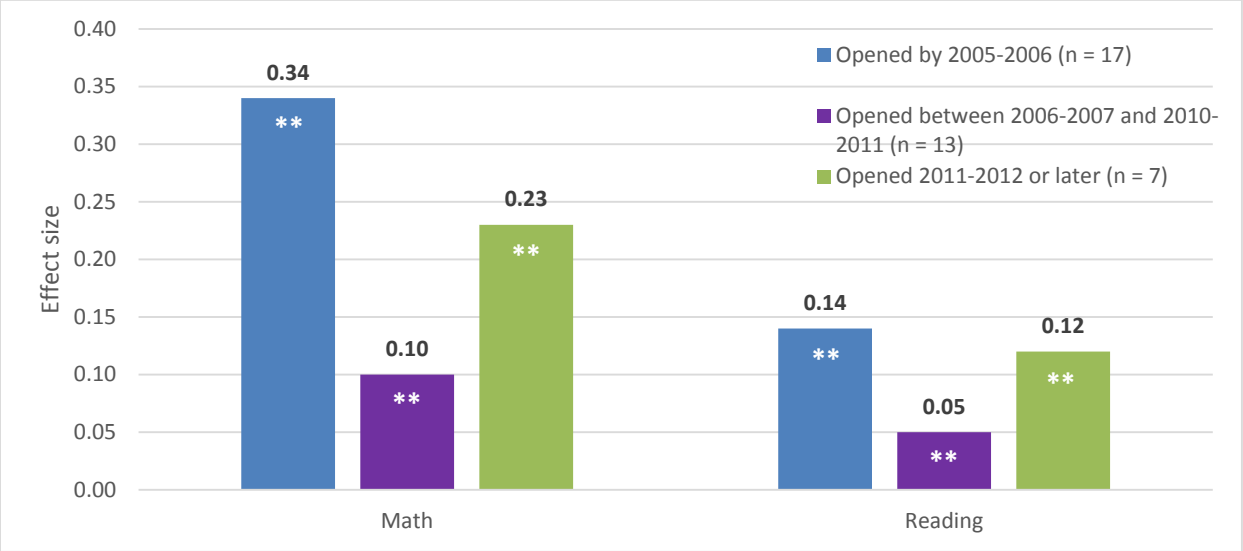
Figure ES.5. Change in the size of the KIPP network and middle school impacts over time



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. Impact estimates are calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. They are estimated separately by school year and plotted using the left-side y-axis. All impacts are statistically significant at the 0.01 level. The year refers to the spring semester of the school year when the achievement exams were taken. The size of the KIPP network is plotted against the right-side y-axis. MS = middle schools.

In fact, the newer KIPP middle schools in our matched student analysis—those opened during the i3 grant period (fall 2011 or later)—have positive impacts on math and reading achievement that are of a similar magnitude of those of the overall impacts for middle schools across the entire study period. When we compare the performance of schools opened during different periods in KIPP's history, we find that the schools opened during the scale up period have impacts that are not quite as large as the oldest KIPP schools (those opened by 2005), but larger than those opened during the period from 2006 to 2010 (Figure ES.6).

Figure ES.6. Impacts of KIPP middle schools on students two years after enrolling, by year opened



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study, based on the year the school opened. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

What are the impacts of KIPP high schools on student achievement?

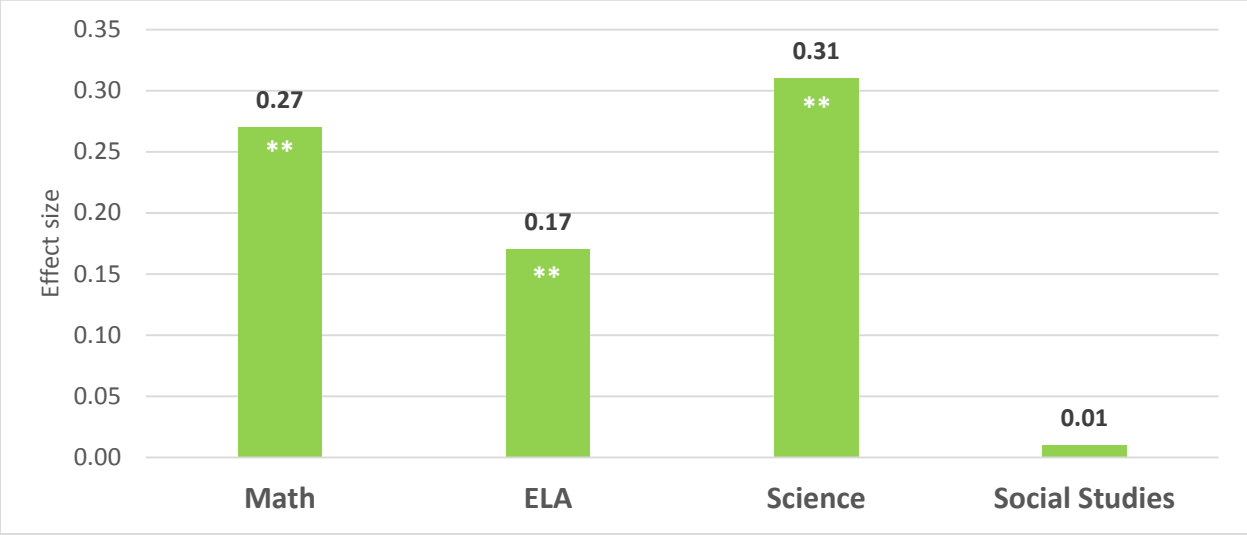
Since students enter KIPP high schools via two routes—from KIPP middle schools and non-KIPP middle schools—we use different quasi-experimental designs to measure impacts on the two groups of students. For the one-third of KIPP high-school students who entered the KIPP network for the first time in grade 9, we use a matched-student design similar to that described above for middle schools. We identify a comparison group for these new entrants, based on demographic characteristics and baseline test scores from grades 7 and 8, of students who attend other charter or traditional public high schools. For the two-thirds of KIPP high school students who also attended a KIPP middle school, we use a matched-school design, comparing outcomes for KIPP middle school students who had the option to attend a KIPP high school with outcomes for a similar set of KIPP middle school students who did not have this option. Whether or not students have the option to enter a KIPP high school depends on the location and timing of their enrollment in KIPP middle schools—in some places and years, the KIPP high school option is present and in others it is absent. Students who do not attend a KIPP high school enroll in a variety of other high schools, including other charter, private, magnet, or boarding schools, in addition to their traditional public school options. This design assumes that aside from the presence/absence of the KIPP high school option, the treatment and comparison groups are similar, on average.

We include 14 KIPP high schools in our matched-student analysis of new entrants and 8 high schools in our matched-school analysis of continuing KIPP students; 4 high schools are included in both designs. Across designs, the high school sample includes 82 percent of all KIPP high schools in operation in 2014–2015. We measure student achievement outcomes using state

assessments for the analysis of impacts on new entrants. For the analysis of impacts on continuing KIPP students, state test scores are less consistently available, because many in the comparison group were attending private high schools or public schools outside the jurisdictions providing data. We therefore measure achievement in the analysis of continuing KIPP students by administering a TerraNova assessment in the third follow-up year after high school entry (typically grade 11).

For new entrants to the network, KIPP high schools have positive, statistically significant, and educationally meaningful impacts on achievement in math, ELA, and science. Having the opportunity to attend a KIPP high school boosts new entrants' high school math scores by 0.27 standard deviation units, a statistically significant impact representing an increase from the 48th to the 59th percentile for the typical student (Figure ES.7). Impacts in ELA and science are 0.18 and 0.31 standard deviations, respectively, and are also significant. Relative to outcomes for the matched comparison group, these impacts are equivalent to an increase from the 47th to the 54th percentile in ELA and from the 42nd to the 54th percentile in science. The average impact in social studies (0.01) is close to zero and not statistically significant. The magnitude of the impact on graduation after four years is positive (four percentage points, not shown), but also not statistically significant.

Figure ES.7. KIPP high school impacts for new entrants



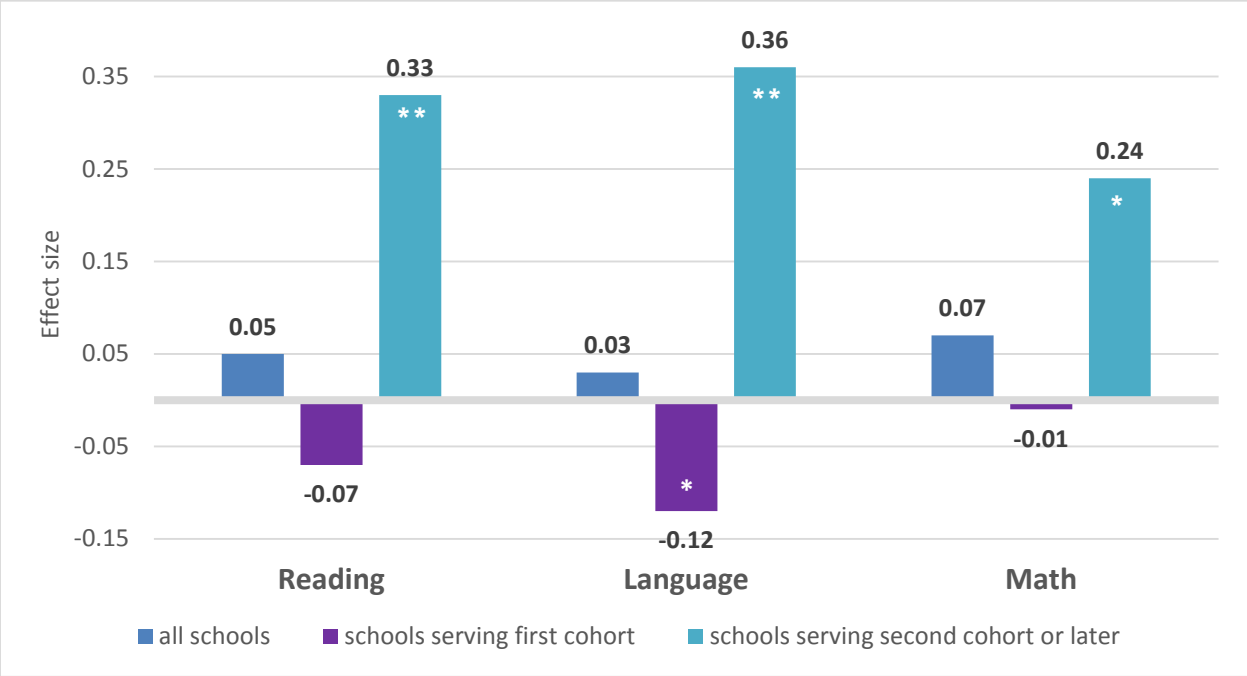
Notes: Model: Matched-student design. Outcome: State test scores. Maximum sample size: 14 schools; 1,861 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

For students continuing from KIPP middle schools, the achievement impacts of KIPP high schools are not statistically significant on average, but these impacts vary by school. For continuing students (compared with KIPP middle school graduates without access to a KIPP high school), the average impacts of KIPP high schools on TerraNova tests in reading, language, and math are positive, but small and not statistically significant (Figure ES.8). These results underestimate the full impact of actually attending a KIPP high school, because all students with the opportunity to attend are included in the treatment group, but not all of them in fact attended. Five of the eight KIPP high schools in this analysis were brand new and serving their first cohort

of students when we measured their impacts. There are significant differences between the impacts of these new KIPP high schools and those of more experienced KIPP high schools. For the five new schools, impacts are negative in all three subjects and statistically significant in language. For the three more experienced high schools, impacts are positive and statistically significant in all three subjects, with magnitudes ranging from 0.24 to 0.36. These more positive impacts for more experienced high schools could imply that KIPP high schools become more effective as they gain experience. Because we do not have data to measure impacts of high schools in multiple years under this design, we cannot determine if KIPP high schools increase their impacts on continuing students as the schools gain experience.

Continuing students with the option to attend a KIPP high school are less likely to drop out of high school. The overall dropout rate is very low, but is significantly lower for the treatment group—1 percent for those who had the chance to attend a KIPP high school and 3 percent for those who did not.

Figure ES.8. KIPP high school impacts for continuing students



Notes: Model: Matched-school design. Outcome: TerraNova test. Sample size: eight schools; 933 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

What are the impacts of KIPP schools on student outcomes other than achievement?

We administered surveys to students and parents to measure impacts on key outcomes other than achievement at all three grade levels.

KIPP elementary and middle schools have positive impacts on school satisfaction, particularly among parents. At both the elementary and middle school levels, being offered

admission to KIPP leads to increases in parents' satisfaction with their child's school. More than three-quarters of elementary parents in the treatment group rate their child's school as excellent, compared to about half of parents in the control group. At the middle school level, 56 percent of treatment group parents and 28 percent of control group parents rate the school as excellent. These findings are consistent with previous research on KIPP in particular and oversubscribed charter middle schools in general (Tuttle et al. 2013; Gleason et al. 2010). Similarly, KIPP has significant positive impacts on a parent-based index capturing satisfaction with school facilities, academics, safety, and discipline. KIPP also has significant positive impacts on several other satisfaction measures, including indices of school efforts to engage parents at both the elementary and middle school level and middle school students' perceptions of their schoolmates. Evidence of KIPP impacts on satisfaction do not extend to the high school level, however, as none of eight measures of impacts on student-reported satisfaction at that level were statistically significant.

KIPP high schools have positive effects on several aspects of college preparation, including discussions about college, applying to college, and coursetaking. KIPP high schools have positive and significant impacts on measures related to school assistance in planning for college, including the frequency of discussions about college at school, students being more likely to have in-depth discussions at school about how to pay for college, and teacher or counselor assistance with planning for college. In addition to assistance provided by the school, KIPP high schools have a positive and significant effect on college preparation activities undertaken by students, as well as on whether the student applied to at least one college or university by spring of senior year—93 percent of treatment students did so, compared with 88 percent of comparison students. Students with the opportunity to attend a KIPP high school enroll in schools more likely to offer advanced placement (AP) or international baccalaureate courses (97 percent versus 89 percent), and the number of AP courses and exams students have taken or intend to take is correspondingly higher.

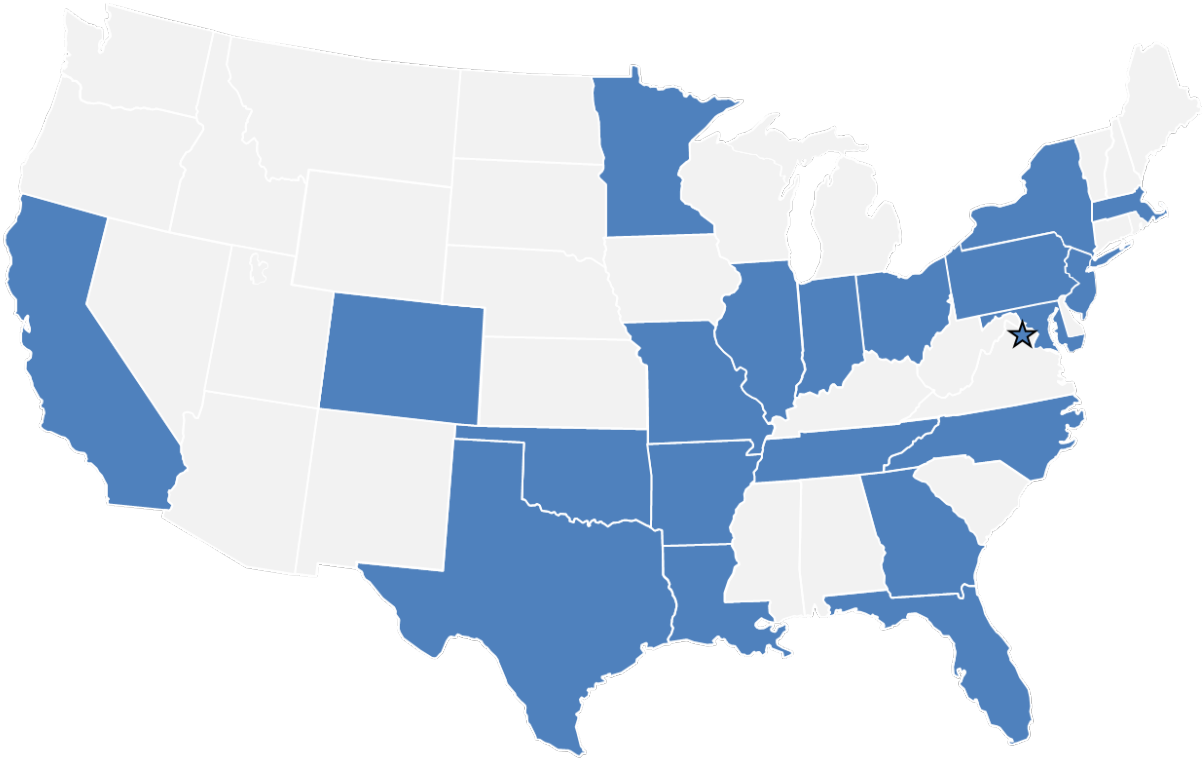
On average, across grade levels, KIPP schools have no statistically significant impact on most survey measures of student motivation and engagement, behavior, or educational aspirations. At all three grade levels, KIPP did not significantly affect measures of motivation and engagement related to student self-control, academic motivation, academic confidence, grit, school engagement, or effort in school, including student reports of the time spent on homework. Student behavior was measured only at the elementary and middle school levels; we find no evidence that KIPP schools affect behavior, including indices of positive behaviors, undesirable behaviors, peer pressure, illegal activities, parental concerns about their child, frequency of school disciplinary actions (according to the parent), and the extent to which the child is well-adjusted. We measure educational goals and aspirations using responses from both parents and the students themselves. In general, the educational goals and aspirations among these elementary, middle, and high school students are high in both the treatment (KIPP) and comparison (non-KIPP) groups. At the high school level, for example, 84 percent of students reported that they think they will graduate from college. For 12 of 13 outcomes, the estimated impact of KIPP is not statistically significant. The single exception is among parents of students at KIPP elementary schools, who are 10 percentage points more likely than the comparison group to believe their child is very likely to complete college (81 versus 71 percent).

I. BACKGROUND

A. The KIPP network of schools

KIPP is a national network of public charter schools with 162 elementary, middle, and high schools operating in the 2014–2015 school year, serving 59,495 students. Nearly all KIPP network schools are operated as part of one of 29 autonomous regional organizations in 20 states and the District of Columbia (Figure I.1).[1] KIPP regions oversee schools in a specific metropolitan or geographic area, providing support on leadership practices, human resources, business operations, technology, and development.

Figure I.1. States in the KIPP network



KIPP schools emphasize rigorous academics and character instruction with the ultimate goal of preparing students to succeed in college and beyond. The KIPP Approach is distinguished by seven key principles, which evolved from the Five Pillars, a set of operating principles which historically guided KIPP schools:[2]

1. A belief that all students can learn and achieve
2. A focus on college graduation as the ultimate goal

[1] Two middle schools are autonomous single sites.

[2] A full description of the KIPP Approach is available at www.kipp.org/our-approach.

3. An emphasis on providing rigorous academics while simultaneously developing student character
4. A belief that visionary, empowered leaders are central to the development and operation of successful schools
5. A belief that excellent teachers are critical to help students succeed in school and beyond
6. A belief that empowered leaders and teachers should leverage existing knowledge and resources when exercising their autonomy
7. A focus on continuous learning and improvement

All 162 KIPP schools in 2014-2015 are public charter schools, and nearly all have been charter schools since they opened. Thus, KIPP schools have greater autonomy to set their own policies than do most traditional public schools, but are accountable to their authorizers for achieving satisfactory performance. Over time, 15 KIPP schools have closed or lost their affiliation with KIPP— in several cases, for performance reasons. KIPP schools have open enrollment policies, but students must choose to apply to and enroll in a KIPP school, and may return to a district school at any time.

B. Findings from prior research

Prior research has consistently found that attending a KIPP school leads to positive effects on student achievement. These positive findings have, in part, fueled KIPP's rapid growth. Using quasi-experimental methods, Mathematica's prior study of 43 KIPP middle schools found positive impacts of KIPP on student achievement across four academic subjects in each of a student's first four years after enrollment and for all examined student subgroups (Gleason et al. 2014). For example, entering a KIPP middle school led to increases in student achievement in math and reading of 0.36 and 0.21 standard deviations after three years, respectively. These are large effects, equivalent to about 90 percent of a year of extra learning in math and about two-thirds of a year of extra learning in reading (Hill et al. 2008). Experimental impacts from this study based on randomized admission lotteries for a much smaller sample of schools and cohorts were consistent with these findings (Tuttle et al. 2013).

Other studies using strong research designs have also found positive impacts on student achievement that were educationally important and statistically significant. In a study of charter management organizations (CMOs), KIPP schools in Washington, D.C., were identified as demonstrating significantly positive two-year impacts in both math and reading, exceeding the impact of the average CMO (Furgeson et al. 2012; Lake et al. 2012). A propensity-score analysis of three KIPP Bay Area (California) middle schools found positive impacts of 0.16 to 0.86 standard deviations on student achievement (Woodworth et al. 2008). Finally, Angrist et al. (2010) used an experimental design based on a randomized admission lottery at KIPP Academy Lynn Middle School (Massachusetts) and found that a year of attendance significantly increased achievement scores by 0.35 standard deviations in math and 0.12 in reading.

Skeptics of KIPP argue that these schools rely on selective admission, attrition, and replacement of students to produce positive achievement results. However, data on student characteristics provide little evidence that KIPP "creams" or selectively enrolls higher-performing students at the middle school level (Tuttle et al. 2013). The typical KIPP student

scored at the 45th percentile within the district—that is, below the district average—in reading and math prior to entering KIPP.[3] Nearly all KIPP students (96 percent) are either black or Hispanic, and more than four-fifths (83 percent) are from households with incomes low enough to make them eligible for free or reduced-price school meals—percentages that exceed the averages at the (non-KIPP) elementary schools they attended prior to enrolling in KIPP middle schools. In contrast, KIPP students are somewhat less likely than students at these feeder elementary schools to have received special education services (9 versus 13 percent) or to have been classified as having limited English proficiency (10 versus 15 percent) when they were in elementary school. Patterns of student attrition from KIPP middle schools are similar to those at nearby non-KIPP public schools (Nichols-Barrer et al. 2015). However, unlike traditional public schools in surrounding districts, when students exit, KIPP schools tend to replace them with higher-achieving students, and fewer students are replaced in the later years of middle school. Still, KIPP's positive achievement impacts do not appear to be explained by advantages in the prior achievement of KIPP students, even when attrition and replacement patterns are taken into account (Nichols-Barrer et al. 2015).

In addition to affecting students' academic achievement, KIPP schools may influence student behaviors and attitudes related to long-term academic success. Using a lottery-based analysis of middle schools, Tuttle et al. (2013) found that KIPP middle schools have no statistically significant effect on a variety of measures of student attitudes that may be related to long-run academic success, including indices of student-reported self-control, academic self-concept, school engagement, effort/persistence in school, and educational aspirations. KIPP similarly does not have a significant effect on several measures of student behavior, including self-reported illegal activities, an index of good behavior, and parent reports of behavior problems. KIPP does have a significant effect on a student-reported measure of undesirable behavior, with KIPP students more likely to report behaviors such as losing their temper, arguing or lying to their parents, or giving their teachers a hard time. By contrast, KIPP leads to a significant increase in the amount of time students report spending on homework and has a positive effect on students' and parents' satisfaction with school.

Despite the growing body of research on the effects of KIPP, three important sets of questions remained prior to this report. First, little research has explored the long-term effects of KIPP middle schools. The KIPP Foundation found that, as of spring 2015, 45 percent of KIPP middle school graduates had earned a four-year college degree in 10 or more years, compared to the national average of less than 10 percent of students from low-income families (KIPP Foundation 2015). These findings are based on the two original KIPP academies in Houston and New York. In addition, these findings have not been supported by studies using more rigorous designs. Second, existing research does not tell us anything about impacts of KIPP schools at levels other than middle school. As KIPP expands up and down, will the model that has been successful in middle schools work at the elementary and high school levels as well? Third, we do not know much about whether KIPP middle schools have been able to maintain their positive impacts as the size of the overall network has grown. Tuttle et al. (2013) measured the impacts of a group of 41 KIPP middle schools that opened in or before 2009–2010, but the overall network

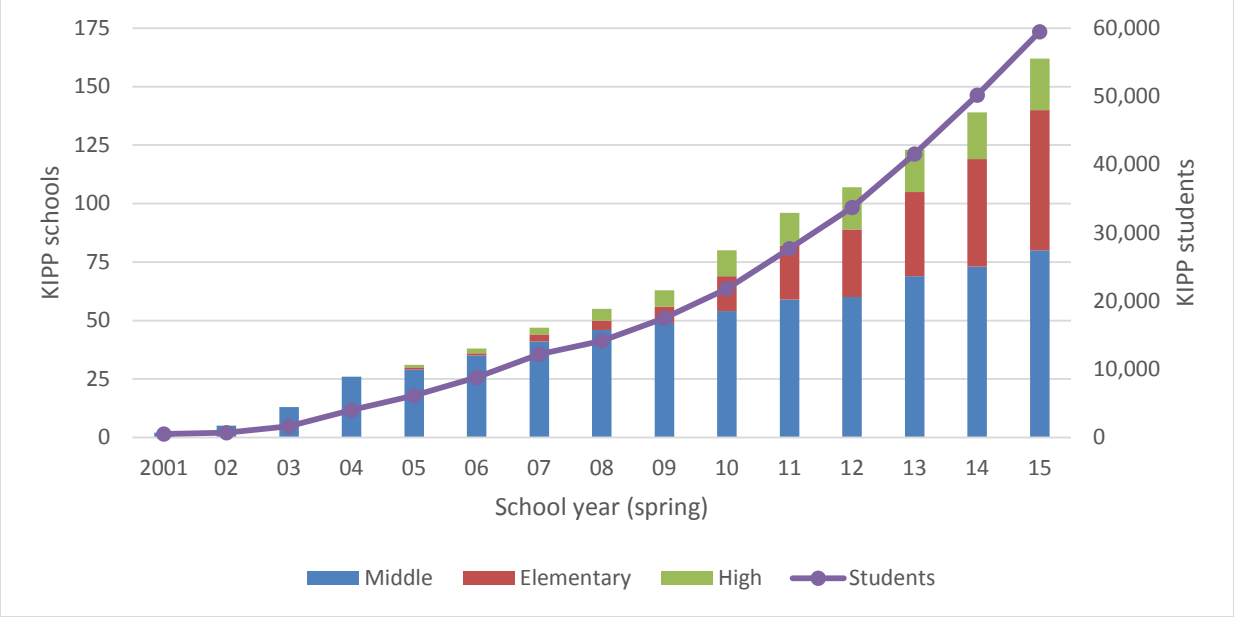
[3] The average achievement level of KIPP middle school students before they entered a KIPP school was also significantly lower than the average among the other students at the elementary schools they attended (Tuttle et al. 2013).

now includes four times that many schools. This study directly addresses the second and third questions.

C. The Investing in Innovation grant program

In 2000, KIPP founders Mike Feinberg and Dave Levin—in partnership with Doris and Don Fisher, founders of Gap Inc.—established the KIPP Foundation to support the expansion of the KIPP network from its first two schools (opened in 1994). The KIPP Foundation guides the network's growth by selecting and training school leaders, promoting the program model, and providing supports and services to KIPP schools and regions on legal, real estate, technology, finance, corporate governance, operations, communications, marketing, and development matters. With the support of the KIPP Foundation, the network has grown dramatically (Figure I.2). Initially, it included only middle schools serving grades 5 to 8 in an increasing number of cities across the country. KIPP expanded into the elementary and high school levels in 2004, with the first such schools opening in Houston. Since then, the majority of growth in the KIPP network has occurred through the opening of new elementary and high schools in cities where KIPP middle schools are already operating.

Figure I.2. Number of KIPP schools and students, by year



Source: KIPP Foundation.

Note: Fifteen schools that closed or left the KIPP network are not included. The network plans to open 21 new schools in fall 2015.

Given its success to date, a key question is whether KIPP can maintain its positive effects if it serves a larger set of students in a larger set of schools. To what extent is KIPP's early success related to the fact that it was a modestly sized organization and focused on the middle school grades? Can this success be maintained when the network has to recruit and retain a larger group of teachers and administrators, as well as attract and serve a larger number of students at different levels? More generally, stakeholders and policymakers wonder whether KIPP's

approach can be adopted more broadly in the charter school sector, or expand to traditional public schools. If the positive impacts of KIPP disappear as the network grows, this would suggest that the model may not work well if adopted by larger school systems. However, if impacts remain positive and strong as the KIPP network continues to expand, then the KIPP model may offer greater promise for public schools more generally.

In October 2010, the KIPP Foundation was awarded a five-year, $50 million Investing in Innovation (i3) scale-up grant by the U.S. Department of Education. The award was one of four i3 scale-up grants that year intended to fund expansion of programs demonstrating strong evidence of previous effectiveness in improving student achievement and attainment. The foundation planned to use the i3 grant to scale up its network while sustaining KIPP's positive impacts. The 2011–2012 school year represented the first full year of i3 grant implementation and marks the beginning of the "scale-up period," which ran through 2014–2015.

KIPP's goals for the proposed scale-up project were threefold. Activities during the scale-up period were focused on (1) increasing the pipeline of effective leaders prepared to lead KIPP schools, (2) placing these newly trained leaders into new or existing KIPP schools grounded in the KIPP Approach, and (3) equipping other public schools to adopt leadership practices that had been successful at KIPP. Table I.1 provides further details of these goals.

Table I.1. Summary of KIPP's project goals

Goal 1	Train 1,000 leaders, including approximately 250 principals, who will each open a new school or assume the leadership of an existing school during the grant period (includes approximately 60 principals outside the KIPP network), and 750 future leaders who will start on the path to school leadership.
Goal 2	Increase annual school openings by at least 50 percent, accelerating from opening an average of 10 schools per year in the last five years to 15 to 18 schools per year during the grant period. Accelerated growth will allow 50,000 students to be served in urban and rural KIPP schools by the end of the grant period, increasing to 66,000 students as those schools reach full enrollment.
Goal 3	Share proven KIPP leadership practices with non-KIPP schools (a) in the urban and rural school districts in which KIPP schools are located and (b) in other growing charter management organizations. By adopting these shared leadership practices, these non-KIPP schools will deepen and expand their own principal pipelines to benefit millions more students.

Source: KIPP i3 grant application; see http://www2.ed.gov/programs/innovation/2010/narratives/u396a100031.pdf.

D. Research questions

As the KIPP network continues to grow into new communities and grades, it faces the dual challenge of effectively serving more students while also building a solid pipeline of leaders to sustain its success. To evaluate its success in achieving these goals, the KIPP Foundation contracted with Mathematica Policy Research to conduct an independent evaluation of the KIPP network and its expansion under the i3 scale-up grant.

The key objective of the evaluation is to measure the impact of KIPP schools on student achievement as the network scales up the number of schools, students, and grades served. In particular, the evaluation addresses the following primary research questions:

- What are the impacts on student achievement of KIPP elementary schools and KIPP high schools?
- What are the impacts of the scaled-up network of KIPP middle schools on student achievement? Are the previous positive impacts in middle schools maintained within the larger network?
- How do impacts of new KIPP middle schools established during the i3 scale-up period compare with those of previously established KIPP middle schools, on average?

In addition, the evaluation addresses a series of exploratory questions of interest:

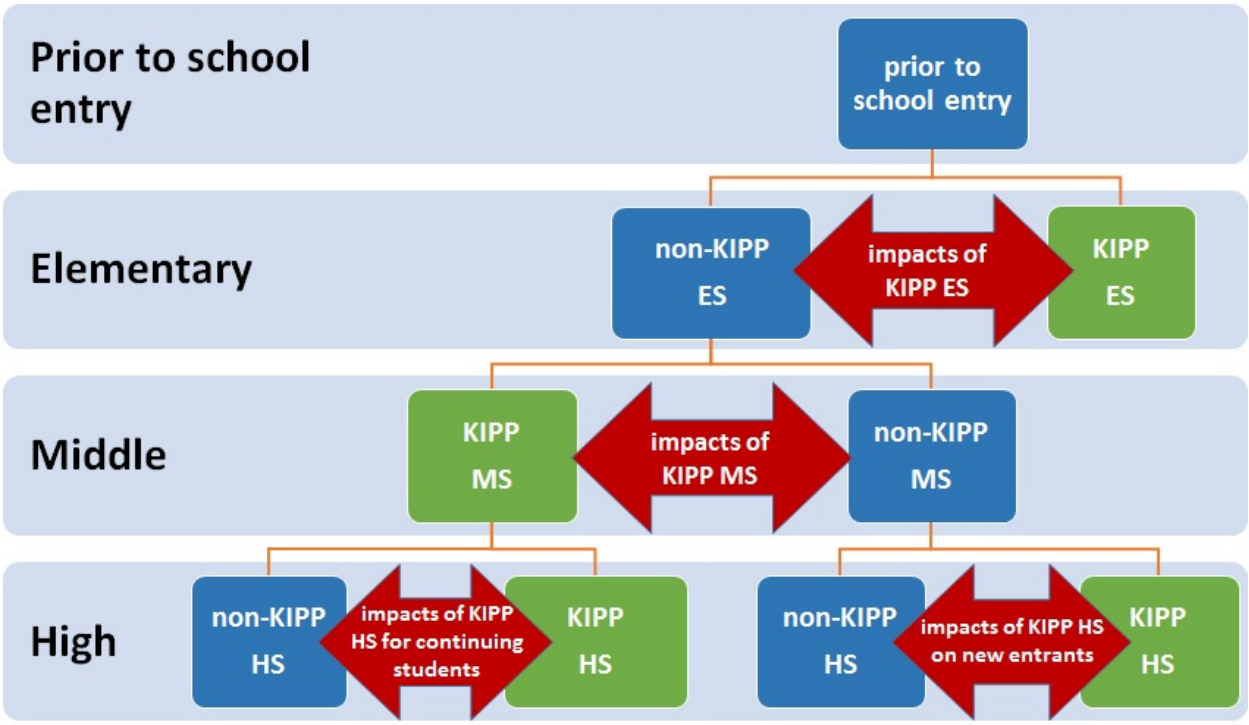
- Do the impacts of KIPP schools on student achievement vary across schools?
- What are the impacts of KIPP schools on student outcomes other than achievement, including:
  - Motivation and engagement?
  - School experiences and satisfaction?
  - Behavior?
  - Goals, aspirations, and college preparation?
  - Likelihood of dropping out of high school?

In the next chapter, we provide an overview of the research designs, data, and samples of schools and students that we employ to answer each of these questions. (More detail on our samples and methods may be found in the appendices to this volume). Chapter III discusses the impacts of KIPP elementary, middle, and high schools on the achievement of their students. Finally, in Chapter IV, we investigate the impacts of KIPP schools on outcomes outside of test scores.

II. STUDY DESIGN

To measure the effect of KIPP on student achievement, we employ the most rigorous possible study design at each school level. Figure II.1 summarizes the designs we use to estimate the effects of KIPP at the elementary, middle, and high school levels. The figure captures students' progression through their elementary and secondary school years, where they may have an opportunity to enter a KIPP school at various points. Among students following a similar schooling path up to each of these points, we compare those who enter KIPP with otherwise similar students who do not. Each of these comparisons serves as a basis for one of our study designs, and is represented in the figure by a red arrow. In each case, the treatment group (represented by green boxes) and the comparison group (represented by blue boxes) come from the same pool of students at the prior level. To measure the effects of KIPP middle schools, for example, we follow students who were in non-KIPP elementary schools and compare those who enter KIPP with those who remain in non-KIPP schools during their middle school years. In comparing outcomes for the KIPP and non-KIPP groups, we use either experimental or quasi-experimental methods to ensure that the groups are otherwise similar at baseline, so that outcome differences we later observe can be credibly attributed to attendance at a KIPP school.

Figure II.1. Overview of study designs, by grades served



An overview of each of these study designs, including information on the samples, outcome data, and years of follow-up, is presented in Table II.1. We discuss each of these in detail in the following pages. A complete list of the schools included in the study is found in Table II.2.

Table II.1. Overview of study designs, by grade level

Grade level	Design	Sample: Schools	Sample: Treatment group	Sample: Comparison group	Sample: Number of cohorts and students	Outcome data (year collected)
Elementary (entry at grade PK3 or K)	Lottery-based	Oversubscribed elementary schools (n = 8)	Consenting lottery winners (n = 473)	Consenting lottery losers (n = 624)	1 cohort (1,097 students)	• Study-administered test (WJ-III; year 3) • Parent surveys (year 2)
Middle (entry at grade 5 or 6)	Lottery-based	Oversubscribed middle schools (n = 16)	Consenting lottery winners (n = 459)	Consenting lottery losers (n = 432)	1 cohort (891 students)	• State tests (years 1 through 3) • Student surveys (year 2) • Parent surveys (year 2)
	Matched-student	All middle schools with available data (n = 37)	KIPP MS students with valid baseline test scores (n = 20,312)	Matched comparison students never attending KIPP MS (n = 20,312)	2 to 13 cohorts (40,624 students)	• State tests (years 1 through 4)
High (entry at grade 9)	Matched-student (new entrants)	All high schools with available data (n = 14)	Students entering KIPP from a non-KIPP middle school with valid baseline test scores (in grade 8) (n = 1,380)	Matched comparison students never attending KIPP HS (n = 1,380)	1 to 9 cohorts (2,760 students)	• State tests (year 1, 2, or 3) • State graduation indicator (year 4)
	Matched-school (continuing students)	All high schools with an appropriate comparison cohort (n = 8)	Students from a KIPP middle school with the opportunity to attend a KIPP high school (n = 464)	Students from a KIPP middle school without the opportunity to attend a KIPP HS (n = 469)	1 cohort (933 students)	• Study-administered test (TerraNova; year 3) • Student surveys (year 4)

Notes: The year(s) reported for the outcome measures indicate the number of years after potential entry to a KIPP school during which the outcome data were collected. WJ-III refers to the Woodcock-Johnson III assessment.

We also conducted a matched-student analysis of the cumulative impacts of KIPP middle and high schools on students who attended both types of schools; these results are presented in Appendix C.

Table II.2. Study schools

Region	Level	School	Lottery -Based	Matched -Student	Matched-School
KIPP Atlanta	middle	KIPP South Fulton Academy	X		
KIPP Atlanta	middle	KIPP STRIVE Academy		X	
KIPP Atlanta	middle	KIPP Vision Academy		X	
KIPP Atlanta	middle	KIPP WAYS Academy	X	X	
KIPP Atlanta	high	KIPP Atlanta Collegiate		X	
KIPP Austin	middle	KIPP Austin Academy of Arts & Letters	X	X	
KIPP Austin	middle	KIPP Austin Beacon Prep		X	
KIPP Austin	middle	KIPP Austin College Prep	X	X	
KIPP Austin	middle	KIPP Austin Vista Middle School		X	
KIPP Austin	high	KIPP Austin Collegiate		X	X
KIPP Baltimore	middle	KIPP Ujima Village Academy	X		
KIPP Bay Area	middle	KIPP Summit Academy	X		
KIPP Bay Area	high	KIPP King Collegiate High School			X
KIPP Bay Area	high	KIPP San Jose Collegiate			X
KIPP Charlotte	middle	KIPP Charlotte		X	
KIPP Colorado	middle	KIPP Montbello College Prep		X	
KIPP Colorado	middle	KIPP Sunshine Peak Academy		X	
KIPP Colorado	high	KIPP Denver Collegiate High School		X	
KIPP Dallas-Fort Worth	middle	KIPP TRUTH Academy	X	X	
KIPP DC	elementary	KIPP DC: LEAP Academy	X		
KIPP DC	middle	KIPP DC: AIM Academy		X	
KIPP DC	middle	KIPP DC: KEY Academy		X	
KIPP DC	middle	KIPP DC: WILL Academy		X	
KIPP DC	high	KIPP DC: College Preparatory		X	X
KIPP Delta	middle	KIPP Blytheville College Prep. School		X	
KIPP Delta	middle	KIPP Delta College Preparatory School		X	
KIPP Delta	high	KIPP Delta Collegiate High School		X	
KIPP Eastern North Carolina	middle	KIPP Gaston College Preparatory	X	X	
KIPP Eastern North Carolina	high	KIPP Pride High School		X	
KIPP Houston	elementary	KIPP SHARP College Prep Lower School	X		
KIPP Houston	elementary	KIPP SHINE Prep	X		
KIPP Houston	middle	KIPP 3D Academy		X	
KIPP Houston	middle	KIPP Academy Middle School (Houston)		X	
KIPP Houston	middle	KIPP Courage College Prep		X	
KIPP Houston	middle	KIPP Intrepid Preparatory School		X	
KIPP Houston	middle	KIPP Liberation College Prep		X	
KIPP Houston	middle	KIPP Polaris Academy for Boys		X	
KIPP Houston	middle	KIPP Sharpstown College Prep	X	X	
KIPP Houston	middle	KIPP Spirit College Prep		X	
KIPP Houston	middle	KIPP Voyage Academy for Girls		X	
KIPP Houston	high	KIPP Generations Collegiate		X	

Table II.2 *(continued)*

Region	Level	School	Lottery -Based	Matched-Student	Matched-School
KIPP Houston	high	KIPP Houston High School		X	
KIPP Houston	high	KIPP Northeast College Preparatory		X	
KIPP Houston	high	KIPP Sunnyside High School		X	
KIPP L.A.	elementary	KIPP Raíces Academy	X		
KIPP L.A.	middle	KIPP LA College Preparatory School	X		
KIPP Massachusetts	middle	KIPP Academy Boston Middle School		X	
KIPP Massachusetts	middle	KIPP Academy Lynn Middle School	X	X	
KIPP Massachusetts	high	KIPP Academy Lynn Collegiate High School	X		
KIPP Memphis	middle	KIPP Memphis Academy Middle		X	
KIPP Memphis	middle	KIPP Memphis Collegiate Middle		X	
KIPP Memphis	high	KIPP Memphis Collegiate High		X	
KIPP New Jersey	elementary	SPARK Academy	X		
KIPP New Jersey	high	Newark Collegiate Academy			X
KIPP NYC	elementary	KIPP Academy Elementary School	X		
KIPP NYC	elementary	KIPP Infinity Elementary School	X		
KIPP NYC	middle	KIPP Academy Middle School (New York)	X	X	
KIPP NYC	middle	KIPP AMP Middle School		X	
KIPP NYC	middle	KIPP Infinity Middle School	X	X	
KIPP NYC	middle	KIPP STAR Harlem Middle School	X	X	
KIPP NYC	middle	KIPP Washington Heights Middle School		X	
KIPP NYC	high	KIPP NYC College Prep High School		X	X
KIPP Philadelphia	elementary	KIPP Philadelphia Elementary Academy	X		
KIPP Philadelphia	middle	KIPP Philadelphia Charter School	X		
KIPP Philadelphia	middle	KIPP West Philadelphia Preparatory	X		
KIPP Philadelphia	high	KIPP DuBois Collegiate Academy			X
KIPP San Antonio	middle	KIPP Aspire Academy		X	
KIPP San Antonio	middle	KIPP Camino Academy		X	
KIPP San Antonio	high	KIPP University Prep High School		X	X

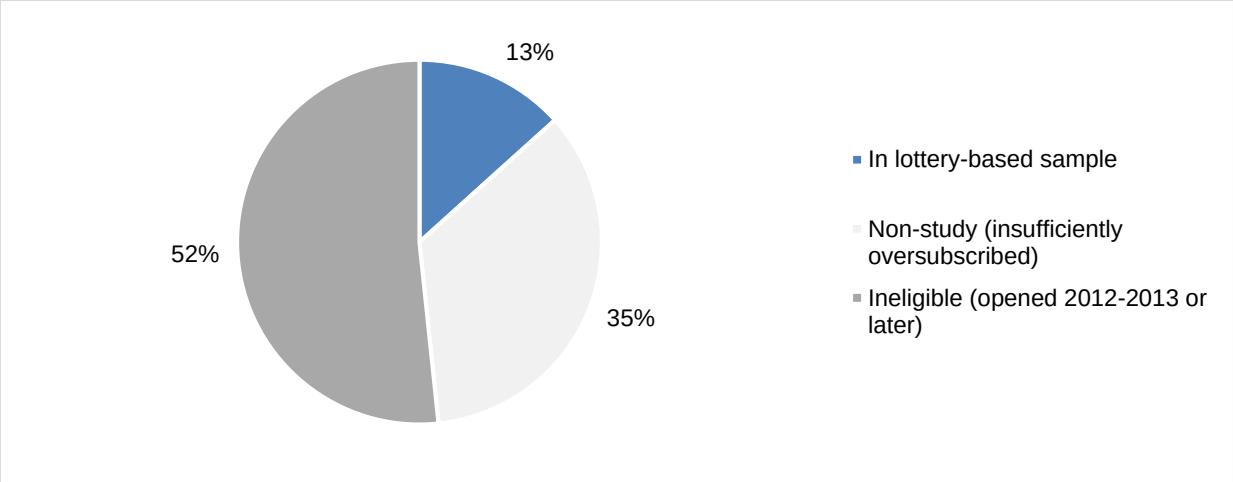
A. Elementary schools

At the elementary school level, we use our most rigorous approach: a **lottery-based design** capitalizing on school admissions lotteries that randomly determine which KIPP applicants are admitted, thereby creating a randomized experiment. Students winning admission to the KIPP school through the lottery form the study's treatment group; those with a poor lottery draw who are not offered admission form the control group. This design ensures that there are no systematic differences in treatment and control group students in terms of key baseline characteristics such as motivation and parental support, prior achievement, and demographics. At the time of admission to KIPP, treatment and control group students are distinguishable only by the luck of their lottery draws, which means that any subsequent differences in their outcomes can be attributed to the impact of attending KIPP.

1. Sample of schools and students

To be included in the lottery-based design, schools need to have substantially more applicants in an entry grade than they have seats to serve students (that is, they must be oversubscribed). In oversubscribed KIPP elementary schools, admissions lotteries are typically held at either pre-kindergarten for 3-year-old students (PK3) or kindergarten (K), depending on the school. Of the 29 KIPP elementary schools open in fall 2011, 8 were sufficiently oversubscribed to be in the analysis (3 with an entry grade of PK3, 5 with an entry grade of kindergarten).[4] Our elementary school sample thus includes 28 percent of KIPP elementary schools in operation during the 2011–2012 school year (the year for which the lotteries for admission occurred) and 13 percent of all KIPP elementary schools in operation by the end of the scale-up period in 2014–2015 (Figure III.2).

Figure II.2. Elementary schools in study sample (2014–2015)



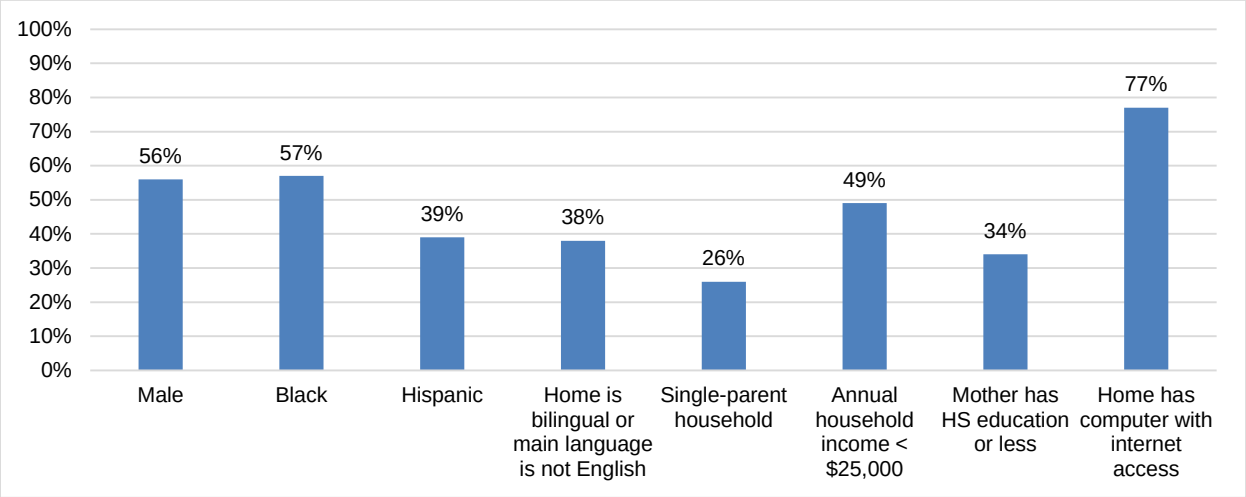
The treatment group comprises 473 students who participated in a lottery and won an offer of admission; the control group consists of 624 students who likewise participated in the lottery but did not receive an offer of admission. Not all students enrolling in a study school in the entry grade are included in the analysis, because some students were admitted outside the lottery (for example, if a student had a sibling already enrolled); in our sample of schools, 61 percent of open slots were filled via lottery. Lottery winners are included in the treatment group regardless of whether they ultimately enrolled in the KIPP school. Similarly, lottery losers are included in the control group regardless of whether they ultimately enrolled in a KIPP school. As a result, the lottery-based design produces estimates of the impact of an offer of admission to a KIPP school (typically referred to as an intent-to-treat, or ITT, estimate) rather than the impact of actually attending a KIPP school. Most lottery winners do attend a KIPP school, however, and most of those not offered admission never attend one. Among students in our sample, 79 percent of lottery winners (treatment group) attended the KIPP school to which they applied, while 6

[4] Of the remaining 21 schools, 11 also conducted lotteries for admission but either exhausted their waitlists (that is, made admissions offers to all lottery participants) or did not provide a sufficiently large treatment or control group. Another six opened in fall 2011 and did not conduct lotteries in spring 2011 (prior to opening).

percent of those who did not win an admissions lottery (control group) ended up attending a KIPP school.

Using information from a baseline survey of the parents of students applying to the KIPP elementary schools included in the lottery-based analysis, we can describe the average characteristics of the students those schools serve (Figure II.3). Most of the students in the elementary school sample are black or Hispanic (96 percent). Parents reported that a language other than English was the main language at home or that the home was bilingual for 38 percent of students, and 34 percent of students' mothers had completed a high school education or less. Approximately half of students are from households with annual incomes of $25,000 or less, and just over one quarter of students are from single-parent households. (Appendix D provides additional detail on sample members' characteristics, types of schools attended, and the baseline equivalence of our samples.)

Figure II.3. Characteristics of KIPP elementary school applicants



Notes: Data obtained from baseline survey of parents of applicants to KIPP elementary schools conducted in spring 2011. Sample includes data obtained for 422 students in the treatment group of the lottery-based baseline sample.

2. Data and methods

To measure elementary school students' academic achievement in reading and math, we administered the Woodcock-Johnson III (WJ-III) assessment in the spring of the third follow-up year after the lottery, when most students who applied at PK3 were in kindergarten and most who applied at kindergarten were in grade 2.[5] Students' WJ-III scores were standardized (into z-scores) using information on the performance of a nationally representative norming population.

[5] Based on grade-level appropriateness, we administered the Letter-Word Identification (Test 1) and Passage Comprehension (Test 9) tests in reading and the Calculation (Test 5) and Applied Problems (Test 10) tests in math. The Calculation test was administered only to the sample of students who participated in lotteries for schools with a kindergarten entry grade. We selected the WJ-III because, relative to other tests for this age-range, it (1) posed a low testing burden on young students in terms of the amount of time it takes to administer and (2) have a reliability for students ages 6 to 9 of over 0.90 for the reading tests and greater than 0.80 for the math tests (McGrew et al. 2007).

Thus, each student's score represents his or her achievement level relative to the national average for students at that grade level: scores greater than zero represent above-average achievement in the domain being tested and scores less than zero represent below-average achievement. Students' behavioral outcomes were measured using a survey of the parents of sampled students conducted in spring of the second follow-up year, when most students who applied at PK3 were in pre-kindergarten as 4-year-olds (PK4) and most students who applied at kindergarten were enrolled in grade 1.[6]

To calculate the impacts of KIPP elementary schools, we compare average outcomes for students in the treatment and control groups using a regression analysis that accounts for the baseline characteristics of sample members. Although a regression approach is not strictly necessary in a randomized experiment—we could simply compare the outcomes of the treatment and control groups—incorporating additional information on baseline characteristics can improve the precision of the results. For more detail on our lottery-based elementary school methodology, see Appendix D.

B. Middle schools

We use two different approaches for measuring the impacts of KIPP middle school on student outcomes. First, we implement the **lottery-based design** described above in sufficiently oversubscribed KIPP middle schools. Second, we employ a **matched-student design** in a broader set of KIPP middle schools, including some that are oversubscribed and some that are not. In this design, we identify a treatment group of students who enter KIPP middle schools in grade 5 or 6 and use propensity-score matching to define a comparison group of students—not attending a KIPP school—who most closely "match" the treatment group in terms of demographic characteristics and (most importantly) baseline test scores from grades prior to KIPP entry.[7] This approach explicitly accounts for differences between the two groups only to the extent that those differences are related to characteristics included in the matching process. Nonetheless, the approach has been validated, successfully replicating lottery-based results when baseline test scores are included among the characteristics used to select the comparison group (Tuttle et al. 2013; Fortson et al. 2015; Gill et al. forthcoming). Moreover, the matched-student design allows us to include many more KIPP middle schools in our sample.[8]

[6] In many cases, the surveys in our study included multiple items capturing the same underlying construct we wished to measure, so we created indices that summarize responses on related data items. We used principal component analysis to identify which group of related items to include in each index. Additional details on all indices are included in Appendix B.

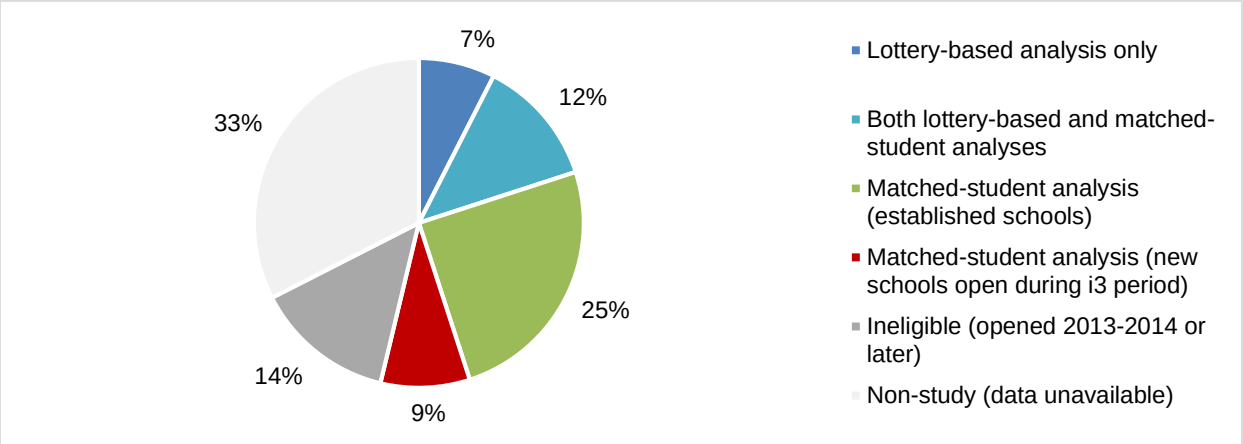
[7] This approach is not feasible for elementary schools in the KIPP network, since there are no available test scores for all students in a district prior to entry at kindergarten or pre-kindergarten.

[8] The matched-student sample includes a combination of schools that were included in our previous report on KIPP middle school impacts (Tuttle et al. 2013) and newer schools that could not be included in that report. Some middle schools that were included in the previous report could not be included here because we could not obtain current data for them. In total, 25 of 37 schools in this report were also included in the analyses for the prior report.

1. Sample of schools and students

Of the 60 KIPP middle schools open in 2011–2012, 16 (27 percent) were sufficiently oversubscribed to include in the lottery-based analysis.[9] We include 37 middle schools in the matched-student design, 10 of which are also in the lottery-based design. The matched-student design includes cohorts of students entering the KIPP middle schools in any year covered by our data.[10] In particular, this analysis includes between 2 and 13 cohorts of entrants, depending on the school.[11] Across the lottery-based and matched-student designs, our middle school sample therefore includes 53 percent of all KIPP middle schools in operation as of the end of the scale-up period in 2014–2015 (Figure II.4), including 7 of 21 new KIPP middle schools (33 percent) that opened during the scale-up period (the 2011–2012 school year or later). Of the other 14 new KIPP middle schools, 11 opened too late (fall 2013 or later) to be included in the matched-student analysis, which requires a minimum of two cohorts of students in each school.

Figure II.4. Middle schools in study sample (2014–2015)



The analytic sample in KIPP middle schools in the lottery-based design includes 891 students, comprising a treatment group of 459 students offered admission and a control group of 432 students not offered admission.[12] As in the elementary school lottery-based design, we measure the impact of admission to—rather than attendance at—a KIPP school, although admission and attendance are closely related. Among lottery participants, 72 percent of treatment group students and 5 percent of control group students attended a KIPP middle school. The matched-student analysis sample includes 40,624 students with valid test score outcome data.

[9] Of the remaining 43 schools, 19 also conducted lotteries for admission but either exhausted their waitlists or did not provide a sufficiently large treatment or control group for the analysis.

[10] Twenty-two KIPP middle schools open in spring 2011 were not included in the matched-student design because it was not possible to acquire the necessary data from the school districts or states in which those schools operated.

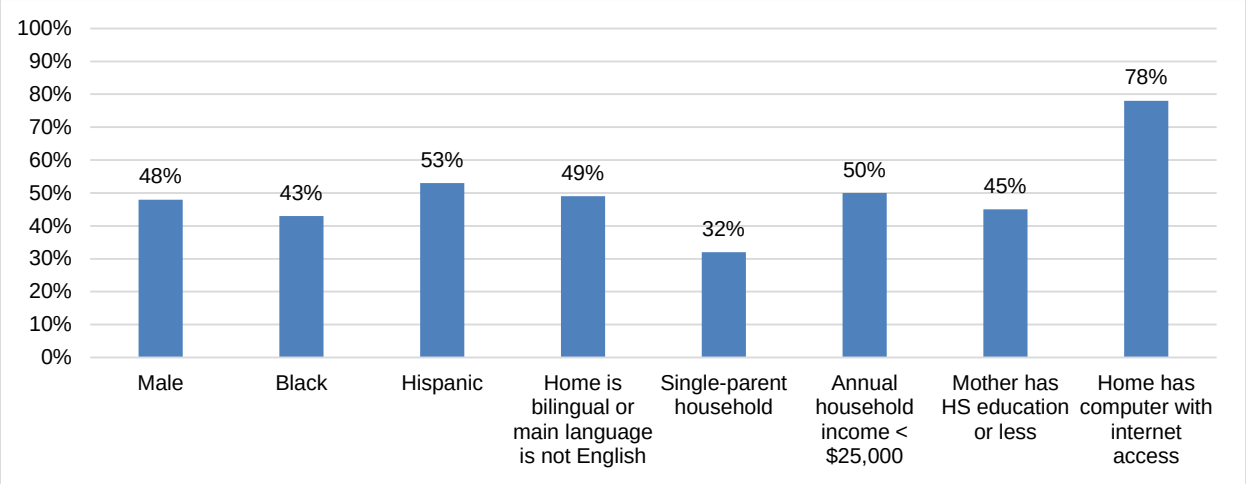
[11] Our outcome data represent the 2001–2002 through 2013–2014 school years, although the specific years and number of cohorts included vary by individual school (based on the years covered by the data provided by each jurisdiction as well as the year in which each school opened). Detail is provided, by school, in Appendix F.

[12] In the entry-grade lotteries for our sample schools, 68 percent of open slots were filled via the lottery; the remaining third of students in the entry grades in those schools were admitted outside the lottery and are therefore not included in the analytic sample.

The treatment group includes 20,312 students who entered a KIPP middle school included in the study at some point over the period 2001–2013; the comparison group of their matched counterparts who did not enter a KIPP school also includes 20,312 students, because we identified one comparison student for each KIPP student.

Applicants to KIPP middle schools in our study sample resemble applicants to KIPP elementary schools in our study sample based on observable characteristics (Figure II.5). Most students in the middle school lottery-based sample are Hispanic (53 percent) or black (43 percent), with less than 5 percent of students in other racial and ethnic categories. About half of students' parents reported that a language other than English was the main language at home or that the household was bilingual, and a similar proportion of students' mothers (45 percent) had completed a high school education or less. In addition, about half of students are from households with incomes of $25,000 or less, and roughly one quarter are from single-parent households. On average, students' math and reading scores in the year prior to the admission lottery were 0.12 and 0.21 standard deviations below their states' mean score, respectively. Appendix E provides additional detail on sample members' characteristics and evidence of baseline equivalence.

Figure II.5. Characteristics of KIPP middle school applicants



Notes: Data obtained from baseline survey of parents of applicants to KIPP middle schools conducted in spring 2011. Sample includes data obtained for 387 students in the treatment group of the lottery-based baseline sample.

Demographic characteristics of the middle school students in the treatment group of the matched-student sample are similar to those in the lottery-based sample; most students are Hispanic (47 percent) or black (51 percent). Most KIPP middle school students qualify for free or reduced-price lunch (89 percent), a proxy for having low family income. Almost 7 percent have special education needs and 10 percent are limited English proficiency students. On average, students' math and reading scores in the year prior to the admission lottery were 0.10 standard deviations below their states' mean score in both subjects. For a detailed discussion of the characteristics of KIPP middle school students relative to their surrounding districts, see Tuttle et al. (2013). Appendix F provides additional detail on sample members' characteristics and the baseline equivalence of our samples.

2. Data and methods

For middle schools in both designs (lottery-based and matched-student), we measured academic achievement scores on statewide assessments drawn from state- or district-provided administrative records. Students' scores on state tests were standardized (converted into z-scores) using statewide means and standard deviations, so scores represent students' achievement level relative to the typical student in the state at their grade level. We collected test score outcomes corresponding to the first, second, and third follow-up year after the lottery for the lottery-based sample and the first through the fourth year after the treatment group entered a KIPP school for the matched-student sample.

For the lottery-based design, we also measured a series of behavioral outcomes via surveys of students and their parents in spring of the second follow-up year, when most study students were in grade 6 (among those who applied for entry in grade 5) or grade 7 (among those who applied for grade 6). We did not measure survey-based outcomes for students in the matched-student design.

In both designs, the primary impact analysis compares average outcomes for the treatment and control groups using a regression approach that accounts for the baseline characteristics of sample members. In the lottery-based design, the main purpose of these control variables is to improve the statistical precision of the KIPP impact estimates. Because the matched-student analysis is quasi-experimental, these control variables play the more important role of accounting for any initial differences between the treatment and comparison groups before the former group entered a KIPP school. As described above, all students initially assigned to the treatment group in both the lottery-based and matched-student designs remained treatment group members throughout the analysis, regardless of whether they were enrolled at a KIPP school.

The specific regression analysis used for the matched-student design was established in our previous work and we also previously conducted a variety of sensitivity tests to check the robustness of the approach (Tuttle et al. 2013). These sensitivity tests were designed to address specific threats to the validity of our impact estimates; that is, possible reasons that our matched-student design could lead to biased or misleading estimates of the impact of KIPP middle schools, including:

- **Students leaving the district/state or moving to private schools (attrition from the sample).** Because the outcome measure is based on administrative records that cover only public school students in the district (or state, in some cases), students in our initial sample who moved to a private school or out of the jurisdiction are lost from our sample. If this occurs at a different rate among KIPP students than it does among comparison group students, this could lead to attrition bias in our impact estimates.
- **Students retained in grade.** When students are retained in grade, they take a different state test than others in their original cohort, making it difficult to compare the achievement levels of students who have been retained with those who have not.
- **Misspecification of the regression**. In the matched-student analysis, we rely on the covariates included in the regression to account for all of the relevant differences (aside

from KIPP attendance) that might influence student outcomes. Failing to properly specify this model could mean that the resulting impact estimates are biased.

The specification tests employed in Tuttle et al. (2013) provide evidence that our regression analysis addresses these threats to validity appropriately. The details of our approach to handling these issues, as well as other aspects of our methodology for the analyses of middle school impacts, are shown in Appendices E and F.

C. High schools

Unlike students in KIPP elementary and middle schools, most students do not enter a KIPP high school via a lottery. Among students enrolling in a KIPP high school in our sample, two-thirds (68 percent) came directly from attending a KIPP middle school in grade 8. As a result, we could not employ a lottery-based design at the high school level during the scale-up period, and instead use a quasi-experimental approach to measure the impacts of KIPP high schools. Moreover, since students can enter KIPP high schools via two routes—from KIPP middle schools and non-KIPP middle schools—we use two different quasi-experimental designs.

To measure the impact of KIPP high schools for the subset of students (32 percent) who enter the KIPP network for the first time in grade 9 (new entrants), we use a **matched-student design** similar to that described above for middle schools. In particular, we use propensity-score matching to identify a comparison group for these new entrants on the basis of demographic characteristics and baseline test scores from grades 7 and 8. As with the middle school matched-student design, we measure outcomes for all students in the initial sample for whom we have outcome data: treatment students who entered KIPP in grade 9 but ultimately moved back to a non-KIPP school are kept in the treatment group.[13]

To measure the marginal effect of KIPP high schools on student outcomes for those students who also attended a KIPP middle school ("continuing students"), we use a **matched-school design**. This design involves two different but complementary sets of comparisons, each comparing outcomes for a set of KIPP middle school students who had an option to attend a KIPP high school with outcomes for a similar set of KIPP middle school students who did not have an option to attend a KIPP high school. In each case, treatment and comparison group students are similar in that they all made the decision to enter a KIPP middle school, but they differ in that only the treatment group had the opportunity to enter a KIPP high school. Whether or not students had the option to enter a KIPP high school depends on the location and timing of their enrollment in KIPP middle schools—in some places and years, the KIPP high school option was present and in others it was absent. Specifically:

[13] For students who attended a KIPP high school after also attending a KIPP middle school, we can use a similar matched-student approach, but must conduct the matching at an earlier baseline (grade 4, prior to middle school entry and when both groups attended non-KIPP schools), as well as restricting the treatment group to students who entered a KIPP middle school and then persisted into a KIPP high school. This analysis provides a cumulative measure of the effects of attending both a KIPP middle and high school on outcomes at the high school level, but does not allow us to distinguish between the separate impacts of KIPP middle and high schools. For this reason, as well as other limitations of this approach, the results from this analysis and their limitations are not discussed in detail in Chapter III, but are included in Appendix C.

- For the first matched-school approach, we focus on a set of KIPP high schools in their first year of operation, and examine students graduating from the same KIPP middle school *in adjacent cohorts*. The treatment group in this design includes KIPP middle school students in grade 8 who had the option to attend the local KIPP high school in its first year of operation. The comparison group includes the previous cohort of KIPP students from the same middle school in grade 8 who did not have the option to attend the local KIPP high school because it had not yet opened. This design includes all five KIPP high schools that opened during the 2008–2009, 2009–2010, or 2010–2011 school years and that were served by KIPP middle schools with at least one cohort of students enrolling in grade 9 before the KIPP high school opened (that is, all schools with an eligible comparison group).
- For the second matched-school approach, the treatment group includes KIPP middle school students in grade 8 who had the option to attend the local KIPP high school. The comparison group includes KIPP students in grade 8 *from different middle schools* (in the same year) in regions with no KIPP high school open at the time. In other words, we compare students graduating from KIPP middle schools that fed into a KIPP high school with those graduating from KIPP middle schools that had no high school to feed into. To define a sample that was equivalent at baseline (grade 8), we identified the comparison KIPP middle schools that most resembled the treatment middle schools on the basis of average school-level characteristics. Then, within these matched sets of schools, we conducted student-level propensity score matching to identify the individual comparison students who most closely matched the treatment students.[14] For this design, we include all five KIPP high schools that opened prior to 2010 and were fed by KIPP middle schools with credible matched schools—two of which were also included in the approach using adjacent cohorts discussed above, for a total of eight unique schools in the matched-school designs.

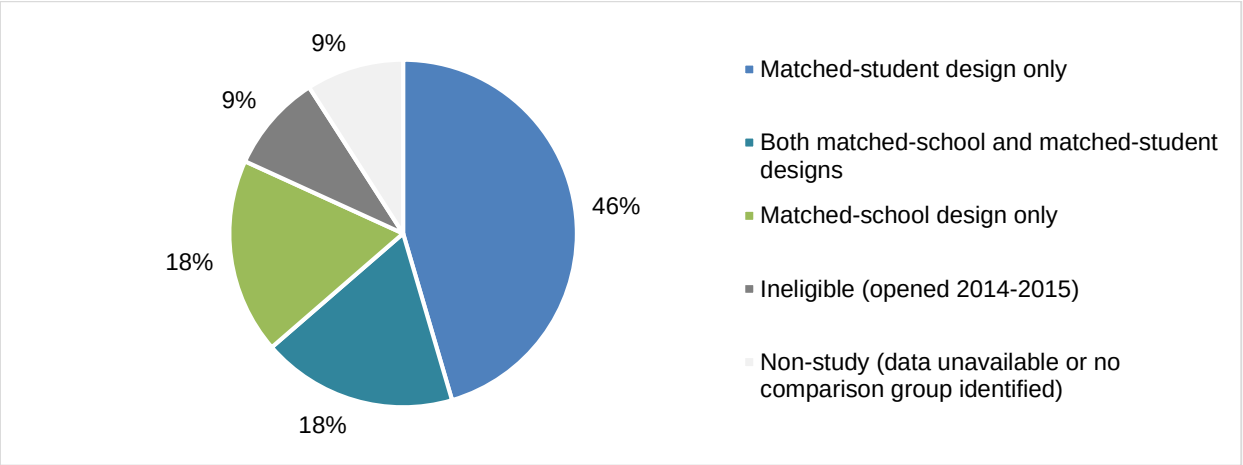
These designs assume that, aside from the presence/absence of the KIPP high school option, the treatment and control groups are similar to one another, on average. This assumption is based on the argument that the presence/absence of the high school option is largely beyond the control of individual students. The two matched-school designs ended up producing similar estimates of the impact of KIPP high schools for those high schools included in both approaches. Thus, we combine the results of the two matched-school designs for the purposes of this report. For more detail on this methodology, and results that are separate by model, see Appendix H.

1. Sample of schools and students

We include 14 KIPP high schools in our matched-student analysis of new entrants, and 8 high schools in our matched-school analysis of continuing KIPP students (including 4 that were also in the matched-student analysis). Across all the designs, our high school sample includes 82 percent of all KIPP high schools in operation as of the end of the scale-up period in 2014–2015 (Figure II.6). For information regarding the specific schools in these study samples, see Appendix A.

[14] Although we employed school-level matching to identify the eligible pool of comparison schools, we did not require a one-to-one match at the school level. Therefore, two different students from the same treatment school may each be matched to comparison students from different comparison schools, based on their individual characteristics.

Figure II.6. High schools in study sample (2014–2015)



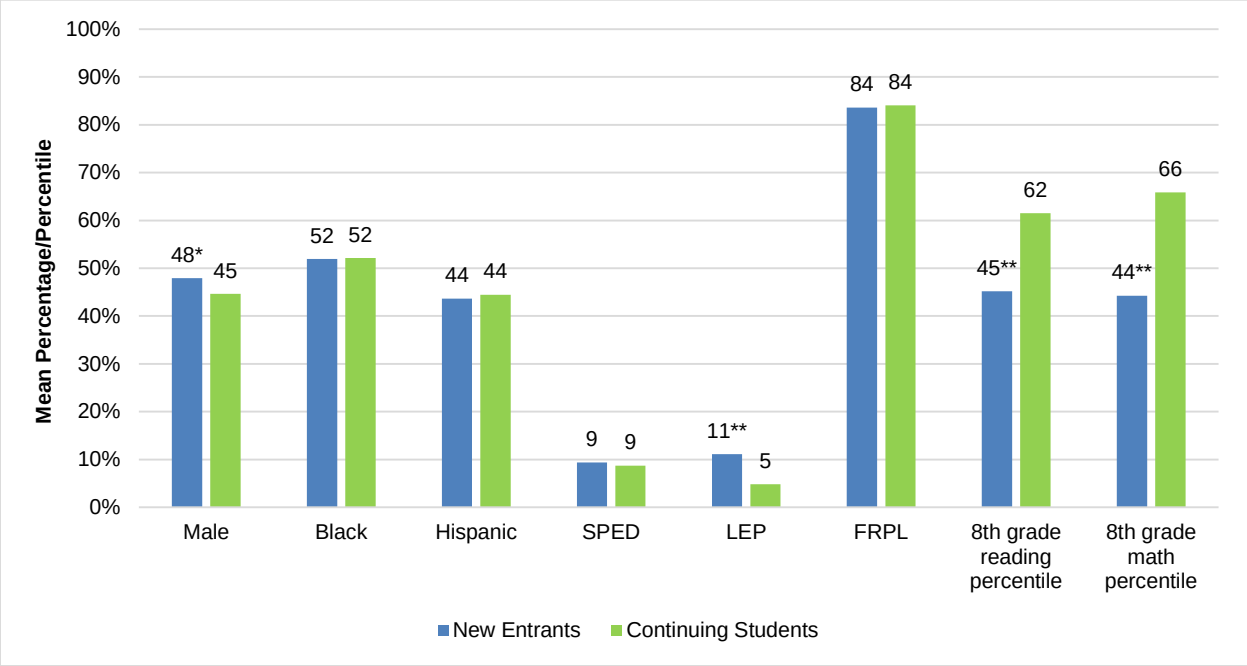
In the matched-student design of the impact of KIPP high schools for new entrants, we have a sample of 2,760 students for whom we have valid test score outcomes. The treatment group includes 1,380 students who entered one of the 14 KIPP high schools in the analysis for the first time in grade 9 at some point over the period 2005–2013. The comparison group of their matched counterparts includes 1,380 students who entered grade 9 at a non-KIPP school over this same period. In the matched-school design of the impact of KIPP high schools for continuing students, our sample includes 933 students with valid test score outcomes. The treatment group in this design includes 464 KIPP middle school students who had the opportunity to enter grade 9 in one of the 8 KIPP high schools in the analysis between fall 2007 and fall 2010. The comparison group in the matched-school design includes 469 KIPP middle school students who did not have the opportunity to enter a KIPP high school.[15] Although all treatment group students in the matched-student analysis entered a KIPP school initially, treatment group students in the matched-school analysis may or may not have entered the KIPP high school they had the option of entering. The matched-school analysis therefore measures the effect of the *availability* of a KIPP high school rather than *enrollment* in a KIPP high school. However, rates of enrollment in KIPP high schools among treatment group students are generally high, at 71 percent overall and ranging from 59 to 83 percent across the feeder middle schools for high schools in our sample. The comparison group attended a wide variety of non-KIPP high schools, including private, magnet, boarding, traditional public, or non-KIPP charter schools (see Appendix H for more detail).

The two different groups of KIPP high school students—new entrants and continuing students—are similar in several respects, but differ in important ways. Demographic characteristics of new entrants and continuing students are similar, with roughly equal proportions of black and Hispanic students, special education students, and students who qualify for free or reduced-price lunch (Figure II.8). New entrants are slightly more likely than continuing students to be male, and are twice as likely to have limited English proficiency. Most strikingly, the two groups differ with respect to their achievement level prior to high school

[15] In the matched-school sample, the size of the treatment group differs slightly from the size of the control group because students were not matched one-to-one in the matched-school design relying on adjacent cohorts.

entry. KIPP high school students who attended a KIPP middle school were much higher achieving prior to high school than those who are new entrants, with baseline (grade 8) scores 11 percentile points higher in reading and 16 points higher in math. As a result, although continuing KIPP high school students were higher achieving in grade 8 than students at non-KIPP high schools, new entrants into KIPP high schools had lower grade 8 achievement scores than non-KIPP high school students. Appendix G provides additional detail on sample members' characteristics and evidence of baseline equivalence for the matched-student analysis of new entrants, and detail on sample members' characteristics and evidence of baseline equivalence for the matched-school analysis of continuing students is included in Appendix H.

Figure II.8. Characteristics of continuing students and new entrants to KIPP high schools



Notes: Bars represent mean characteristics for two different types of KIPP high school students—continuing students and new entrants. Differences between the two KIPP samples are statistically significant at the 0.05 level (*) or the 0.01 level (**), two-tailed test, as noted by the asterisks in the graph. Each KIPP high school is given equal weight to calculate the overall average and statistical significance.

2. Data and methods

We used administrative data from jurisdictions (states or districts) hosting at least one KIPP high school to measure student achievement outcomes based on state assessments for the matched-student analysis of new entrants to KIPP high schools. State assessments tend to be administered differently in high schools than in middle schools. A majority of the high school tests are end-of-course exams in a single subject such as algebra I or biology that students may take at different points during their high school careers.[16] We obtained data on students' test scores from grade 9, 10, or 11, and the analysis is based on the first end-of-course test score in a

[16] A small number of jurisdictions provided data from end-of-grade exams administered to all students in a given grade.

given subject observed for each student (that is, we disregard retests for students who took the same subject in multiple years). We use as outcomes only those subject tests that were taken by a majority of students and by roughly equal proportions of KIPP and non-KIPP students, and where the timing of the test did not differ greatly between the treatment and matched comparison group.[17]

Our analytic approach for the matched-student analysis of new entrants to KIPP high schools mirrors our approach to the middle school matched-student analysis, with a few exceptions. When conducting propensity score matching, we matched the treatment group of new entrants to a comparison group of students in the same grade and district who did not go on to attend a KIPP high school but instead attended a non-KIPP public high school. We matched these groups on grade 7 and 8 test scores and demographic characteristics. As in the middle school analysis, KIPP students transferring out of KIPP high schools remained in the treatment group. For a detailed discussion of our analytic methods, including additional details about the tests administered, see Appendix G.

For the matched-school design of continuing KIPP students, we used a different data source to measure the student achievement outcome, because many students in this sample attended either private schools or schools in other states, such that our district or state administrative data did not include achievement test scores for these students. Thus, we measured achievement levels for students in this sample by administering a TerraNova assessment in the third follow-up year after high school entry (typically grade 11) to these students.[18] To measure outcomes reflecting student behavior and attitudes for students in the matched-school sample, we administered a student survey conducted in spring of the fourth follow-up year after high school entry (typically grade 12).

To estimate the impact of KIPP high schools on achievement and non-achievement outcomes for our matched-school designs, we compare average outcomes for the treatment and comparison groups using a regression that controls for baseline characteristics of matched sample members. See Appendix H for a detailed discussion of these methods.

[17] Because of the characteristics of high school assessments, it is important to remember that the impact estimates for each of these high school outcomes may not be directly comparable to the impact estimates for KIPP middle schools. For instance, there may be differences between the types of knowledge tested in high school end-of-course exams (which tend to be offered in multiple different grade levels) relative to the middle school exams (which have a separate test designed for each grade level). Similarly, high school exams tend to measure knowledge in academic areas (such as geometry or biology) that are more specialized than the general types of mathematics and literacy knowledge measured in statewide middle school tests.

[18] We administered Form G, Level 21/22 in reading, language, and math to the students in the sample, calculating z-scores that were standardized to capture student achievement relative to that of a nationally representative norming population.

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

III. KIPP'S IMPACTS ON ACHIEVEMENT AND GRADUATION

Network-wide, KIPP schools have positive, statistically significant, and educationally meaningful impacts on student achievement, particularly at the elementary and middle school grades. KIPP elementary schools have positive impacts on students' reading and math achievement. KIPP middle schools also have positive impacts on reading and math, as well as on achievement in science and social studies. For KIPP high schools, impacts on achievement are positive for some schools and groups of students, including new entrants to the network, but not statistically significant for others. We do not find evidence that KIPP high schools have statistically significant impacts on the probability that a student graduates in four years, but they do reduce the likelihood of dropping out of high school by a statistically significant amount.

Across the network as a whole, average impacts of KIPP at the middle school level (where the longest trends can be examined) have been consistently positive in both reading and math from 2005 through 2014, even as the system has grown rapidly. Average impacts have declined since 2007, but newly opened KIPP middle schools are producing positive impacts that are generally similar to those achieved by older KIPP middle schools when they were in their first years of operation. Older KIPP middle schools have shown some small declines in impacts, though they remain significantly positive.

A. Elementary school achievement

The KIPP network initially included only middle schools. The first KIPP elementary school opened in 2004–2005; by 2010–2011 when the i3 grant was awarded, there were 23 KIPP elementary schools. During the scale-up period, the number of KIPP elementary schools almost tripled, to 60 in operation as of the 2014–2015 school year, making them a much more important part of the KIPP network. No previous research has provided evidence on the effects of KIPP elementary schools.

Based on the randomized design described in Chapter II, we estimate the impact of being offered admission to a KIPP elementary school on students' early achievement in reading and mathematics, measured three years after the schools' admissions lotteries. As described previously, these outcomes are measured at different points in students' elementary school careers—at kindergarten for students who applied to enter KIPP at PK3 and in grade 2 for those who applied to enter KIPP in kindergarten. We find the following:

KIPP elementary schools have positive, statistically significant, and educationally meaningful impacts on three of four measures of students' reading and mathematics skills. On tests administered three years after entry, being offered admission to a KIPP elementary school has positive and statistically significant impacts on two measures of students' reading achievement (WJ-III Letter-Word Identification and Passage Comprehension) three years after random assignment (Table III.1). The Letter-Word Identification assessment requires students to orally name letters and words; in the Passage Comprehension assessment, students are required to read a printed passage and orally name a missing key word that makes sense in the context of

that passage.[19] Being offered admission to a KIPP elementary school is estimated to lead to an increase of 0.25 standard deviation units on the Letter-Word Identification test and 0.22 on the Passage Comprehension test.[20] The impact of an offer of admission on Letter-Word Identification is approximately equivalent to a student moving from the 78th percentile (the percentile corresponding to the control group students' mean score) to the 84th percentile; the impact on Passage Comprehension scores is equivalent to a move from the 48th percentile to the 57th percentile.

Table III.1. Impacts of offer of admission to KIPP elementary school

Outcome (Z-Scores)	Mean, lottery winner	Mean, lottery loser	Impact estimate	Number of schools	Number of students
Math achievement					
Calculation [a]	0.48	0.20	0.28** (0.11)	5	371
Applied Problems	0.04	-0.03	0.07 (0.05)	8	652
Reading achievement					
Letter-Word Identification	1.01	0.76	0.25** (0.07)	8	651
Passage Comprehension	0.18	-0.04	0.22** (0.07)	8	648

Notes: Outcomes are measured on Woodcock-Johnson III, administered in the spring of the third follow-up year. All impacts in this table are intent-to-treat (ITT), based on regression models that pool all lottery elementary schools and that control for baseline covariates. Standard errors are in parentheses. Measures of the complier average causal effect (CACE, sometimes referred to as a treatment-on-treated or TOT estimate) for each outcome are provided in Appendix D. Means for lottery losers are unadjusted; means for the lottery winners are equal to lottery losers' mean plus the estimated impact.

[a] Subtest administered only to students in grade 2 at the time of assessment.

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

In math, being offered admission to a KIPP elementary school has a positive and statistically significant impact on students' Calculation score (0.28 standard deviation units), but the impact on the Applied Problems score is smaller and not statistically significant (Table III.1). In the Calculation test, administered only to students in the grade 2 sample, students are required to solve printed mathematical calculations (for example, "3 - 1 = ?"). In the Applied Problems test, students are required to perform math calculations in response to orally presented problems.[21] The magnitude of the offer of admission to a KIPP elementary school on students' Calculation

[19] For kindergarteners, the Passage Comprehension test begins with a series of questions requiring the student to choose the appropriate rebus (representation of a word or phrase using pictures or symbols that suggest its syllables) based on an image, or choose the correct image based on a printed phrase.

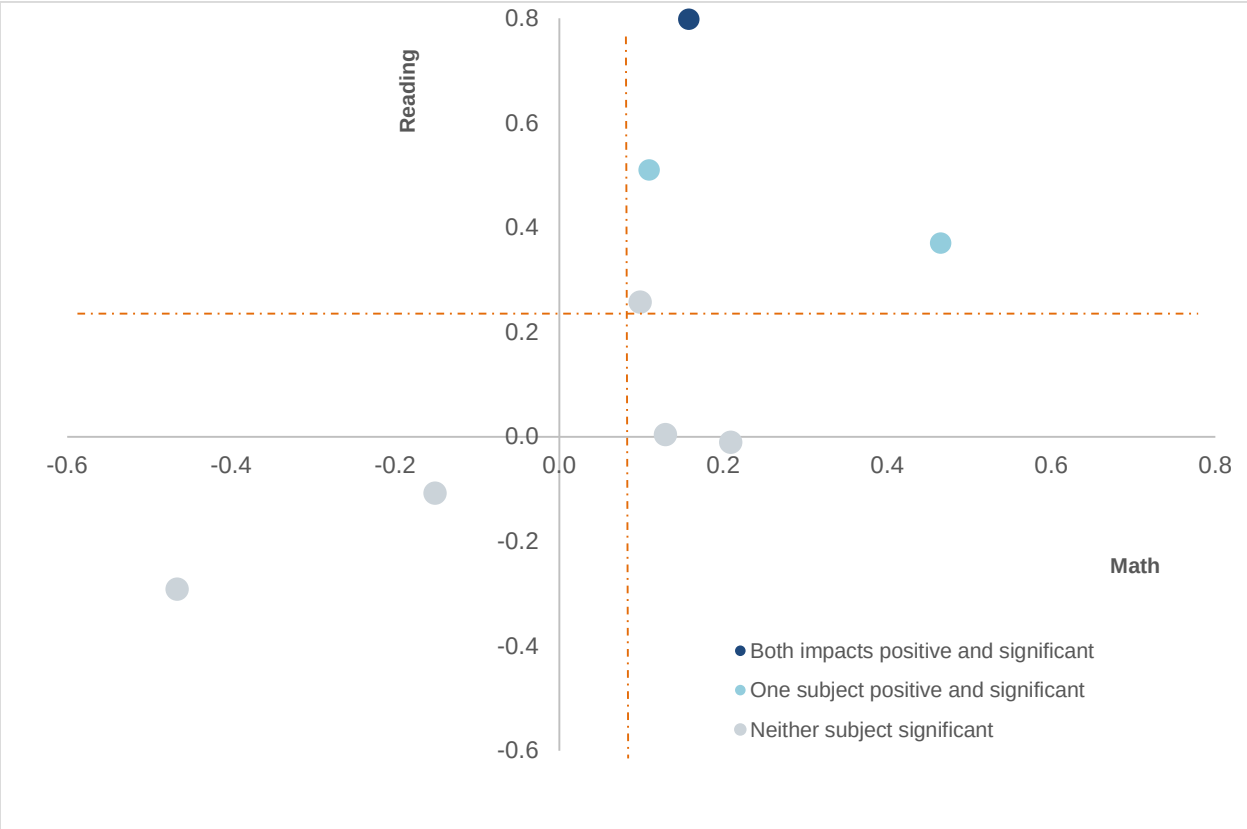
[20] Recall that these are intent-to-treat (ITT) estimates of the effect of an offer of an admission to a KIPP elementary school, regardless of whether the student ultimately enrolled in a KIPP elementary school. In other words, the treatment group includes students that did not actually attend KIPP schools.

[21] For kindergarteners, the test begins with a series of questions requiring the student to identify the number of instances an object appears in an image containing other objects.

scores is 0.28 standard deviation units, approximately equivalent to moving a student from the 58th percentile to the 68th percentile.

The impact of KIPP at the elementary school level varies by school. Among the eight KIPP elementary schools we examine, three have positive and statistically significant impacts on at least one measure of academic achievement (Figure III.1). Two additional schools have marginally significant positive impacts on one outcome ($p < 0.10$). There are no statistically significant impacts in the other three schools. In these schools, the point estimates for one school are positive and the estimates for two schools are negative in both subjects.

Figure III.1. Distribution of elementary school reading and math impact estimates after three years



Note: Each circle represents the average math impact estimate (across tests) and average reading impact estimate (across tests) for one KIPP elementary school. Dark blue circles indicate that at least one impact in both subjects is statistically significant and positive at the 0.05 level, two-tailed test. Light-blue circles indicate that at least one impact in only one of the two test subjects is statistically significant and positive. Grey indicates that all impacts are statistically indistinguishable from zero. No school had a statistically significant and negative impact in either subject in any test. The orange lines represent the average impacts across KIPP elementary schools.

B. Middle school achievement

Previous research finds that KIPP middle schools have positive impacts on student achievement, boosting the test scores of students above the levels they would have achieved had they attended non-KIPP schools (Gleason et al. 2013; Angrist et al. 2010; Woodworth et al. 2008). However, these studies measured KIPP impacts at a time when the network was limited in size, prior to the recent increase during the i3 scale-up period. The number of middle schools increased 33 percent between 2011–12 and 2014–15, while the number of schools across the entire network increased 51 percent over the same period.

The growth in the number of KIPP schools could present challenges for the network, making it difficult to maintain the positive impacts on student achievement found in prior studies of KIPP middle schools. When multiple KIPP middle schools serve the same district region, the network must recruit a larger number of students to attend KIPP middle schools; identify, recruit, and train a larger number of effective leaders; bring on board and retain a larger number of qualified teachers; and address bureaucratic issues that arise when a group of related schools must operate in the same educational market. Failing to adequately meet these challenges could result in a watering down of positive KIPP impacts.

A key objective of this report is to determine whether the KIPP network has maintained positive impacts among its middle schools. Do the existing schools, which have been successful in the past, produce equally positive impacts when faced with the challenges caused by a growing network? Are newly opened KIPP middle schools as successful as the schools KIPP has previously opened? Using both lottery-based and matched-student designs for measuring the impacts of KIPP middle schools, we find the following:

Consistent with prior research, KIPP middle schools have positive, statistically significant, and educationally meaningful impacts in math, reading, science, and social studies. Based on both study designs, KIPP middle schools have positive and statistically significant impacts on students' state test scores in math and reading, beginning two years after students are admitted (Table III.2).[22] For example, the lottery-based design suggests that being admitted to a KIPP middle school leads to an increase in students' average score in math of 0.24 student standard deviation units after two years, equivalent to a student moving from the 40th percentile to the 50th percentile in the distribution of his or her state math test. The two-year impact in reading of 0.18 standard deviation units is equivalent to a student moving from the 37th percentile to the 44th percentile in the distribution. The impact estimates are similar for the matched-student analysis based on a larger sample of 37 schools and multiple cohorts per school, suggesting that KIPP middle schools lead to an increase in average math scores of 0.23 standard deviations and reading scores of 0.10 standard deviations two years after admission.[23]

[22] See Appendices E and F for additional evidence on the impacts of KIPP middle schools on student achievement, including sensitivity tests that examine whether the impact estimates change with different modeling assumptions and methods.

[23] Note that our matched-student analyses weight students equally across cohorts for a given school.

Table III.2. Middle school impacts on state tests in math and reading

Outcome (Z-scores)	Years after admission 1	2	3	4
Middle schools (lottery-based analysis)				
Math achievement	0.10*	0.24**	0.18**	na
	(0.05)	(0.06)	(0.07)	
Reading achievement	0.03	0.18**	0.14*	na
	(0.05)	(0.05)	(0.06)	
Number of schools	15	15	14	
Number of students	608	563	458	
Middle schools (matched-student analysis)				
Math achievement	0.06**	0.23**	0.29**	0.27**
	(0.01)	(0.01)	(0.01)	(0.02)
Reading achievement	0.00	0.10**	0.15**	0.16**
	(0.01)	(0.01)	(0.01)	(0.01)
Number of schools	37	37	31	30
Number of students	36,798	29,386	23,433	14,425
New middle schools (matched-student analysis)[a]				
Math achievement	0.04	0.23**	na	na
	(0.02)	(0.04)		
Reading achievement	0.04	0.12**	na	na
	(0.02)	(0.03)		
Number of schools	7	7		
Number of students	2,471	1,205		

Source: State and district administrative records data

Note: Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Lottery-based estimates measure the impact of an offer of admission to a KIPP middle school (the ITT estimate) and are based on regression models that pool all lottery schools and control for baseline covariates. The impact estimates from the matched-student design use a similar regression model. A matched-student impact estimate was calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. Standard errors are in parentheses. na=not available.

[a] Estimates report the impact of KIPP on students who enrolled in a KIPP middle school that was founded in 2011–2012 or later.

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

The newer KIPP middle schools in our matched-student analysis—those that opened during the i3 grant period (fall 2011 or later)—have positive impacts on math and reading achievement that are of a similar magnitude as those of the overall middle school impact estimates. In year 2, for example, impacts on math test scores are 0.23 standard deviations among both newer KIPP middle schools and in the overall sample (Table III.2).[24] The magnitude

[24] The estimated impact in the overall sample of the matched-student design reflects the average estimated impact of the KIPP middle schools included in the sample, with each school weighted equally. An alternative approach for measuring impacts would have given greater weight to KIPP middle schools established earlier and that served a greater number of KIPP students (in other words, an approach that would have weighted schools according to the number of cohorts or number of students from the school in our data). Under that weighted scheme, the estimated impacts of KIPP middle schools based on the matched-student design would have been greater than 0.23.

of the impacts in reading in year 2 are also similar across the two samples (0.12 for newer schools compared to 0.10 for all middle schools). These results provide evidence that the new KIPP middle schools in our data appear to be just as effective in increasing student achievement as the typical middle school in the study, even though the new schools are in their first years of operation.

Because science and social studies tests are not administered annually during middle school, we based estimates of the impacts of KIPP middle schools on science and social studies achievement on a single score in each subject—the highest middle school grade where science or social studies scores could be observed for more than one cohort of KIPP students in our sample. On average, KIPP middle schools have a positive and statistically significant impact of 0.25 standard deviations in both science and social studies, equivalent to moving the average student from the 48th percentile to the 58th percentile in science and from the 51st to the 61st percentile in social studies (Table III.3).

Table III.3. Middle school impacts on state tests in science and social studies

Outcome (Z-scores)	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
Science achievement	0.07	-0.18	0.25** (0.01)	32	18,433
Social studies achievement	0.10	-0.15	0.25** (0.02)	18	10,440

Source: State and district administrative records data.

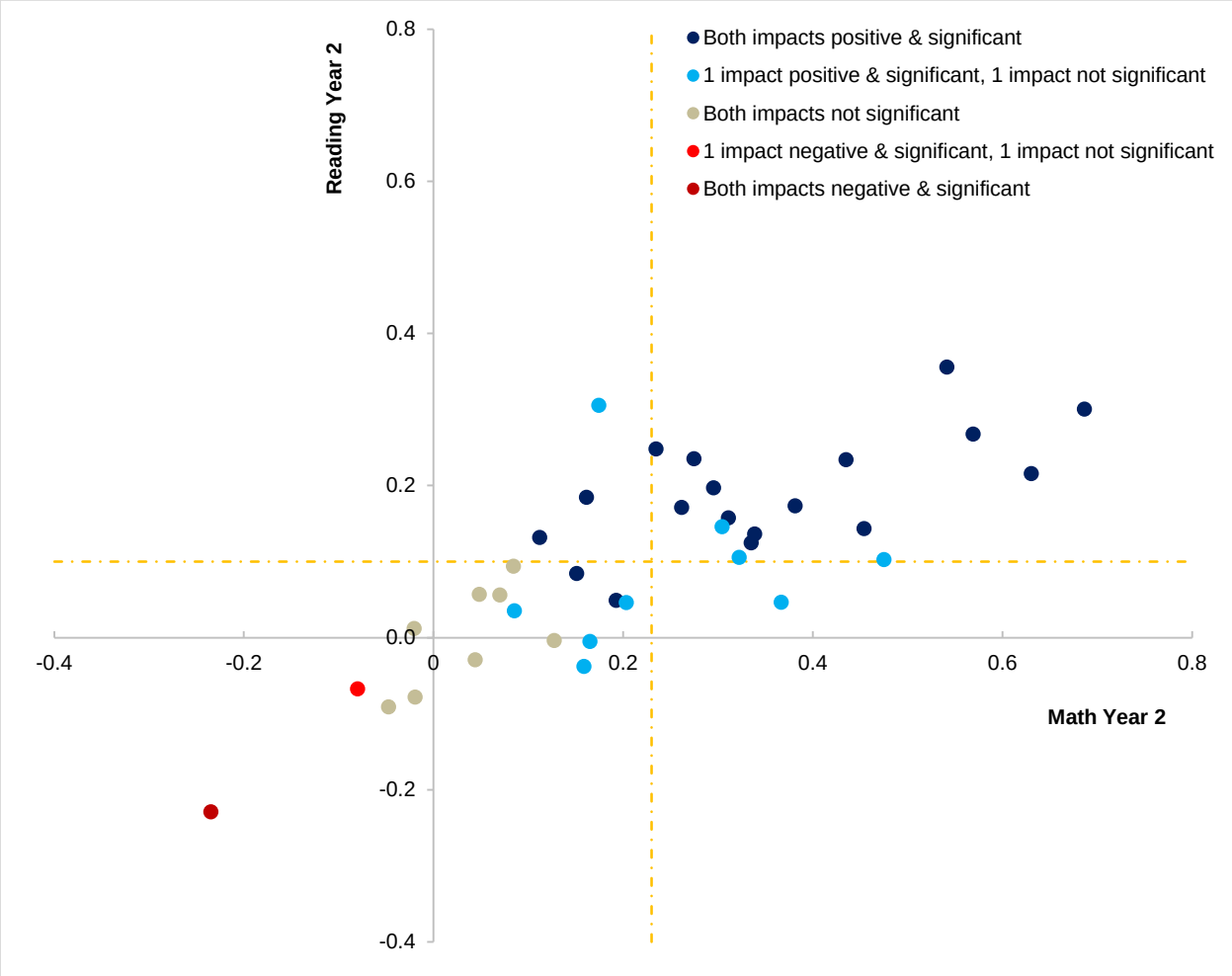
Notes: Impacts represent the cumulative effect of attending KIPP, not the marginal effect of an additional year of treatment. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. An impact estimate was calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. Standard errors are shown in parentheses. The grade level of middle school exams used for this analysis varied by jurisdiction. For each subject and site, we selected the highest middle school grade level where science or social studies was observed for more than one cohort of KIPP students.

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

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

We also examined the distribution of school-level impacts for our matched-student sample of 37 schools (Figure III.2). Each point plots one KIPP middle school's math and reading impact estimates two years after students enroll, based on all years covered by our data. Overall, 18 schools have positive and statistically significant impacts in both math and reading and nine schools have a positive and statistically significant impact in either math or reading (but not both). Eight schools have no statistically significant impacts, one school has a negative and statistically significant impact in only one subject, and one school has statistically significant negative impacts in both math and reading.

Figure III.2. Distribution of middle school reading and math impact estimates after two years



Notes: Impacts calculated separately for each middle school; each circle represents one middle school with two-year cumulative impacts plotted for reading (y-axis) and math (x-axis). Impact estimates are noted as statistically significant at the 0.05 level, two-tailed test. The dashed orange lines represent the average impact in math (vertical line) and reading (horizontal line).

C. Patterns of KIPP middle school impacts over time

Across the KIPP network, the average impacts of middle schools were positive and statistically significant throughout the 10-year period for which we have data, though higher in earlier years than recent years. To examine changes in the effectiveness of the network over time, we focused on trends at the middle school level, since the KIPP network has always included middle schools but added elementary and high schools only in recent years. Table III.4 shows estimates of the average impact of all operating KIPP middle schools where we have data *in each school year*, as measured by the schools' cumulative effect on students two years after initial KIPP enrollment. The number of KIPP middle schools included in the average impact increases from 8 schools in 2005 to 34 schools by 2014. KIPP middle schools have had positive and statistically significant impacts in both math and reading for all years from 2005 to 2014. KIPP-wide average

impacts were largest in 2007 and earlier (when no more than 17 schools were included in the sample), especially in math, ranging from 0.38 to 0.50 standard deviations, compared with 0.16 to 0.30 standard deviations between 2008 and 2014. In 2013 and 2014, when these two-year impacts fully reflect the performance of KIPP schools during the scale-up period and 30–34 schools are included in the sample, math impacts are 0.22 and 0.24 standard deviations. Differences over time in the average impacts of KIPP middle schools on reading achievement are less pronounced.

Table III.4. Impacts of KIPP middle schools on students two years after enrolling, by calendar year

Year	Math impact estimate	Reading impact estimate	Number of schools	Number of students	Average age of schools (years)
2005	0.38** (0.04)	0.11** (0.04)	8	848	2.9
2006	0.44** (0.03)	0.19** (0.03)	13	1,420	3.9
2007	0.43** (0.02)	0.21** (0.02)	17	1,908	4.1
2008	0.30** (0.02)	0.13** (0.02)	20	2,383	4.5
2009	0.28** (0.02)	0.17** (0.02)	23	2,974	4.9
2010	0.26** (0.02)	0.14** (0.02)	23	3,250	5.9
2011	0.26** (0.02)	0.09** (0.02)	20	2,933	6.0
2012	0.16** (0.02)	0.07** (0.02)	23	3,307	6.2
2013	0.22** (0.02)	0.09** (0.02)	30	4,298	6.9
2014	0.24** (0.01)	0.08** (0.01)	34	5,118	6.7

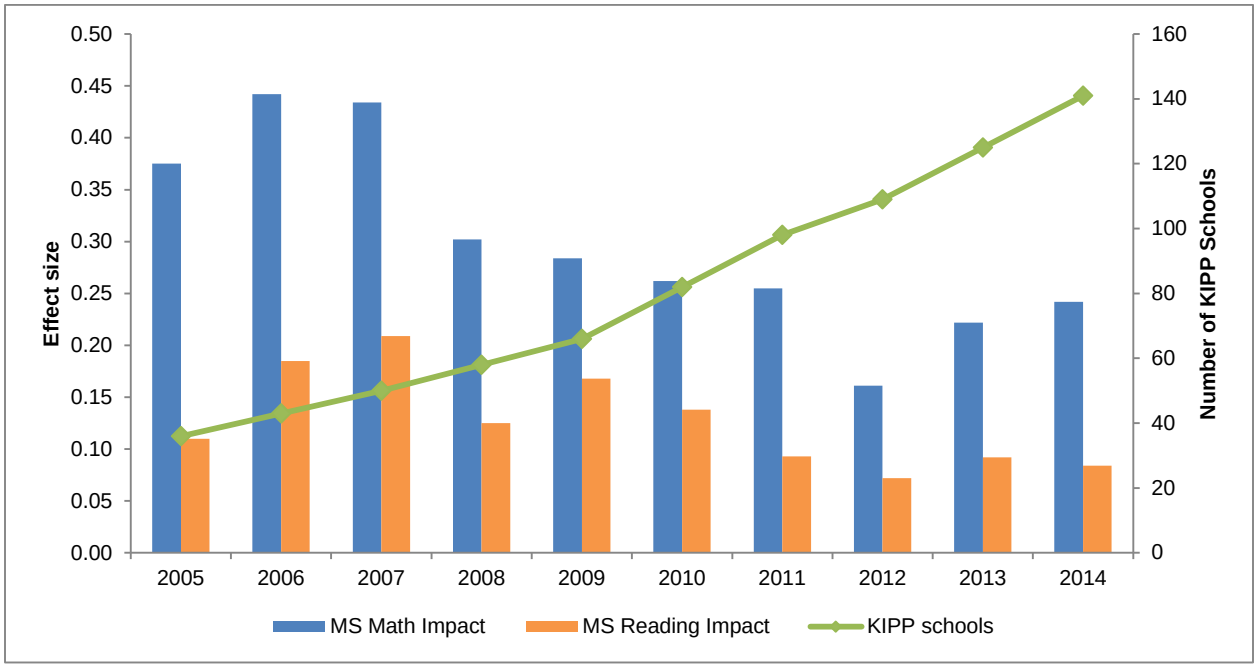
Notes: Impact estimates are estimated separately by school year; each table row includes an impact estimate based solely on the test scores of students affiliated with a KIPP middle school in that school year as compared to the scores of those students' matched comparison group in that school year. Impact estimates are calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. Standard errors are shown in parentheses. The year refers to the spring semester of the school year when the achievement exams were taken. Schools were omitted from the sample when fewer than 20 treatment students could be observed in a given year.

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

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

Several factors may explain these trends in impacts, including changes in the number and composition of schools in the sample, the relative performance of newer versus older schools, changes over time in the effectiveness of existing KIPP schools as the network has expanded, and changes in the effectiveness of schools attended by comparison students. Overall, KIPP's student achievement impacts moderated during a time of high growth in the network, although this moderation in impacts did not accelerate during the i3 scale-up period (Figure III.3). Although middle schools were part of the growth story, the largest expansion occurred in the number of KIPP elementary and high schools.

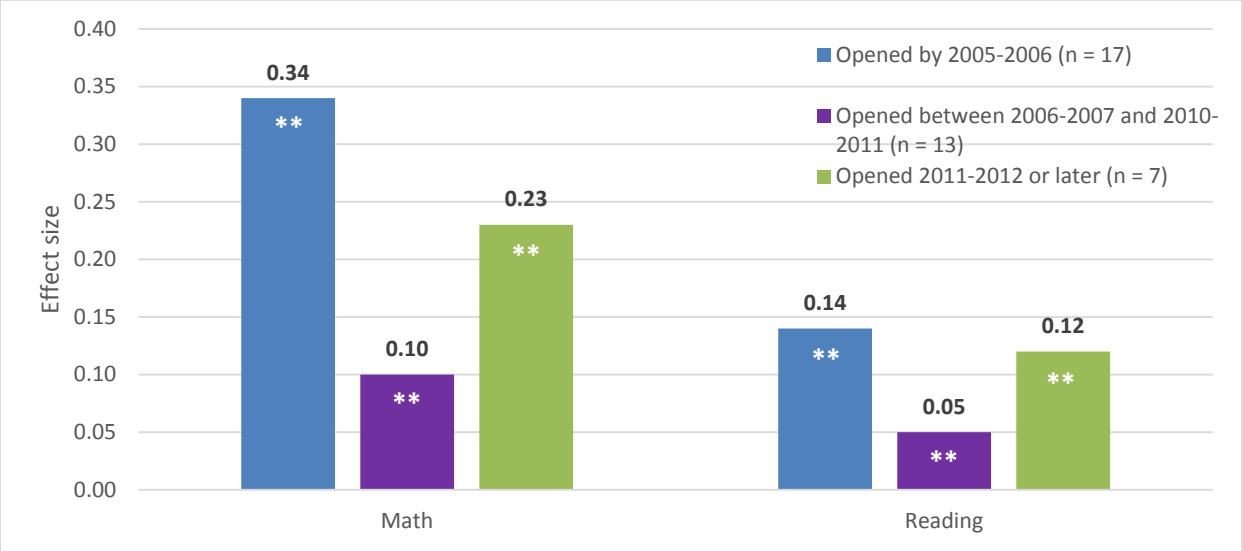
Figure III.3. Change in the size of the KIPP network and middle school impacts over time



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. Impact estimates are calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. They are estimated separately by school year and plotted using the left side y-axis. All impacts are statistically significant at the 0.01 level. The year refers to the spring semester of the school year when the achievement exams were taken. The size of the KIPP network is plotted against the right-side y-axis.

To further explore the overall trends in the network-wide estimates of KIPP impacts over time, we examined the relative performance of KIPP middle schools that opened at different points in the network's history. Figure III.4 presents impacts of KIPP schools (using all years of data to measure each school's effectiveness) separately for groups of schools opened during different periods. Average impacts are highest for the group of schools that opened first—in the 2005–2006 school year and earlier, and lowest for the group of schools that began operating between 2006–2007 and 2010–2011, when the network started significant expansion into the elementary and high school levels (by adding more than one school at each level per year beginning in 2006–2007) as well as into new cities. Schools that opened during the scale-up period—fall 2011 or later—fall in between, but the magnitude of their impacts is closer to the earliest KIPP schools.

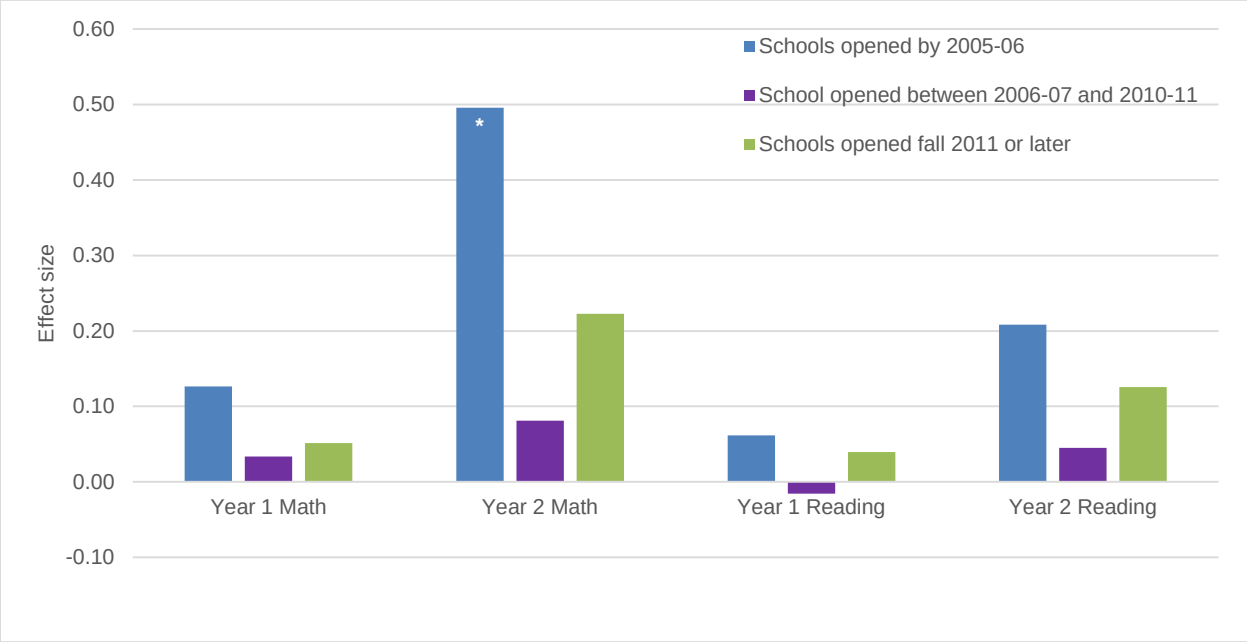
Figure III.4. Impacts of KIPP middle schools on students two years after enrolling, by year opened



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study, based on the year the school opened. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

KIPP middle schools that opened in fall 2011 or later—during the scale-up period—performed about the same in their first two years of operation as KIPP middle schools that opened prior to 2011, on average. The comparison shown in Figure III.4 is based on all years of data available for a given school, but a fairer way to compare more recently opened schools with established KIPP middle schools is to estimate their impacts in their first years of operation. Figure III.5 shows impacts in the first two years of operation for the oldest KIPP schools (opened by fall 2005), more recent KIPP schools (opened between fall 2006 and fall 2010) and schools that opened in the last few years (in fall 2011 or later). This analysis sheds light on whether schools opening during the period of KIPP network expansion facilitated by the i3 scale-up grant were any more or less effective from the outset than schools opening during earlier periods of KIPP growth. KIPP middle schools opened during the scale-up period had higher average impacts in math and reading than schools that opened between fall 2006 and fall 2010, but the differences were not statistically significant. The earliest KIPP schools had larger average student achievement impacts during their first two years of operation than new KIPP middle schools. However, only the year 2 math impact estimate is a statistically significant difference.

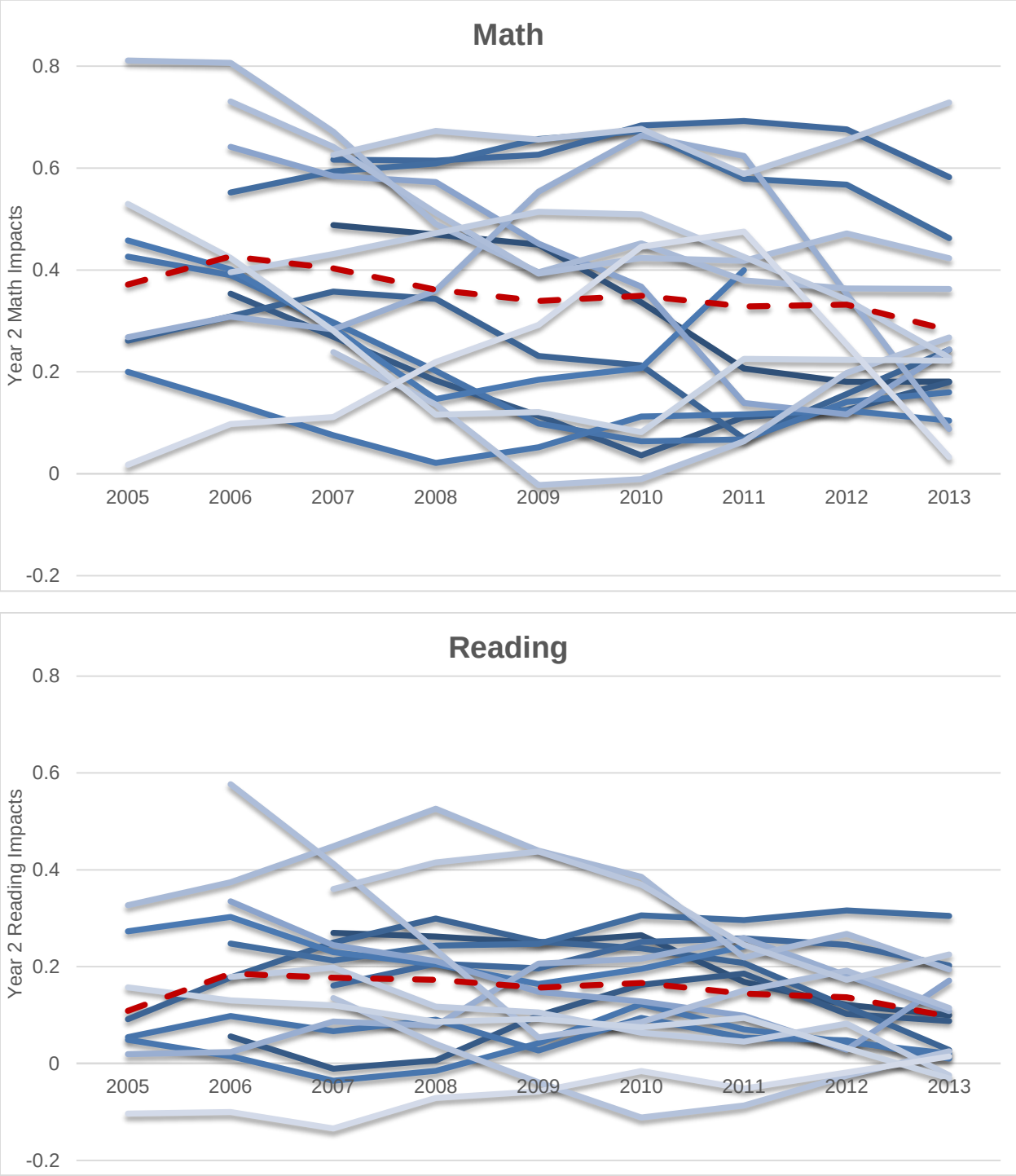
Figure III.5. Differences in achievement outcomes within first two years of operation



Notes: This figure reports the one-year and cumulative two-year impacts of KIPP on students who enrolled in a KIPP middle school during the first two years of that school's operation. Results are divided between schools opened by fall 2005, schools opened from fall 2006 through fall 2010, and schools opened since fall 2011 that are part of the study sample. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. The impact estimates are estimated separately by school and school year. The average impact estimates reported here assign an equal weight to each of the school-level impact estimates. The single statistically significant difference in impacts with the post-2011 schools (in year 2 math) at the .05 level, as measured by a two-tailed t-test, is noted with an asterisk (*).

Finally, we examined the school-level trends in our sample to see how the performance of individual schools affects the average annual impacts described above, focusing on the set of 17 schools open by 2005–2006 that we observe for several years. Because the sample size for a given school in a given year is small, we "smoothed" the trends in these figures by averaging the estimate in a given year with the school's impact estimate in the prior year and subsequent year (Figure III.6). Each shaded blue line represents the smoothed impact trend for a particular KIPP middle school (lines were shaded differently to make it easier to follow an individual line over time; different shades of blue do not signify differences in impacts). The dotted red line represents the average of all the individual smoothed school-level estimates. As shown in these figures, the overall trend in the average impacts of the network is relatively stable over time. There is a slight decline in both reading and math impacts as the schools (and the network) age.

Figure III.6. Achievement impacts of KIPP middle schools over time, two years after enrollment, by school (schools opened by fall 2005)



Notes: This figure reports school-level estimates of the cumulative two-year impact of KIPP on math (top) and reading (bottom) test scores for schools in our sample that were opened by the 2005-2006 school year.. The impact estimates are estimated separately by school year for each school in the study sample. Each blue line represents a different KIPP middle school and the dotted red line represents the average of all the blue lines. The smoothed impact estimates average the estimate in a given year with the school's impact estimate in the prior year and subsequent year.

D. High school achievement

KIPP opened its first high school in 2004–2005; as of the 2014–2015 school year, the network included 22 high schools. Unlike the first KIPP middle and elementary schools that opened, KIPP high schools were designed from the outset to serve students continuing within the network, so that students completing grade 8 at a KIPP middle school would be able to remain within the network throughout their high school years. In addition to these continuing students, the typical KIPP high school also serves some students entering the network from non-KIPP middle schools.

As was the case with KIPP elementary schools, KIPP high schools could not simply adopt the educational approach used by KIPP middle schools and apply it to older students. The curriculum at the high school level is different, student issues differ, and college preparation activities are especially important. Moreover, schools face the challenge of serving the two groups of students described above, continuing students and new entrants, who differ in terms of their familiarity with the KIPP culture and approach and may differ in terms of their initial level of academic preparation.

We measure the impacts of KIPP high schools on student achievement using two approaches, both of which are quasi-experimental. To estimate impacts of these schools on continuing students, we compare two groups of KIPP middle school graduates—those with and those without an available KIPP high school to enter upon completing middle school. To estimate impacts for new entrants to the network, we use a matched-student design, much like the matched-student design used to estimate impacts of KIPP middle schools.

The estimated impacts of KIPP high schools may also be influenced by the counterfactual condition—the set of non-KIPP students included in the analysis to represent what would have happened to the treatment group of KIPP students had they not been able to attend a KIPP high school. In the absence of a KIPP high school, network middle schools try to connect graduates with high schools most likely to encourage college preparation activities, and many graduates choose options other than traditional public high schools. In the matched-school comparison group, for example, 25 percent of students attend traditional public high schools, 38 percent attend non-KIPP charter high schools, 14 percent attend magnet schools, and 14 percent attend private or boarding schools. By comparison, in regions with KIPP high schools in our sample, an average of 70 percent of KIPP middle school students (the treatment group in our matched-school design) enroll in a KIPP high school, 5 percent attend other charter high schools, 3 percent attend private or boarding schools, 5 percent attend magnet schools, and 13 percent attend traditional public high schools. In the new entrant analysis, since the treatment group of new entrants come from non-KIPP middle schools in the district, we define the comparison group to be other students at non-KIPP middle schools who remain at non-KIPP public schools in their high school years—either traditional, charter, or magnet.

We find that the impacts of KIPP high schools differ for different groups of students and for different schools:

For new entrants to the network, KIPP high schools have positive, statistically significant, and educationally meaningful impacts on achievement in math, ELA, and science. Attending a KIPP high school boosts new entrants' high school math scores by 0.27 standard deviation units, a statistically significant impact representing an increase from the 48th to the 59th percentile for the typical student (Table III.5). Impacts in ELA and science are 0.18 and 0.31 standard deviations, respectively, and are also significant. Relative to outcomes for the matched comparison group, these impacts are equivalent to an increase from the 47th to 54th percentile in ELA and from the 42nd to 54th percentile in science. The average impact in social studies (0.01) is close to zero and not statistically significant.

Table III.5. Impacts of KIPP high school on achievement for new entrants

	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
Mathematics achievement	0.22	-0.04	0.27** (0.04)	12	1,489
ELA achievement	0.11	-0.07	0.17** (0.04)	14	1,861
Science achievement	0.11	-0.20	0.31** (0.04)	12	1,383
Social studies achievement	-0.14	-0.15	0.01 (0.06)	8	643

Source: State and district administrative records data.

Notes: Impacts were calculated separately for each KIPP high school. In a given high school, the outcome may be either an end-of-course exam (e.g., algebra), or an end-of-grade exam (e.g., grade 10 mathematics). Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Standard errors are shown in parentheses.

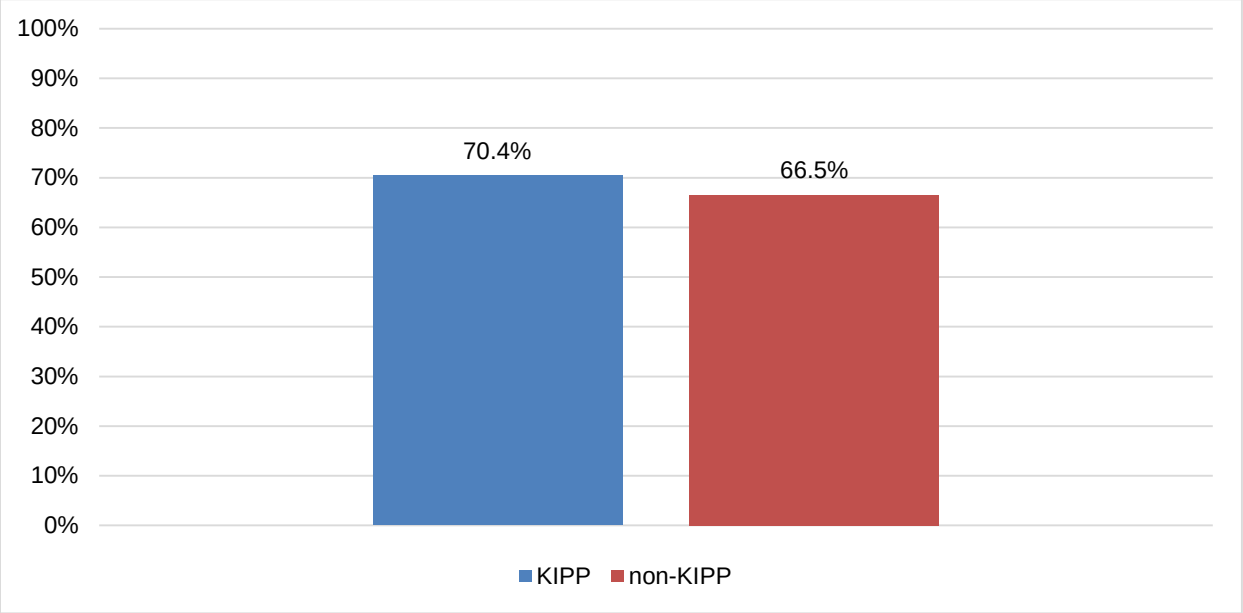
* Impact estimate is statistically significant at the 0.05 level, two-tailed test.

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

For new entrants to KIPP high schools, we also examine the probability of graduating within four years of entry. We find that this group of KIPP high schools did not significantly affect four-year graduation rates among new entrants (Figure III.7). The estimated impact on graduation rates is positive but small (4 percentage points) and statistically indistinguishable from zero.

There are two important limitations to the study's graduation analyses. First, the graduation indicator we use cannot distinguish between dropouts (who did not graduate) and students who left the data for some other reason such as transferring to private school or to a different school district (and who may or may not have graduated from high school). Both of these groups of students are classified as "non-graduates" in the analysis. Second, our propensity score matching approach relies on the assumption that the pre-KIPP characteristics of sample members observed in our data (middle school test scores and demographic attributes) fully capture attributes that are associated both with selection into KIPP and the outcomes of interest. In our graduation analyses, unlike our analyses of achievement, we have no baseline measure of the outcome being examined (graduation).

Figure III.7. Impact of KIPP high school on graduation among new entrants

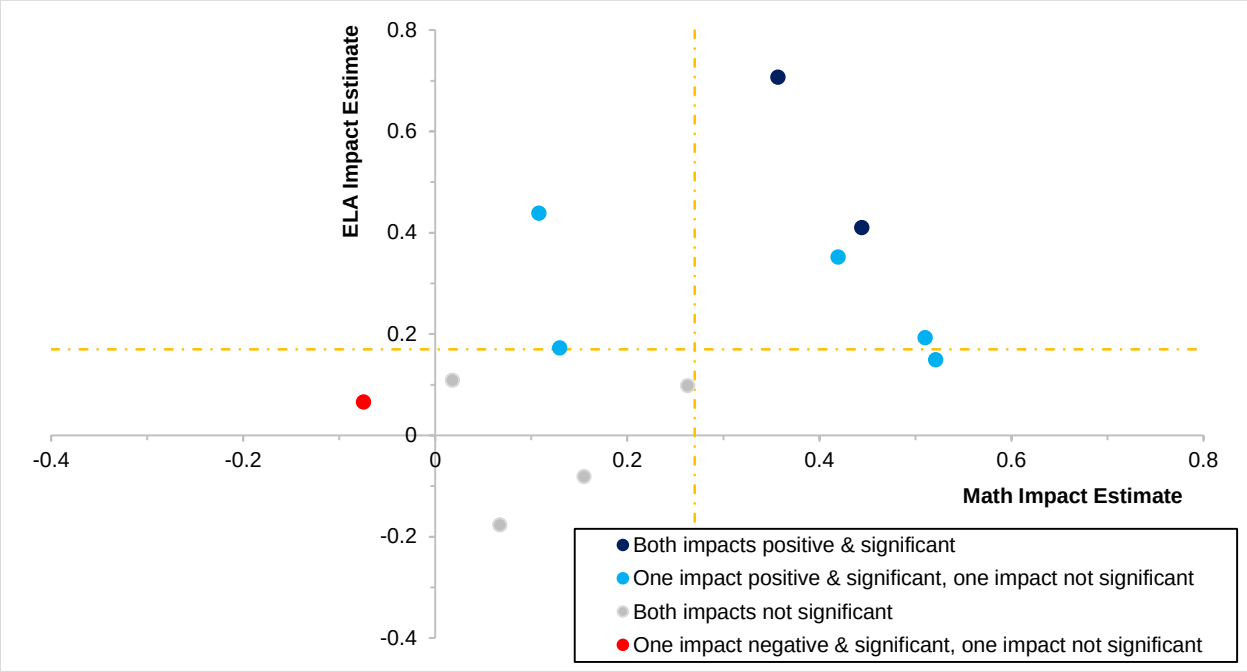


Notes: Impacts were calculated separately for the 7 KIPP high schools with available data for this outcome. Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Difference is not statistically significant at the 0.05 level, two-tailed test.

As is the case with both KIPP elementary and middle schools, the impact estimates for new entrants vary across the different KIPP high schools in our sample (Figure III.8). Of the 12 schools with both math and ELA impact estimates, two high schools have positive and statistically significant impacts on new entrants in both math and ELA, five schools have a positive and significant impact on new entrants in one subject but not the other, four schools have no significant impacts on new entrants in either subject, and one school has a negative and significant impact on new entrants in one subject.[25]

[25]The sample size for each school varies greatly, since different KIPP high schools have different numbers of cohorts and different numbers of students entering KIPP for the first time in grade 9 in our data. Thus, in some schools, impacts are estimated relatively imprecisely.

Figure III.8. Distribution of KIPP high school impacts on new entrants in ELA and math



Note: Each circle represents the math and ELA impact estimate for one KIPP high school. Dark blue circles indicate that impacts in both subjects are statistically significant and positive at the 0.05 level, two-tailed test. Light blue circles indicate that the impact in only one of the two test subjects is statistically significant and positive. Grey indicates that both impacts are statistically indistinguishable from zero. Light red circles indicate that the impact in only one of the two test subjects is statistically significant and negative. No school has a statistically significant and negative impact in both subjects. The dashed orange lines represent the average impact in math (vertical line) and ELA (horizontal line).

For students continuing from KIPP middle schools, the achievement impacts of KIPP high schools are not statistically significant on average, but these impacts vary by school. For continuing students, we measure KIPP high schools' impacts on student achievement in reading, language, and math using their scores on the TerraNova standardized assessment. These impacts are positive but small (none is larger than 0.07 standard deviation units) and not statistically significant (Table III.6).

Table III.6. Impacts of KIPP high schools on achievement for continuing students

	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
TerraNova Reading achievement	0.21	0.16	0.05 (0.05)	8	933
TerraNova Language achievement	0.04	0.01	0.03 (0.06)	8	933
TerraNova Math achievement	0.01	-0.06	0.07 (0.06)	8	933

Source: Study-administered test.

Notes: Impacts were calculated separately for each KIPP high school. Marginal impacts of opportunity to attend a KIPP high school are ITT estimates, measured on the TerraNova assessment. Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Standard errors are shown in parentheses.

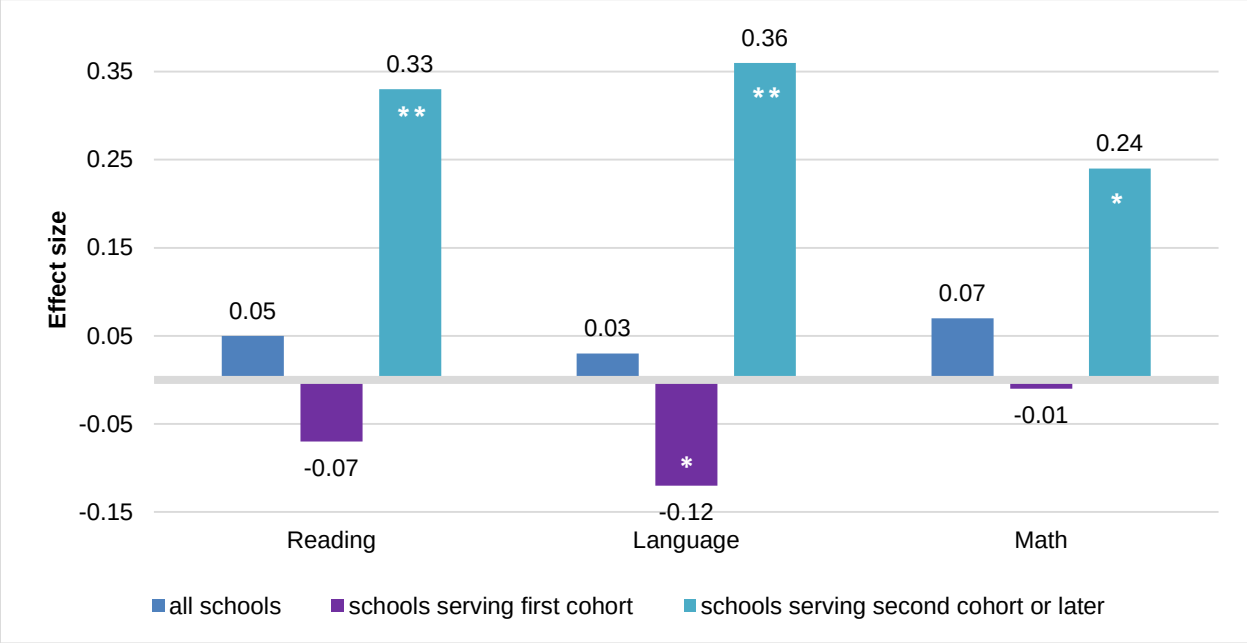
* Impact estimate is statistically significant at the 0.05 level, two-tailed test.

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

Five of the eight KIPP high schools included in this analysis were brand new and serving their first cohort of students when we measured their impacts. It turns out that there are significant differences between the estimated impacts of these new KIPP high schools versus those with more experience that are serving a later cohort of students. For the set of five new high schools where we observe only the first cohort of students, impacts are negative in all three subjects, and statistically significant in language (Figure III.9). For the three high schools where we observe only a later cohort, by contrast, impacts are positive and statistically significant in all three subjects, with magnitudes ranging from 0.24 to 0.36 standard deviation units. These more positive impacts for the more experienced high schools could imply that KIPP high schools become more effective as they gain experience. We examined this possibility in the high schools in the analysis of new entrants, but did not find clear evidence that impacts improved after the first cohort of students: impacts were somewhat lower for the first cohort than for later cohorts, but the difference was not statistically significant. The more positive impacts could result from some other characteristic of the more experienced KIPP high schools that distinguishes them from the new schools. Because we do not observe impacts of high schools in multiple years under this design, we cannot distinguish among these hypotheses.

Continuing students with the option to attend a KIPP high school are less likely to drop out of high school. The overall dropout rate in the entire sample is very low (only 15 students reported dropping out), but is significantly lower for the treatment group—1 percent for those who had the chance to attend a KIPP high school and 3 percent for those who did not. The magnitude of this impact (two percentage points) is similar to the magnitude of the impact of KIPP high schools on the likelihood of graduating from high school after four years among new entrants (four percentage points), but is measured more precisely.

Figure III.9. High school impacts for continuing students, by cohort examined



Notes: Model: Matched-school design. Outcome: TerraNova test. Sample size: eight schools; 933 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

1. Why do results differ for different analyses of KIPP high school impacts?

As we have shown, the matched-student analysis finds that KIPP high schools have positive impacts for new entrants to the network, whereas the matched-school analysis finds that, on average, impacts are not statistically significant for continuing students. There are four potential explanations for these differences, related to differences in the KIPP high schools included in the analyses, differences in the test-score measures, differences in the students, and differences in the experiences of the comparison groups.

First, the new entrant analysis includes impacts for a different set of high schools than the matched-school analysis (see Appendix A). Differences between the two sets of impact estimates could simply reflect differences between the effectiveness and age of various KIPP schools included in the analyses at the time outcomes were measured at each school. Although we did not find clear evidence that KIPP high schools improve with experience, data are not sufficient to fully assess whether high school age or other characteristics of the high schools included in the analyses might be driving the results.

Second, the assessments measuring student achievement differ in the two analyses. The type of learning measured by the TerraNova assessment (the outcome for continuing students in the matched-school analysis) may differ from that measured by standardized end-of-course subject tests (the outcome for new entrants in the matched-student analysis). In addition, we administered the TerraNova assessment to students three years after entering high school (usually in grade 11), whereas the timing of the state exams varied by jurisdiction and exam but were administered as early as grade 9 in some cases.

Third, the fact that two analyses are measuring impacts for different sets of students—new entrants to KIPP versus continuing students—may explain these differences. KIPP high schools may have more positive impacts for students with the kinds of characteristics typical of new entrants than for students with characteristics typical of continuing students. As discussed in Chapter II, new entrants have much lower baseline (grade 8) test scores than continuing students. If KIPP high schools are delivering a model that is more effective for students with lower prior achievement, this could help to explain why the impact estimates for new grade 9 entrants are more substantial than the impacts for students who already had extensive exposure to KIPP in middle school. A related explanation is that continuing students may have already benefited substantially from their time in a KIPP middle school, so attending KIPP during high school represents a smaller change for them than it does for new entrants.[26]

Fourth, estimated impacts may differ due to differences in the experiences of the non-KIPP comparison students in the two analyses. In the matched-student analysis, new entrants to KIPP are being compared to students who attended non-KIPP middle schools in the district (including non-KIPP charter schools) and remained in non-KIPP schools in their high school years. In the matched-school analysis, continuing students (those with an opportunity to attend a KIPP high school) are compared with those who did not have that opportunity, the majority of whom (at least 66 percent) attended private or boarding schools, magnet schools, and non-KIPP charter schools in the district. If those schools tend to be more effective than the high schools attended by comparison students in the matched-student analysis, then the KIPP schools would show a smaller impact even if they were equally effective relative to similar comparison schools.

[26] Another question is whether the new entrants to KIPP high schools might benefit from peer effects derived from attending high school alongside the higher-scoring students who arrived from a KIPP middle school. Available evidence from the literature suggests that peer effects are less pronounced during high school than in earlier grades, with a standard deviation increase in achievement of peers (measured against the student-level distribution of test scores) associated with an increase in a student's score of 0.01 to 0.06 standard deviations, depending on the sample and methods used (Lavy et al. 2012; Burke and Sass 2013). In our sample, new entrants are surrounded by peers with middle school scores 0.42 and 0.56 standard deviations higher than their own scores in reading and math, respectively. Based on peer effects reported in the literature, this could produce benefits for new entrants of about 0.00 to 0.03 standard deviations in reading and math. Our impact estimates for new entrants are much larger than this, suggesting that peer effects are unlikely to play a large role in explaining impacts for new entrants.

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

IV. KIPP'S IMPACTS ON STUDENT BEHAVIOR AND ATTITUDES

A student's long-term success depends on more than test scores, so we are interested in whether KIPP schools have effects on student outcomes other than academic achievement. We use data from surveys of students and their parents to investigate how attending a KIPP school influences student behavior, motivation and engagement, school experiences and satisfaction, educational goals and aspirations, and—for high school students—college preparation activities.

Findings on behavioral measures in this chapter should be interpreted with caution. Several outcomes may be affected by reference bias or by the fact that different groups of students or parents may have a consistently different frame of reference when answering survey questions about their attitudes, behavior, and experiences.[27] As a result, when comparing KIPP and non-KIPP students, differences in some outcomes may arise from differences between the groups in their frames of reference (related, for example, to differences in peer groups in KIPP schools and other schools) rather than differences in actual attitudes, behavior, and experiences. Reference bias should be less problematic for objective measures, such as courses high school students report having taken or other specific college preparation activities.

In many cases, the surveys included multiple items capturing the same underlying construct, so we created indices that summarize responses on related data items. We used principal components analysis to identify which group of related items to include in each index. For example, the index indicating the extent to which a student is well adjusted to school represents a parent's average response on seven items related to his or her child's adjustment to various aspects of school, such as getting along with others, liking school, working hard at school, and respecting adults. Details on all indices are presented in Appendix B.

We rely on the same designs for measuring KIPP schools' impacts on student behavior and attitudes as for measuring impacts on achievement outcomes, with one caveat. We did not have the opportunity to administer surveys to students in the matched-student analyses (or to their parents) at the middle school and high school levels. Thus, impacts on these outcomes are based on the lottery-based, randomized design at the elementary and middle school levels and on the matched-school design at the high school level. At the middle school level, we administered surveys to students and their parents in spring of the second follow-up year. Given the age of the students in the other samples, we administered surveys to parents only at the elementary school level (in spring of the second follow-up year) and to students only at the high school level (in spring of the fourth follow-up year).

As with the achievement impacts, the results presented here represent the effect of being offered admission to a KIPP school (that is, winning the lottery) at the elementary and middle school levels, and the effect of initial enrollment in a KIPP high school. Elementary and middle school students are included in the treatment group regardless of whether they actually enrolled in a KIPP school; high school students who initially enrolled at a KIPP school are included in the

[27] For further study on reference bias among educational surveys, see West et al. (2014). In addition, Duckworth and Yeager (2015) provide a review of the literature on limitations on measuring noncognitive outcomes for purposes of evaluating educational interventions.

treatment group even if they subsequently transferred out. These are therefore conservative estimates of KIPP's impact on students who remain enrolled.

Table IV.1 provides an overview of the results, based on separate analyses for KIPP elementary, middle, and high schools for a large number of student behavior and attitude outcomes. In each cell of the table, we report the number of outcomes examined for that level of KIPP school, as well as the number of those impact results that were statistically significant (at the 5 percent level). Cells shaded dark green represent domain-by-grade combinations where at least three outcomes (or all outcomes in cases with fewer than three outcomes examined) are statistically significant, suggesting a positive effect overall. Cells shaded light green include one or two statistically significant positive effects. More detailed tables showing these impacts are presented in Appendix B.

Table IV.1. Summary of significant effects, by domain

Domain	Elementary	Middle	High
Motivation and engagement Time spent on homework; indices of engagement, self-control, academic confidence, grit, and effort in school	1 of 5	2 of 10	0 of 8
Experiences and satisfaction Parent and student feelings about school, perceptions of classmates and teachers, school disciplinary environment, academic difficulty, parent outreach, parent engagement	3 of 9	6 of 14	0 of 8
Behavior Peer pressure, undesirable behaviors, positive behaviors, illegal activities, adjustment to school, parental concerns	0 of 1	0 of 11	--
Goals/aspirations Expectations for on-time high school graduation, college attendance, and college completion	1 of 3	0 of 7	0 of 3
College prep activities Discussions about college, assistance in planning for college, college testing, application to college	--	--	7 of 16
Coursetaking Years of coursework, number of AP/IB/honors courses	--	--	8 of 15

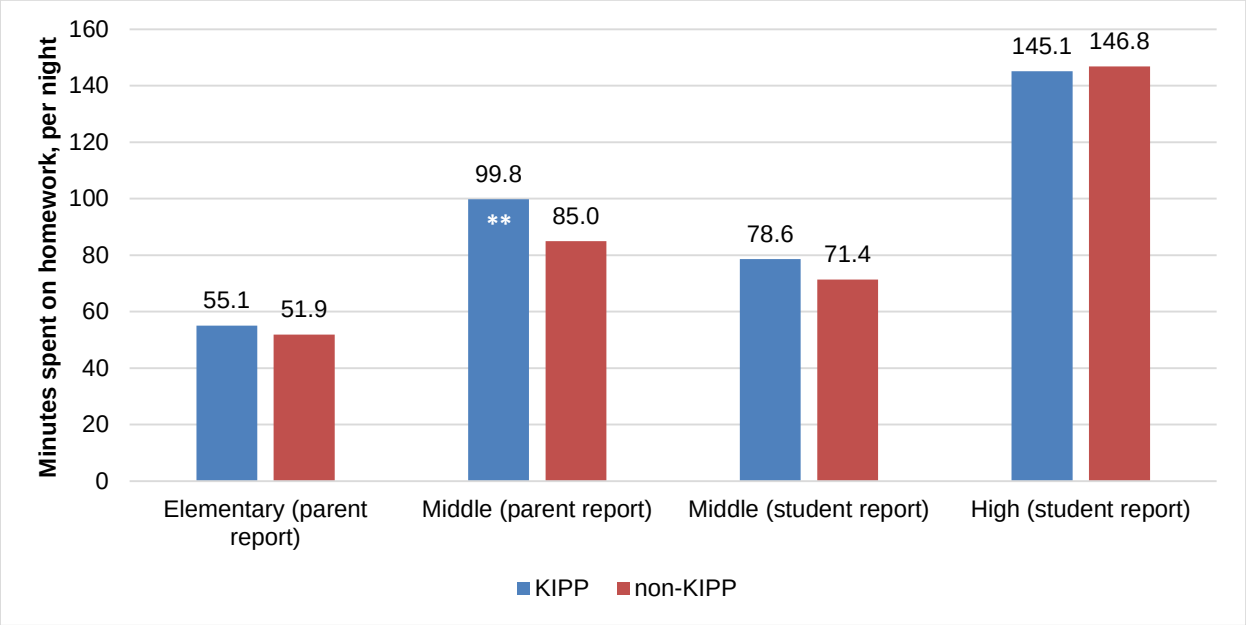
Notes: Each cell indicates the number of statistically significant outcomes from the total number of outcomes measured in a given domain and grade level. Cells shaded dark green represent domain-by-grade combinations where at least three outcomes (or all outcomes in cases with fewer than three outcomes examined) are statistically significant, suggesting a positive effect overall. Cells shaded light green include one or two statistically significant positive effects. Specific results, by outcome, are included in Appendix B.

Across grade levels, KIPP schools have no statistically significant impact on most measures of student motivation and engagement, behavior, or educational aspirations. By and large, we find no statistically significant impact of KIPP elementary, middle, or high schools on most measures included in three of the six categories of non-achievement outcomes we examined—student motivation and engagement, behavior, and goals and aspirations. In the motivation and engagement category, KIPP does not significantly affect measures of student

self-control, academic motivation, academic confidence, grit, school engagement (positive or negative), or effort in school, including student reports of the time spent on homework.

There are a few exceptions to the lack of impacts on outcomes in the motivation and engagement category. At the elementary school level, parents of KIPP students report a higher frequency of positive comments from the student about school and looking forward to going to school. At the middle school level, KIPP has a positive and significant impact on an index of a student's academic collaboration (how much a student enjoys working with, and learns while helping out, fellow students). The estimated impact of KIPP is statistically significant for only one other outcome measure in this category, the amount of time middle-school students spend on homework, according to their parents (Figure IV.2). Parents of treatment group students in middle school report that their children spend an average of almost 15 more minutes on homework each night compared with parents of control group students. This positive impact estimate is not supported by other evidence on homework completion. According to student reports, the impact of KIPP middle schools on homework time is about half as large (about seven minutes) and not statistically significant. Further, parents of KIPP middle school students do not report that their children are any more likely to complete their homework. Note that this result is different from estimates from our earlier study of KIPP middle schools (Tuttle et al. 2013), in which the impacts on time spent on homework were positive and significant according to both the student and parent reports. Further, in that study, the number of minutes spent was higher for both treatment and control groups (118 and 96 minutes, respectively, according to parent reports). At the elementary and high school levels, the estimated impact of KIPP on students' homework time is not statistically significant, regardless of the specific measure examined.

Figure IV.2. KIPP impacts on time spent on homework, by grade level



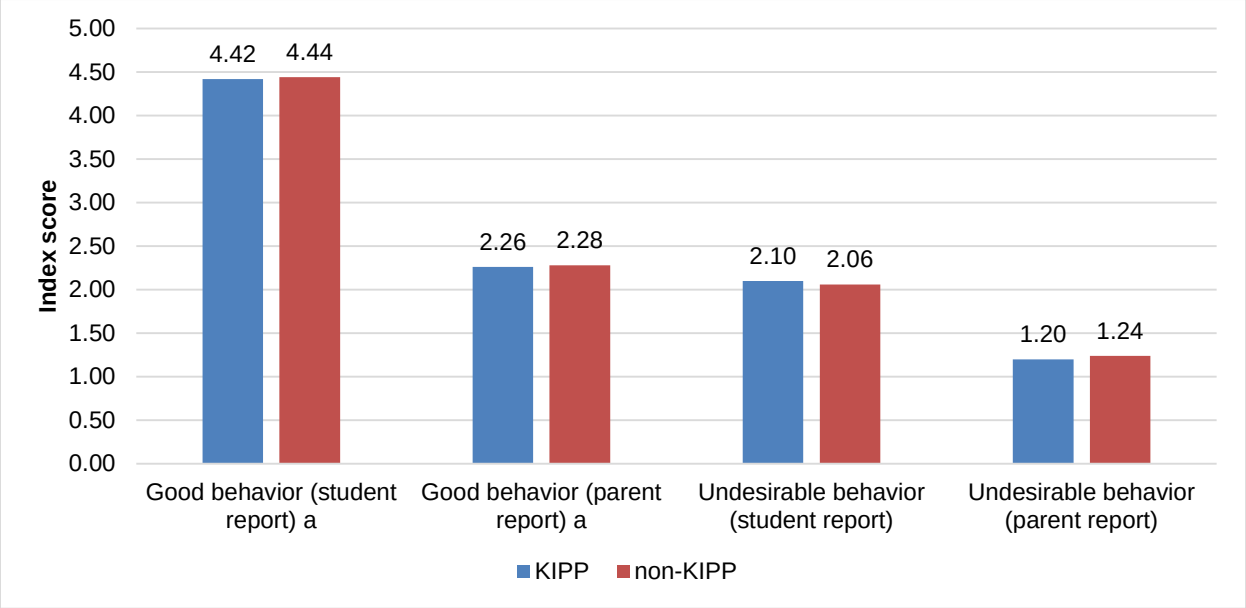
*Significantly different from zero at the .05 level, two-tailed test.

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

Similarly, we find no evidence that attending a KIPP school affects student behavior. At the elementary school level, we measured behavior using an index of items indicating the extent to which the student is well adjusted to school, and find that KIPP does not significantly affect this outcome. We used a range of measures of behavior at the middle school level, including indices of positive behaviors, undesirable behaviors, peer pressure, illegal activities, parental concerns about their child, frequency of school disciplinary actions (according to the parent), and the extent to which the child is well adjusted. In none of these cases is the impact of KIPP statistically significant. For example, Figure IV.3 shows estimates of the mean value of the indices of positive and undesirable behaviors for KIPP (treatment group) and non-KIPP (control group) middle school students, from both parent and student reports. The mean values of each of these student behavior outcomes for the KIPP and non-KIPP groups are nearly identical. This finding is in contrast to estimates from Tuttle et al. (2013), in which we found that KIPP schools led to increases in two indicators of student-reported undesirable behavior.

We measured educational goals and aspirations using responses from parents and students. In general, the educational goals and aspirations among these elementary, middle, and high school students are high in both the treatment (KIPP) and comparison (non-KIPP) groups. At the high school level, for example, 84 percent of students report that they think they will graduate from college. For 12 of 13 outcomes, the estimated impact of KIPP is not statistically significant. The single exception is among parents of students at KIPP elementary schools, who are 10 percentage points more likely than the comparison group to believe their child is very likely to complete college (81 versus 71 percent).

Figure IV.3. KIPP middle school impacts on behavior

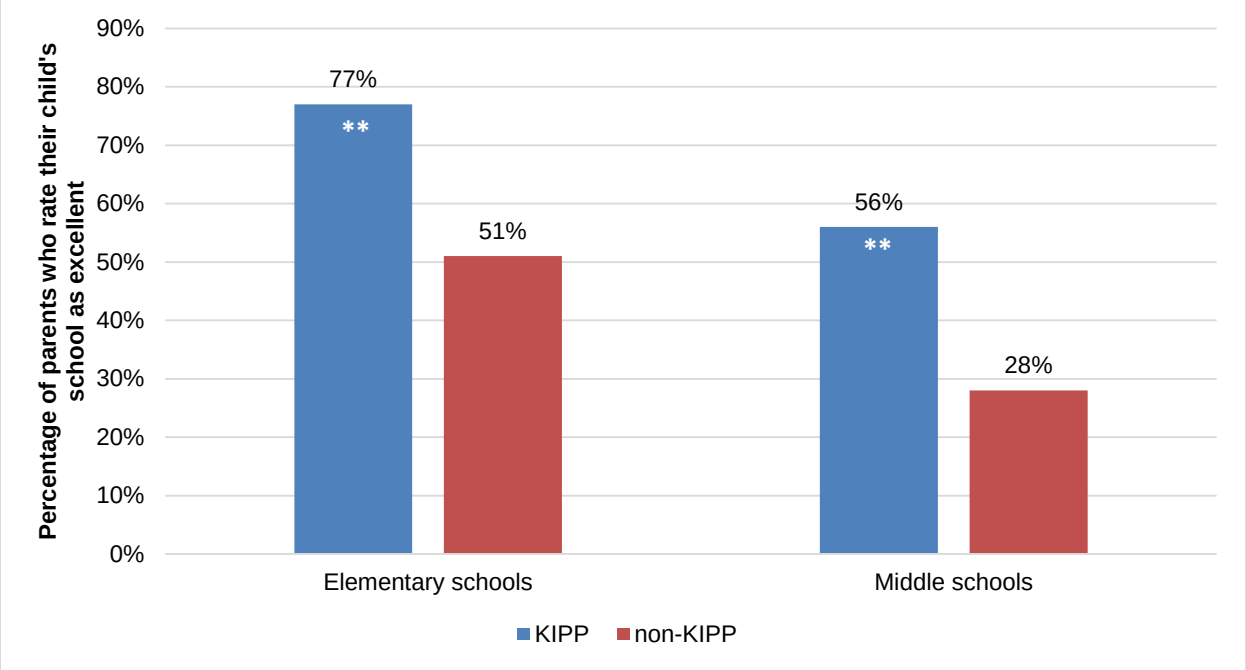


Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information. Indices based on student reports are measured on a scale of 1 (once or twice per year) to 5 (almost every day) and indices based on parent reports are measures on a scale of 1 (never) to 3 (often). None of the differences shown in the figure is statistically significant.

[a] Index has an alpha smaller than 0.7, indicating low reliability

KIPP elementary and middle schools have positive impacts on school satisfaction, particularly among parents. At both the elementary and middle school levels, being offered admission to KIPP leads to increases in parents' satisfaction with their child's school. At the elementary school level, for example, more than three-quarters of parents in the treatment group rate their child's school as excellent, compared to about half of parents in the control group (Figure IV.4). At the middle school level, 56 percent of treatment group parents and 28 percent of control group parents rate the school as excellent. These findings are consistent with previous research on KIPP in particular and oversubscribed charter middle schools in general (Tuttle et al. 2013; Gleason et al. 2010). Similarly, KIPP has significant positive impacts on a parent-based index capturing satisfaction with school facilities, academics, safety, and discipline. KIPP also has significant positive impacts on several other satisfaction measures, including indices of school efforts to engage parents at both the elementary and middle school level; an index of parental perceptions of problems in their child's middle school; and middle school students' perceptions of their schoolmates (see Appendix B). Evidence of KIPP impacts on satisfaction do not extend to the high school level, however, as none of eight measures of student-reported satisfaction at the high school level is statistically significant.

Figure IV.4. KIPP impacts on parental satisfaction



Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

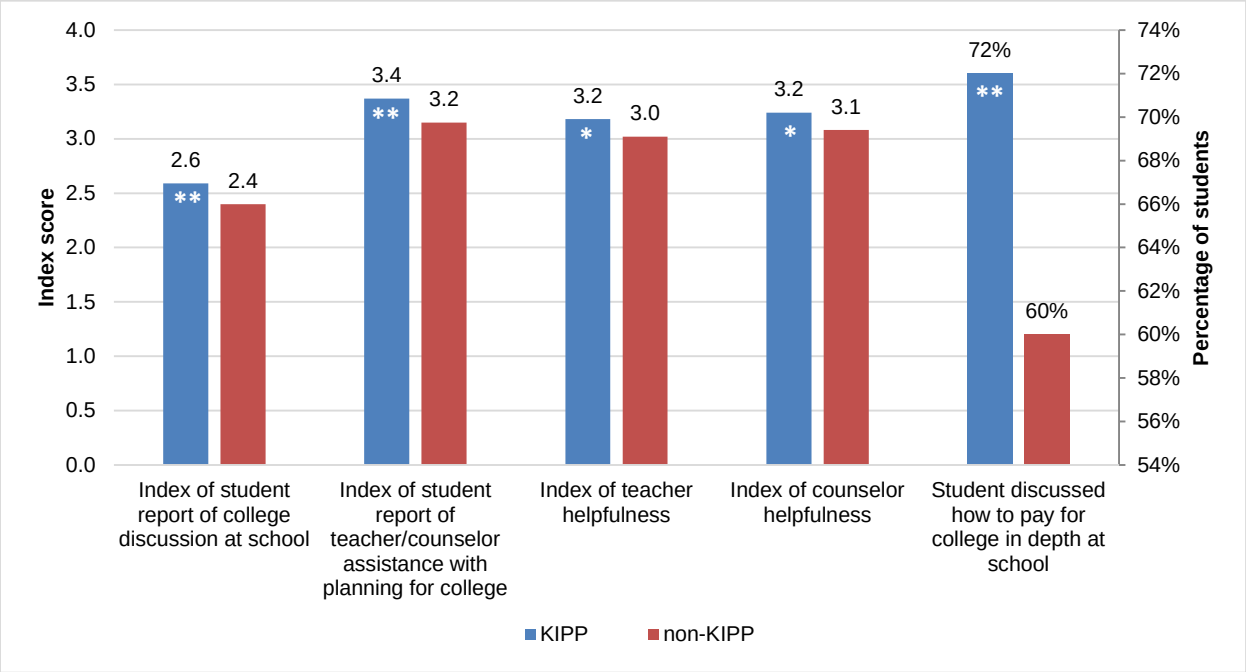
*Significantly different from zero at the .05 level, two-tailed test.

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

KIPP high schools have positive effects on several aspects of college preparation, including discussions about college, applying to college, and coursetaking. Based on surveys administered in spring of the fourth follow-up year—senior year for most students—KIPP high schools have positive and significant impacts on three measures related to school assistance in

planning for college. First, KIPP has a positive impact on the frequency of discussions about college at school, based on an index of student reports. This outcome includes indicators of whether school staff discussed with students nine different topics related to college admissions and readiness (Figure IV.5). Second, KIPP leads to students being more likely have in-depth discussions at school about how to pay for college, with 72 percent of treatment students reporting having these discussions compared to only 60 percent of comparison students. Third, KIPP high schools have a positive impact on an index summarizing student reports of teacher or counselor assistance with planning for college. This finding is supported by estimates of positive KIPP impacts on student reports of the helpfulness and encouragement of their school's teachers and counselors on various aspects related to college preparation (including help in selecting courses that meet graduation requirements and are needed for college admission, help in finding a suitable college and deciding what to do after graduation, encouragement to take AP/honors courses, and encouragement to continue their education through college and beyond).

Figure IV.5. KIPP high school impacts on assistance for college planning



Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

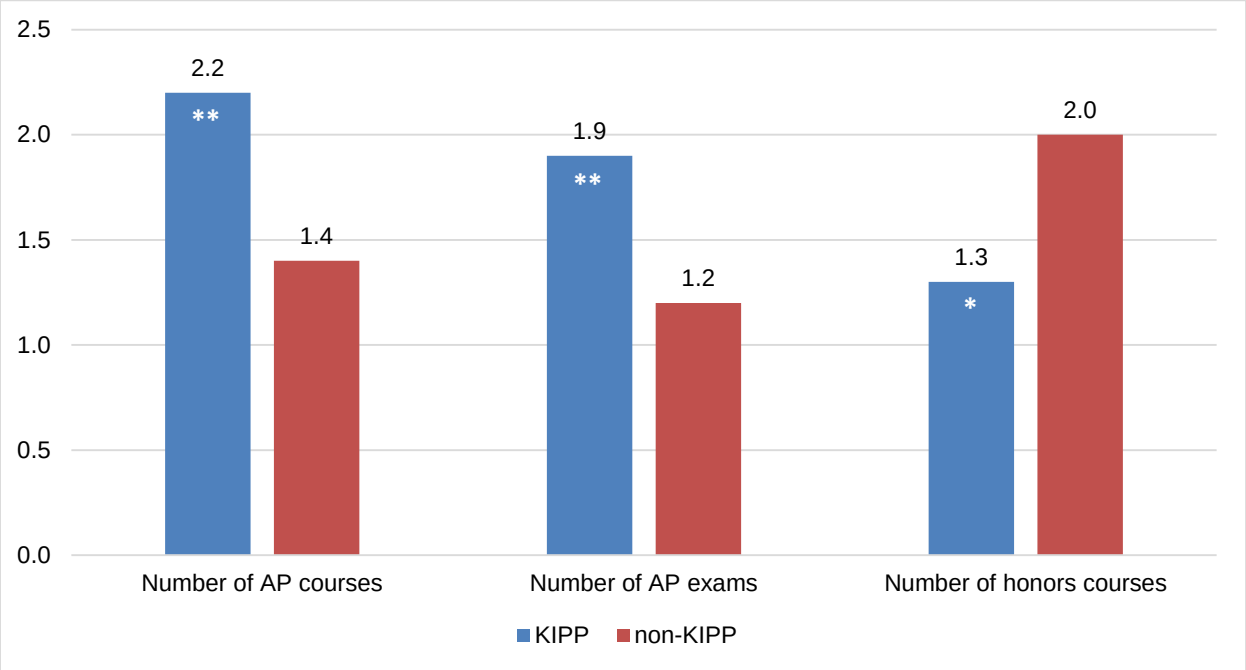
*Significantly different from zero at the .05 level, two-tailed test.

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

In addition to assistance provided by the school, we also measured specific activities students may have engaged in to better prepare themselves for college. KIPP high schools have a positive and significant effect on a summary index of seven different college preparation activities a student may have undertaken.[28] They also have a positive and significant effect on whether the student applied to at least one college or university by spring of senior year—93 percent of treatment students did so compared with 88 percent of comparison students.

A key aspect of students' college preparation involves which courses they take while in high school. Students with the opportunity to attend a KIPP high school enroll in more AP courses and, correspondingly, have taken or intend to take more AP exams (Figure IV.6). This difference is mostly offset by a negative effect of KIPP on the number of honors courses students have taken or plan to take, suggesting that students at KIPP high schools may be substituting AP courses for honors courses. KIPP high schools also have positive and statistically significant effects on the number of years of coursework students report having taken in several subjects, with treatment students averaging 1.4 more high school courses overall than comparison students, including more music, science, foreign language, and history.

Figure IV.6. KIPP high school impacts on advanced coursetaking



Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

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

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

[28] Impacts are positive and significant for three of the seven items included in the index: whether students visited in-state college campuses, whether they visited out-of-state college campuses, and whether they took practice ACT/SAT exams.

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

REFERENCES

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. "Inputs and Impacts in Charter Schools: KIPP Lynn." *American Economic Review: Papers & Proceedings*, vol. 100, May 2010, pp. 1-5.

Burke, Mary A. and Tim Sass. "Classroom Peer Effects and Student Achievement," *Journal of Labor Economics*, vol. 31, no. 1, 2013, pp. 51-82.

Duckworth, Angela L. and David Scott Yeager. "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes." *Educational Researcher*, vol. 44, no. 4, 2015.

Fortson, Kenneth, Philip Gleason, Emma Kopa, and Natalya Verbitsky-Savitz. "Horseshoes, Hand Grenades, and Treatment Effects? Reassessing Whether Nonexperimental Estimators are Biased." *Economics of Education Review*, vol. 44, February 2015, pp. 100-113.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts." Cambridge, MA: Mathematica Policy Research, January 2012.

Gill, Brian, Joshua Ferguson, Hanley Chiang, Bing-Ru Teh, Joshua Haimson, and Natalya Verbitsky-Savitz. "Replicating Experimental Impact Estimates in the Context of Control-Group Noncompliance." *Statistics and Public Policy.* Forthcoming.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. "The Evaluation of Charter School Impacts." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2010.

Gleason, Philip M., Christina Clark Tuttle, Brian Gill, Ira Nichols-Barrer, and Bing-ru Teh. "Do KIPP Schools Boost Student Achievement?" *Education Finance and Policy*, vol. 9, no. 1, 2014, pp. 36-58.

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

KIPP Foundation. "KIPP: 2014 Report Card." 2015. Available at http://www.kipp.org/reportcard. Accessed August 1, 2015.

Lake, Robin, Melissa Bowen, Allison Demeritt, Moira McCullough, Joshua Haimson, and Brian Gill. "Learning from Charter School Management Organizations: Strategies for Student Behavior and Teacher Coaching." Seattle, WA: Center on Reinventing Public Education and Washington DC: Mathematica Policy Research, March 2012.

Lavy, Victor and M. Daniele Paserman. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *The Economic Journal*, vol. 122, 2012, pp. 208–237

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

Nichols-Barrer, Ira, Philip Gleason, Brian Gill, and Christina Clark Tuttle. "Student Selection, Attrition, and Replacement in KIPP Middle Schools." *Educational Evaluation and Policy Analysis*, published online before print, January 2015.

Tuttle Christina Clark, Brian Gill, Phillip Gleason, Virginia Knechtel, Ira Nichols-Barrer, and Alex Resch. "KIPP Middle Schools: Impacts on Achievement and Other Outcomes." Washington, DC: Mathematica Policy Research, February 27, 2013

Tuttle, Christina Clark, Bing-ru Teh, Ira Nichols-Barrer, Brian P. Gill, and Philip Gleason. "Student Characteristics and Achievement in 22 KIPP Middle Schools." Washington, DC: Mathematica Policy Research, June 2010.

West, Martin R., Matthew A. Kraft, Amy S. Finn, et al. "Promise and Paradox: Measuring Students' Non-cognitive Skills and the Impact of Schooling." Center for Education Policy Research, Harvard University. February 2014.

Woodworth, Katrina R., Jane L. David, Roneeta Guha, Haiwen Wang, and Alejandra Lopez-Torkos. "San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement." Final Report. Menlo Park, CA: SRI International, 2008.

APPENDIX A

LIST OF KIPP SCHOOLS IN NETWORK

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

Table A.1. KIPP elementary schools (2014-2015)

School	Region	Year opened	Lottery-based Sample	Ineligible	Non-study
KIPP Academy Boston Elementary	KIPP Massachusetts	2014		X	
KIPP Academy Elementary School	KIPP NYC	2009	X		
KIPP AMP Elementary School	KIPP NYC	2013		X	
KIPP Ascend Primary School	KIPP Chicago	2010			X
KIPP Austin Comunidad	KIPP Austin	2010			X
KIPP Austin Connections Elementary	KIPP Austin	2011			X
KIPP Austin Leadership Elementary	KIPP Austin	2013		X	
KIPP Austin Obras	KIPP Austin	2013		X	
KIPP Believe Primary	KIPP New Orleans	2011			X
KIPP Central City Primary	KIPP New Orleans	2008			X
KIPP Columbus Elementary	KIPP Columbus	2014		X	
KIPP Comienza Community Prep	KIPP LA	2010			X
KIPP CONNECT Houston Primary School	KIPP Houston	2014		X	
KIPP DC: Arts & Technology Academy	KIPP DC	2014		X	
KIPP DC: Connect Academy	KIPP DC	2013		X	
KIPP DC: Discover Academy	KIPP DC	2009			X
KIPP DC: Grow Academy	KIPP DC	2010			X
KIPP DC: Heights Academy	KIPP DC	2011			X
KIPP DC: Lead Academy	KIPP DC	2012		X	
KIPP DC: LEAP Academy	KIPP DC	2007	X		
KIPP DC: Promise Academy	KIPP DC	2009			X
KIPP DC: Quest Academy	KIPP DC	2014		X	
KIPP DC: Spring Academy	KIPP DC	2013		X	
KIPP Delta Elem. Literacy Academy	KIPP Delta	2009			X
KIPP Destiny Elementary	KIPP Dallas-Fort Worth	2013		X	
KIPP Dream Prep	KIPP Houston	2006			X
KIPP East Community Primary	KIPP New Orleans	2014		X	
KIPP Empower Academy	KIPP LA	2010			X
KIPP Esperanza Dual Language Academy	KIPP San Antonio	2014		X	
KIPP Explore Academy	KIPP Houston	2009			X
KIPP GCP Primary	KIPP Eastern North Carolina	2012		X	
KIPP Harmony Academy	KIPP Baltimore	2009			X
KIPP Iluminar Academy	KIPP LA	2013		X	

Table A.1 *(continued)*

School	Region	Year opened	Lottery-based sample	Ineligible	Non-study
KIPP Indy Unite Elementary	KIPP Indianapolis	2014		X	
KIPP Infinity Elementary School	KIPP NYC	2010	X		
KIPP Legacy Preparatory School	KIPP Houston	2010			X
KIPP McDonogh 15 Primary	KIPP New Orleans	2006			X
KIPP Memphis Academy Elementary	KIPP Memphis	2013		X	
KIPP Memphis Collegiate Elementary	KIPP Memphis	2012		X	
KIPP New Orleans Leadership Primary	KIPP New Orleans	2011			X
KIPP PEACE Elementary School	KIPP Houston	2011			X
KIPP Philadelphia Elementary Academy	KIPP Philadelphia	2010	X		
KIPP Raíces Academy	KIPP LA	2008	X		
KIPP SHARP College Prep Lower School	KIPP Houston	2008	X		
KIPP SHINE Prep	KIPP Houston	2004	X		
KIPP STAR Harlem College Prep Elementary	KIPP NYC	2014		X	
KIPP STAR Washington Heights Elementary	KIPP NYC	2011			X
KIPP STRIVE Primary	KIPP Metro Atlanta	2012		X	
KIPP Un Mundo Dual Language Academy	KIPP San Antonio	2012		X	
KIPP Victory Academy	KIPP St. Louis	2014		X	
KIPP Vida Preparatory Academy	KIPP LA	2014		X	
KIPP Vision Primary	KIPP Metro Atlanta	2013		X	
KIPP VOICE Elementary	KIPP Jacksonville	2012		X	
KIPP WAYS Primary	KIPP Metro Atlanta	2014		X	
KIPP ZENITH Academy	KIPP Houston	2009			X
Life Academy at Bragaw, a KIPP School	KIPP New Jersey	2014		X	
Revolution Primary, a KIPP School	KIPP New Jersey	2014		X	
Seek Academy, a KIPP School	KIPP New Jersey	2013		X	
SPARK Academy, a KIPP School	KIPP New Jersey	2009	X		
THRIVE Academy, a KIPP School	KIPP New Jersey	2012		X	
	TOTAL		**8**	**31**	**21**

Note: Because the lotteries for inclusion in the elementary school study sample were conducted for entry in the 2011-2012 school year, schools were ineligible for the lottery-based sample if they opened in fall 2012 or later. Non-study schools are those elementary schools that were open as of 2011-2012, but were insufficiently oversubscribed to support the study design at an entry grade for that school year.

Table A.2. KIPP middle schools (2014-2015)

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP 3D Academy	KIPP Houston	2001		X			
KIPP Academy Boston Middle School	KIPP Massachusetts	2012			X		
KIPP Academy Lynn Middle School	KIPP Massachusetts	2004	X	X			
KIPP Academy (Houston)	KIPP Houston	1995		X			
KIPP Academy Middle School (New York)	KIPP NYC	1995	X	X			
KIPP Academy Nashville	KIPP Nashville	2005					X
KIPP Academy of Innovation	KIPP LA	2014				X	
KIPP Academy of Opportunity	KIPP LA	2003					X
KIPP Adelante Preparatory Academy	KIPP San Diego	2003					X
KIPP AMP Middle School	KIPP NYC	2005		X			
KIPP Ascend Middle School	KIPP Chicago	2003					X
KIPP Aspire Academy	KIPP San Antonio	2003		X			
KIPP Austin Academy of Arts & Letters	KIPP Austin	2009	X	X			
KIPP Austin Beacon Prep	KIPP Austin	2012			X		
KIPP Austin College Prep	KIPP Austin	2002	X	X			
KIPP Austin Vista Middle School	KIPP Austin	2012			X		
KIPP Bayview Academy	KIPP Bay Area	2003					X
KIPP Believe College Prep	KIPP New Orleans	2006					X
KIPP Bloom College Prep	KIPP Chicago	2013				X	
KIPP Blytheville College Prep School	KIPP Delta	2010		X			
KIPP Bridge Charter School	KIPP Bay Area	2002					X
KIPP Camino Academy	KIPP San Antonio	2010		X			
KIPP Central City Academy	KIPP New Orleans	2007					X
KIPP Charlotte	KIPP Charlotte	2007		X			
KIPP Columbus Middle	KIPP Columbus	2008					X

Table A.2 *(continued)*

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP CONNECT Houston Middle School	KIPP Houston	2014				X	
KIPP Courage College Prep	KIPP Houston	2012			X		
KIPP Create College Prep	KIPP Chicago	2012					X
KIPP DC: AIM Academy	KIPP DC	2005		X			
KIPP DC: KEY Academy	KIPP DC	2001		X			
KIPP DC: Northeast Academy	KIPP DC	2014				X	
KIPP DC: WILL Academy	KIPP DC	2006		X			
KIPP Delta College Preparatory School	KIPP Delta	2002		X			
KIPP Endeavor Academy	KIPP Kansas City	2007					X
KIPP Gaston College Preparatory	KIPP Eastern North Carolina	2001	X	X			
KIPP Halifax College Preparatory	KIPP Eastern North Carolina	2014				X	
KIPP Heartwood Academy	KIPP Bay Area	2004					X
KIPP Heritage Academy	KIPP Bay Area	2014				X	
KIPP Impact Middle School	KIPP Jacksonville	2010					X
KIPP Indy College Prep Middle	KIPP Indianapolis	2004					X
KIPP Infinity Middle School	KIPP NYC	2005	X	X			
KIPP Inspire Academy	KIPP St. Louis	2009					X
KIPP Intrepid Preparatory School	KIPP Houston	2008		X			
KIPP LA College Preparatory School	KIPP LA	2003	X				
KIPP Liberation College Prep	KIPP Houston	2006		X			
KIPP McDonogh 15 Middle	KIPP New Orleans	2006					X
KIPP Memphis Academy Middle	KIPP Memphis	2012			X		
KIPP Memphis Collegiate Middle	KIPP Memphis	2002		X			
KIPP Memphis Preparatory Middle	KIPP Memphis	2013				X	
KIPP Memphis University Middle	KIPP Memphis	2014				X	

Table A.2 *(continued)*

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP Montbello College Prep	KIPP Colorado	2011			X		
KIPP Nashville College Prep	KIPP Nashville	2013				X	
KIPP New Orleans Leadership Academy	KIPP New Orleans	2010					X
KIPP North Star Academy	KIPP Minnesota	2008					X
KIPP Philadelphia Charter School	KIPP Philadelphia	2003	X				
KIPP Philosophers Academy	KIPP LA	2012					X
KIPP Polaris Academy for Boys	KIPP Houston	2007		X			
KIPP Prize Preparatory Academy	KIPP Bay Area	2014				X	
KIPP Reach College Preparatory	— none —	2002					X
KIPP San Francisco Bay Academy	KIPP Bay Area	2003					X
KIPP Scholar Academy	KIPP LA	2012					X
KIPP Sharpstown College Prep	KIPP Houston	2007	X	X			
KIPP Sol Academy	KIPP LA	2013				X	
KIPP South Fulton Academy	KIPP Metro Atlanta	2003	X				
KIPP Spirit College Prep	KIPP Houston	2006		X			
KIPP STAR Harlem Middle School	KIPP NYC	2003	X	X			
KIPP STRIVE Academy	KIPP Metro Atlanta	2009		X			
KIPP Summit Academy	KIPP Bay Area	2003	X				
KIPP Sunshine Peak Academy	KIPP Colorado	2002		X			
KIPP TECH VALLEY	KIPP Albany	2005					X
KIPP TRUTH Academy	KIPP Dallas-Fort Worth	2003	X	X			
KIPP Tulsa College Preparatory	KIPP Tulsa	2005					X
KIPP Ujima Village Academy	KIPP Baltimore	2002	X				
KIPP Vision Academy	KIPP Metro Atlanta	2010		X			

Table A.2 *(continued)*

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP Voyage Academy for Girls	KIPP Houston	2009		X			
KIPP Washington Heights Middle School	KIPP NYC	2012			X		
KIPP WAYS Academy	KIPP Metro Atlanta	2003	X	X			
KIPP West Philadelphia Preparatory	KIPP Philadelphia	2009	X				
Rise Academy, a KIPP School	KIPP New Jersey	2006					X
TEAM Academy, a KIPP School	KIPP New Jersey	2002					X
	TOTAL		**16**	**30**	**7**	**11**	**26**

Note: Schools in the matched-student sample that were open in 2010-2011 are considered "established;' those that opened in fall 2011 or later (during the scale-up period) are considered "new." Because a minimum of two years of available data were required for inclusion in any of the middle school study samples, schools that opened in fall 2013 or later are ineligible. Non-study schools are those located in jurisdictions (states or districts) that did not provide data for the matched-student analysis.

Table A.3. KIPP High Schools (2014-2015)

School	Region	Year opened	Matched -student sample	Matched -school sample	In-eligible	Non-study
KIPP Academy Lynn Collegiate High School	KIPP Massachusetts	2011	X			
KIPP Atlanta Collegiate	KIPP Metro Atlanta	2011	X			
KIPP Austin Collegiate	KIPP Austin	2008	X	X		
KIPP Blytheville Collegiate High School	KIPP Delta	2014			X	
KIPP DC: College Preparatory	KIPP DC	2009	X	X		
KIPP Delta Collegiate High School	KIPP Delta	2006	X			
KIPP Denver Collegiate High School	KIPP Colorado	2009	X			
KIPP DuBois Collegiate Academy	KIPP Philadelphia	2010		X		
KIPP Generations Collegiate	KIPP Houston	2011	X			
KIPP Houston High School	KIPP Houston	2004	X			
KIPP King Collegiate High School	KIPP Bay Area	2007		X		
KIPP Memphis Collegiate High	KIPP Memphis	2011	X			
KIPP Nashville Collegiate High School	KIPP Nashville	2014			X	
KIPP Northeast College Preparatory	KIPP Houston	2013	X			
KIPP NYC College Prep High School	KIPP NYC	2009	X	X		
KIPP Pride High School	KIPP Eastern North Carolina	2005	X			
KIPP Renaissance High School	KIPP New Orleans	2010				X
KIPP San Francisco College Preparatory	KIPP Bay Area	2013				X
KIPP San Jose Collegiate	KIPP Bay Area	2008		X		
KIPP Sunnyside High School	KIPP Houston	2010	X			
KIPP University Prep High School	KIPP San Antonio	2009	X	X		
Newark Collegiate Academy, a KIPP School	KIPP New Jersey	2007		X		
	TOTAL		**14**	**8**	**2**	**2**

Note: Because a minimum of one year of available data was required for inclusion in any of the high school study samples, schools that opened in fall 2014 or later are ineligible. Non-study schools are those located in jurisdictions (states or districts) that did not provide data for the matched-student analysis.

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

APPENDIX B

DETAIL ON SURVEY OUTCOMES

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

This appendix presents detailed information about the study's analysis of the impact of KIPP schools on student behavior and attitudes. First, we present information on response rates and the baseline equivalence of the analytic samples for the survey-based outcomes. Next, we define our methodology for constructing survey indices. We then present complete survey impact estimates for elementary, middle, and high schools. We conclude the appendix with tables listing the components of each index.

Detail on sample

Our survey outcome impact estimates come from four different data sources: an elementary school parent survey that includes eight KIPP elementary schools; a middle school parent survey and a middle school student survey, both of which include 16 KIPP middle schools; and a high school student survey that includes eight KIPP high schools. The elementary and middle school surveys were conducted using a randomized controlled trial design (see Appendix D for more information on the elementary school analytic methods and Appendix E for the middle school analytic methods). Table B.1 summarizes response rates by treatment and control group and overall for the RCT designs.

Table B.1. Rates of available outcome data by analytic sample (RCT)

Analytic Sample	Treatment	Control	Overall	# of schools
Elementary School Parent Survey	74%	59%	65%	8
Middle School Student Survey	61%	46%	54%	16
Middle School Parent Survey	75%	61%	68%	16

Note: Sample sizes for individual outcomes vary. The elementary and middle school surveys were administered during the spring of 2013.

The high school student survey was conducted using a quasi-experimental design detailed in Appendix H. The analytic sample for survey outcomes (which combines two models) included 865 students overall, 504 in the treatment group and 361 in the control group. Two schools are included in both models. As a result of this analytical approach, some outcome sample sizes are greater than 865. The survey was administered during the spring of the students' fourth year after completing 8th grade at a KIPP middle school (12th grade for most students) in either 2012, 2013, or 2014.

To check that our four outcome samples are equivalent on observable characteristics, we examined baseline equivalence separately for each. In particular, we compared mean values and proportions of baseline characteristics in the treatment and comparison groups among those with valid outcome data, separately for each survey instrument. For the elementary school parent survey, we examined 21 baseline characteristics including gender, age, race, language spoken, household income, mother's education, and Internet and computer access, and the number of books in the household (Table B.2). We found only one statistically significant difference, which is what we would have expected due to chance alone. For the middle school parent and student surveys, we examined 28 baseline characteristics including baseline and pre-baseline test scores in reading and math, gender, age, race, language spoken, special education status, household income, free and reduced price lunch status, mother's education, computer and Internet access,

parental help with homework, and parental discussions on college (Tables B.3—B.4). There were only two statistically significant differences for the middle school student survey and one statistically significant difference for the parent survey. Both samples were balanced on key test score and demographic variables.

Given these findings, we are confident that the admissions lotteries were conducted correctly and that the treatment and control groups in each analysis sample are similar in terms of their background characteristics, motivation, and prior educational experiences, aside from the outcome of the lottery itself. We address any potential differences by controlling for baseline characteristics in the impact models.

Table B.2. Elementary school parent survey baseline equivalence

Baseline characteristic	Mean lottery winner	Mean lottery loser	Difference	p-value	N_t	N_c	SD_t	SD_c
Female	0.451	0.489	-0.039	0.373	319	329	0.498	0.501
Age of respondent in years	4.12	4.14	-0.012	0.708	304	310	1.148	0.959
White, non-Hispanic	0.010	0.008	0.002	0.777	344	361	0.093	0.091
Hispanic (any race)	0.395	0.407	-0.013	0.649	344	361	0.501	0.492
Black, non-Hispanic	0.550	0.540	0.010	0.724	344	361	0.500	0.499
Other race	0.045	0.044	0.000	0.990	344	361	0.198	0.206
English: main language at home	0.612	0.589	0.023	0.460	306	314	0.500	0.493
Another language is main language at home	0.223	0.197	0.025	0.393	306	314	0.447	0.399
English and another language spoken equally at home	0.166	0.213	-0.048	0.129	306	314	0.395	0.410
One adult in household	0.255	0.307	-0.052	0.149	347	362	0.425	0.462
Family income less than 15K	0.221	0.272	-0.051	0.160	291	305	0.415	0.446
Family income between 15K and 25K	0.236	0.203	0.033	0.359	291	305	0.436	0.403
Family income between 25K and 35K	0.207	0.200	0.007	0.877	291	305	0.422	0.401
Family income between 35K and 50K	0.181	0.184	-0.003	0.931	291	305	0.378	0.388
Family income 50K or greater	0.155	0.141	0.014	0.613	291	305	0.330	0.349
Mother's education: less than HS	0.083	0.171	-0.088**	0.002	303	310	0.316	0.377
Mother's education: HS or GED	0.255	0.242	0.013	0.727	303	310	0.432	0.429
Mother's education: some college	0.327	0.277	0.050	0.193	303	310	0.461	0.448
Mother's education: college	0.334	0.310	0.025	0.569	303	310	0.473	0.463
Student has access to computer with internet at home	0.774	0.792	-0.017	0.610	303	312	0.309	0.331
Number of children's books at home	37.192	40.418	-3.226	0.353	292	292	0.484	0.490

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

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

Table B.3. Middle school student survey baseline equivalence

Baseline characteristic	Mean lottery winner	Mean lottery loser	Difference	p-value	N_t	N_c	SD_t	SD_c
Baseline reading exam (z-score)	-0.150	-0.165	0.015	0.879	169	126	0.888	0.811
Baseline math exam (z-score)	0.009	-0.139	0.147	0.163	169	126	0.931	0.845
Pre-baseline reading exam (z-score)	-0.156	-0.255	0.100	0.304	158	118	0.755	0.815
Pre-baseline math exam (z-score)	-0.130	-0.129	-0.001	0.992	159	118	0.902	0.882
Female	0.494	0.511	-0.017	0.740	270	188	0.501	0.501
Age of respondent in years	10.49	10.546	-0.061	0.271	244	172	0.779	0.668
White, non-Hispanic	0.012	0.020	-0.008	0.620	281	198	0.156	0.141
Hispanic or Latino	0.549	0.581	-0.032	0.340	281	198	0.501	0.495
Black, non-Hispanic	0.368	0.374	-0.006	0.848	281	198	0.486	0.485
Other race	0.071	0.025	0.045*	0.024	281	198	0.295	0.157
English: main language at home	0.476	0.460	0.017	0.633	281	198	0.501	0.500
Another language is main language at home	0.259	0.318	-0.059	0.139	281	198	0.420	0.467
English and another lang. spoken equally at home	0.265	0.222	0.042	0.271	281	198	0.441	0.417
One adult in household	0.292	0.253	0.039	0.355	279	198	0.453	0.436
Free/reduced price lunch status	0.818	0.864	-0.046	0.176	266	184	0.416	0.344
Special education status	0.059	0.110	-0.051	0.103	259	181	0.279	0.314
Family income less than 15K	0.213	0.201	0.012	0.773	240	164	0.416	0.402
Family income between 15K and 25K	0.277	0.250	0.027	0.543	240	164	0.418	0.434
Family income between 25K and 35K	0.157	0.250	-0.093*	0.034	240	164	0.377	0.434
Family income between 35K and 50K	0.185	0.165	0.020	0.643	240	164	0.391	0.372
Family income 50K or greater	0.168	0.134	0.034	0.358	240	164	0.398	0.342
Mother's education: less than HS	0.194	0.242	-0.049	0.215	280	198	0.378	0.430
Mother's education: HS or GED	0.249	0.263	-0.014	0.735	280	198	0.434	0.441
Mother's education: some college	0.248	0.232	0.016	0.703	280	198	0.434	0.423
Mother's education: college	0.309	0.263	0.047	0.257	280	198	0.471	0.441
Student has access to computer with internet at home	0.808	0.801	0.007	0.862	250	171	0.371	0.400
Parent helps student with homework at least 5 days per week	0.663	0.726	-0.064	0.177	248	168	0.455	0.447
Parent discussed college with student at least twice during pre-baseline school year	0.860	0.851	0.009	0.811	248	168	0.361	0.357

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

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

Table B.4. Middle school parent survey baseline equivalence

Baseline characteristic	Mean lottery winner	Mean lottery loser	Difference	p-value	N_t	N_c	SD_t	SD_c
Baseline reading exam (z-score)	-0.167	-0.182	0.015	0.864	201	173	0.929	0.821
Baseline math exam (z-score)	-0.050	-0.180	0.131	0.154	201	173	0.927	0.857
Pre-baseline reading exam (z-score)	-0.161	-0.260	0.099	0.257	192	163	0.780	0.798
Pre-baseline math exam (z-score)	-0.134	-0.172	0.038	0.695	193	164	0.899	0.876
Female	0.507	0.496	0.011	0.797	330	248	0.501	0.501
Age of respondent in years	10.53	10.565	-0.035	0.491	291	215	0.805	0.665
White, non-Hispanic	0.019	0.031	-0.012	0.434	343	262	0.160	0.172
Hispanic or Latino	0.520	0.553	-0.034	0.285	343	262	0.500	0.498
Black, non-Hispanic	0.387	0.366	0.021	0.450	343	262	0.492	0.483
Other race	0.074	0.050	0.025	0.222	343	262	0.299	0.218
English: main language at home	0.498	0.488	0.010	0.763	342	260	0.499	0.501
Another language is main language at home	0.249	0.296	-0.047	0.175	342	260	0.410	0.457
English and another lang. spoken equally at home	0.253	0.215	0.038	0.258	342	260	0.431	0.412
One adult in household	0.295	0.265	0.030	0.434	340	260	0.458	0.442
Free/reduced price lunch status	0.809	0.839	-0.030	0.342	322	242	0.425	0.368
Special education status	0.071	0.105	-0.034	0.220	311	239	0.277	0.307
Family income less than 15K	0.220	0.181	0.039	0.288	285	204	0.416	0.386
Family income between 15K and 25K	0.263	0.250	0.013	0.760	285	204	0.418	0.434
Family income between 25K and 35K	0.168	0.255	-0.087*	0.028	285	204	0.378	0.437
Family income between 35K and 50K	0.183	0.186	-0.003	0.936	285	204	0.393	0.390
Family income 50K or greater	0.166	0.127	0.038	0.237	285	204	0.395	0.334
Mother's education: less than HS	0.182	0.223	-0.041	0.231	341	260	0.368	0.417
Mother's education: HS or GED	0.244	0.258	-0.013	0.717	341	260	0.428	0.438
Mother's education: some college	0.272	0.265	0.007	0.857	341	260	0.453	0.442
Mother's education: college	0.301	0.254	0.047	0.192	341	260	0.464	0.436
Student has access to computer with internet at home	0.813	0.837	-0.025	0.493	301	215	0.370	0.370
Parent helps student with homework at least 5 days per week	0.709	0.698	0.011	0.800	299	212	0.449	0.460
Parent discussed college with student at least twice during pre-baseline school year	0.859	0.863	-0.004	0.905	299	212	0.365	0.344

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

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

For the high school student survey, there were no statistically significant differences between the treatment and comparison group on any baseline characteristics (Table B.5). Assessing baseline equivalence separately for the two different models, the treatment and comparison groups in the model using matched middle schools are equivalent, which is consistent with the model's design due to the matching process. In the model using adjacent cohorts from the same middle school, the groups are also generally equivalent—no differences between the groups are significant at the 5% level, and only one difference (whether a student is Hispanic) is marginally significant.

Table B.5. High school student survey baseline equivalence

Baseline characteristic	Treatment	Comp.	Difference	p-value	N_t	N_c	SD_t	SD_c
Matched middle school model								
Baseline reading exam (z-score)	0.100	0.101	-0.001	0.989	264	264	0.743	0.762
Baseline math exam (z-score)	0.547	0.593	-0.046	0.536	264	264	0.870	0.822
Male	0.409	0.405	0.004	0.930	264	264	0.493	0.492
Black, non-Hispanic	0.508	0.515	-0.008	0.862	264	264	0.501	0.501
Hispanic or Latino	0.356	0.390	-0.034	0.419	264	264	0.480	0.489
Old for grade	0.167	0.144	0.023	0.472	264	264	0.373	0.352
Adjacent cohort model								
Baseline reading exam (z-score)	0.444	0.367	0.077	0.347	206	209	0.825	0.834
Baseline math exam (z-score)	0.810	0.758	0.052	0.547	206	209	0.771	0.986
Male	0.353	0.393	-0.040	0.381	221	229	0.479	0.489
Black, non-Hispanic	0.520	0.555	-0.034	0.468	221	229	0.501	0.498
Hispanic or Latino	0.466	0.384	0.082	0.080	221	229	0.500	0.487
Old for grade	0.231	0.197	0.034	0.377	221	229	0.422	0.398

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

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

Defining indices

Several of the measures in Chapters II, III, and IV of this report are derived from survey items. Many of the outcomes are indices created by combining closely related survey items into a single measure, reducing measurement error, and capturing the breadth of a construct. The indices and their component items are listed at the end of this appendix in Tables B.9 through B.11. The tables also include survey outcomes or measures that are not indices, but are derived from one or more survey items.

The process for creating the indices included a number of steps to maximize reliability and reduce the number of separate outcome variables we examined (that is, to reduce dimensionality). We first identified all items from the surveys that were conceptually related to a specific construct. We used principal component analysis to confirm that the items were related to the underlying construct (and to one another) in the theoretically expected way, excluding

items not related to this underlying construct. We then created an index variable based on the included items to represent the underlying construct. Finally, we computed the standardized Cronbach's alpha, an estimate of the internal consistency or reliability of an index, and dropped indices with alpha values suggesting low reliability. Conventionally, indices with alpha values greater than 0.7 are considered reliable. Following Gleason et al. (2010), we retained indices with alpha values somewhat lower than this threshold but indicate that these indices may have low levels of reliability. Indices with values of alpha below 0.7 are noted in the tables.

Detailed survey outcomes tables

The following tables provide detailed results from our analysis of the impact of KIPP on survey-based (non-achievement) outcomes for elementary, middle, and high school students. Further details on the analytic methods used can be found in Appendix D for the elementary school analysis, Appendix E for the middle school analysis, and Appendix H for the high school analysis.

Table B.6. Estimated impact of offer of admission to KIPP elementary school

Outcome	Mean lottery winner	Mean lottery loser	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Student motivation and engagement								
Minutes spent on homework on typical night, parent report	55.11	51.86	3.26	0.25	341	327	38.31	38.01
Parent says student typically completes homework	0.95	0.92	0.03	0.05	340	327	0.18	0.28
Frequency of negative comments to parent about school (index)	1.26	1.26	0.00	0.99	342	346	0.42	0.42
Frequency of positive comments to parent about school (index)	2.79	2.69	0.10**	<0.01	341	347	0.42	0.48
Index of student development compared to peers	3.25	3.31	-0.06	0.09	341	346	0.46	0.44
Student behavior								
Index indicating student is well adjusted to school	3.66	3.60	0.06	0.08	338	343	0.42	0.42
School experiences and satisfaction								
Index of parental satisfaction with school	3.63	3.36	0.27**	<0.01	341	341	0.49	0.65
Parent rates school as excellent	0.77	0.51	0.26**	<0.01	344	351	0.42	0.50
Index of parental perceptions of problems in student's school	1.92	1.84	0.07	0.44	289	270	1.15	1.07
Index of parent outreach to school[a]	0.70	0.70	0.01	0.77	337	343	0.27	0.26
Index of school efforts to engage parents in school	2.53	2.39	0.14**	<0.01	327	326	0.44	0.53
Index of family engagement at home	2.83	2.81	0.03	0.59	340	348	0.62	0.62
Index of student engagement at home[a]	3.30	3.19	0.11	0.07	343	354	0.68	0.74
Index of parent indicating school is too easy[a]	0.10	0.12	-0.02	0.28	339	316	0.23	0.25
Index of parent indicating school is too hard[a]	0.02	0.04	-0.02	0.12	339	316	0.13	0.14
Education goals and aspirations								
Parent expects student to graduate high school on time	0.96	0.98	-0.02	0.17	338	345	0.21	0.14
Parent wishes student to complete college	0.99	0.98	0.00	0.66	341	351	0.11	0.13
Parent believes student very likely to complete college	0.81	0.71	0.10**	<0.01	340	349	0.40	0.45

Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

[a] Index has an alpha smaller than 0.7, indicating low reliability

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

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

Table B.7. Estimated impact of offer of admission to a KIPP middle school

Outcome	Mean lottery winner	Mean lottery loser	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Student motivation and engagement								
Minutes spent on homework on typical night, student report	78.64	71.42	7.22	0.18	272	176	53.38	50.65
Minutes spent on homework on typical night, parent report	99.79	84.97	14.82**	<0.01	326	244	57.41	52.56
Parent says student typically completes homework	0.91	0.92	-0.01	0.77	327	243	0.23	0.27
Index of positive school engagement	3.37	3.38	-0.01	0.82	279	197	0.46	0.45
Index of negative school engagement[a]	3.79	3.84	-0.05	0.24	280	198	0.47	0.43
Index of self-control	3.53	3.55	-0.02	0.66	280	194	0.50	0.49
Index of academic collaboration	3.30	3.18	0.12*	0.02	279	198	0.49	0.53
Index of effort in school	3.60	3.61	-0.01	0.89	281	198	0.43	0.44
Index of academic confidence	3.30	3.28	0.01	0.84	280	197	0.54	0.54
Index of grit	3.58	3.54	0.05	0.35	272	190	0.51	0.49
Student behavior								
Index of peer pressure for bad behavior	0.03	0.03	0.00	0.94	281	197	0.10	0.13
Index of student reported undesirable behavior	2.10	2.06	0.04	0.71	280	193	1.03	1.08
Index of parent report of undesirable behavior[a]	1.20	1.24	-0.04	0.16	337	255	0.32	0.35
Index of student reported illegal action[a]	0.03	0.02	0.00	0.59	281	196	0.09	0.08
Index of good behavior, student report[a]	4.42	4.44	-0.03	0.80	280	196	0.98	1.07
Index indicating well-adjusted student	3.44	3.49	-0.05	0.23	334	252	0.44	0.43
Parent reported any school disciplinary problems for student	0.35	0.32	0.02	0.53	338	253	0.47	0.47
Index of parent-reported frequency of school disciplinary actions for student[a]	0.22	0.18	0.04	0.17	338	253	0.38	0.31
Student never gets in trouble at school	0.44	0.51	-0.07	0.20	281	196	0.50	0.50
Index of good behavior, parent report[a]	2.26	2.28	-0.03	0.43	322	248	0.39	0.38
Index of parental concerns about student	1.41	1.41	-0.01	0.92	335	250	0.65	0.66
School experiences and satisfaction								
Index of student's feelings about school	3.35	3.27	0.08	0.06	274	196	0.41	0.45
Index of parental satisfaction with school	3.38	3.03	0.35**	<0.01	336	253	0.65	0.67
Parent rates school as excellent	0.56	0.28	0.29**	<0.01	339	256	0.50	0.45
Student likes school a lot	0.54	0.54	0.00	0.99	280	197	0.49	0.50
Index of parental perceptions of problems in student's school	2.00	2.24	-0.23*	0.04	277	184	1.12	1.14
Index of parent outreach to school[a]	0.64	0.64	0.01	0.81	324	251	0.24	0.27
Index of school efforts to engage parents in school	2.34	2.14	0.21**	<0.01	313	241	0.53	0.55
Parent talks to child about school experiences almost every day	0.80	0.79	0.01	0.82	340	258	0.39	0.41

Table B.7 *(continued)*

Outcome	Mean lottery winner	Mean lottery loser	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Parent helps child with homework almost every day	0.41	0.45	-0.04	0.32	339	257	0.49	0.50
Index of student perceptions of schoolmates	2.88	2.72	0.16**	<0.01	275	194	0.45	0.51
Index of student perceptions of teachers	3.48	3.41	0.07	0.08	281	195	0.39	0.42
Index of student perceptions of school disciplinary environment	3.37	3.24	0.13**	<0.01	281	196	0.44	0.41
Academic difficulty, parent report								
Index indicating school is too easy[a]	0.11	0.12	-0.01	0.73	323	238	0.23	0.24
Index indicating school is too difficult[a]	0.09	0.08	0.01	0.72	323	238	0.19	0.19
Education goals and aspirations								
Student expects to graduate high school on time	0.97	0.98	-0.01	0.41	281	198	0.14	0.14
Parent expects student to graduate high school on time	0.93	0.95	-0.03	0.12	337	254	0.21	0.21
Parent wishes student to complete college	0.98	0.99	-0.01	0.52	339	256	0.14	0.11
Parent believes student very likely to complete college	0.69	0.69	-0.01	0.90	331	253	0.45	0.46
Student reports having more than 2 discussions about college at school	0.46	0.47	-0.01	0.82	281	196	0.50	0.50
Student reports having more than 2 discussions about college at home	0.70	0.65	0.05	0.28	281	196	0.46	0.48
Parent report of having more than 2 discussions about college	0.84	0.82	0.02	0.60	338	256	0.37	0.38

Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information. Some of the items presented in this table will be presented in an Appendix of the final report.

[a] Index has an alpha smaller than 0.7, indicating low reliability

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

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

Table B.8. Estimated impact of opportunity to attend KIPP high school

Outcome	Mean (T)	Mean (C)	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Student motivation and engagement								
Minutes spent on homework on a typical night	145.14	146.80	-1.66	0.85	410	398	103.37	94.43
Index of positive school engagement	3.19	3.18	0.01	0.89	491	492	0.61	0.59
Index of negative school engagement	1.75	1.67	0.07	0.32	493	500	0.89	0.81
Index of self control	4.43	4.38	0.05	0.45	401	394	0.71	0.68
Index of academic collaboration	3.05	3.07	-0.02	0.70	492	493	0.66	0.61
Index of academic motivation	3.21	3.20	0.01	0.84	486	491	0.54	0.55
Index of academic confidence	3.31	3.36	-0.05	0.26	493	501	0.56	0.55
Index of effort in school	3.28	3.22	0.06	0.32	492	496	0.65	0.68
Student dropped out of school	0.01	0.03	-0.02*	0.04	504	513	0.06	0.17
Education goals and aspirations								
Student thinks they will graduate college or go further in his/her education	0.84	0.85	-0.01	0.71	496	503	0.37	0.36
College preparation activities								
Index of student reports of discussions about college at school	2.59	2.40	0.19**	<0.01	471	481	0.52	0.58
Student discussed how to pay for college in depth at school	0.72	0.60	0.11**	<0.01	489	491	0.44	0.49
Index of student reports of teachers/counselor assistance with planning for college	3.37	3.15	0.22**	<0.01	477	481	0.77	0.87
Index of student reports of parent assistance with planning for college	3.04	2.94	0.10	0.22	475	473	0.93	1.03
Index of teacher helpfulness and encouragement	3.18	3.02	0.15*	0.02	477	478	0.71	0.77
Index of counselor helpfulness and encouragement	3.24	3.08	0.16*	0.03	466	471	0.79	0.85
Student reports having completed an education plan	0.86	0.86	0.00	1.00	459	466	0.34	0.33
Student reports having completed a career plan	0.75	0.71	0.04	0.36	459	466	0.44	0.45
College testing								
Student has taken PSAT or PLAN	0.82	0.80	0.02	0.50	485	500	0.37	0.40
Student has taken or plans to take SAT or ACT	0.89	0.85	0.03	0.26	487	500	0.31	0.36
Student has taken SAT or ACT	0.83	0.78	0.05	0.14	487	500	0.37	0.42
Student plans to take SAT or ACT	0.07	0.11	-0.04	0.13	487	500	0.25	0.31
Student reports participating in college prep program	0.25	0.21	0.04	0.21	484	477	0.44	0.40
Index of college preparation activities	0.88	0.80	0.08**	<0.01	460	459	0.19	0.23
Number of colleges or universities to which the student applied	6.99	6.90	0.09	0.84	490	496	4.95	5.83
College Application Process								
Student applied to at least 1 college/university	0.93	0.88	0.05*	0.04	490	496	0.24	0.32

Table B.8 *(continued)*

Outcome	Mean (T)	Mean (C)	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Accepted to at least 1 of the respondent's top 3 choices for college/university	0.81	0.76	0.05	0.12	483	487	0.38	0.43
Applied early admission/action/decision to one of top 3 college choices	0.41	0.42	-0.01	0.80	484	483	0.49	0.50
Coursetaking								
School offers honors courses	0.80	0.84	-0.04	0.24	426	426	0.40	0.37
School offers AP or IB courses	0.98	0.88	0.09**	<0.01	461	434	0.15	0.32
Number of honors courses student took or planned to take	1.25	1.95	-0.70*	0.01	411	440	1.97	2.99
Number of AP courses student took or planned to take	2.23	1.33	0.90**	<0.01	498	453	2.10	1.92
Number of AP exams student took or planned to take	1.96	1.14	0.82**	<0.01	480	435	2.11	1.77
How many years of the following subjects will the student complete:								
English or language arts	3.97	3.91	0.06	0.12	469	463	0.46	0.48
Math	3.93	3.87	0.06	0.20	470	466	0.51	0.59
Science or engineering	3.78	3.54	0.23**	<0.01	466	457	0.78	0.89
Social studies or history	3.79	3.42	0.37**	<0.01	469	460	0.59	0.83
Foreign language	2.92	2.73	0.19*	0.02	467	451	0.90	0.97
Physical education	2.36	2.30	0.06	0.56	467	451	1.33	1.29
Art	1.50	1.53	-0.02	0.82	462	460	1.09	1.22
Music	1.33	0.90	0.44**	<0.01	451	435	1.29	1.28
Technical or vocational education	0.53	0.61	-0.08	0.44	413	416	1.02	1.14
Total number of high school courses student took or plans to take	24.31	22.91	1.40**	<0.01	470	466	4.67	5.31
School experiences and satisfaction								
Index of student's feelings about school	3.39	3.40	-0.01	0.77	490	494	0.51	0.52
Student likes school a lot	0.37	0.45	-0.08	0.06	500	507	0.49	0.50
Index of student perceptions of schoolmates	3.01	3.04	-0.03	0.55	493	502	0.55	0.59
Index of student perceptions of teachers	3.33	3.34	-0.01	0.89	483	501	0.56	0.57
Index of school disciplinary environment	3.40	3.38	0.02	0.62	490	501	0.51	0.50
Student was suspended or expelled within the last school year	0.12	0.11	0.00	0.91	491	499	0.32	0.32
Student feels is getting a good education at the high school (strongly agree)	0.54	0.54	0.00	0.98	497	495	0.49	0.50
Index of helpfulness of high school in career planning	2.64	2.50	0.14	0.08	487	485	0.94	0.93

Note: Estimates are from a model that pools impacts across schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information. All indices have alpha values larger than 0.7. Standard errors are adjusted for the fact that some students are included in the sample multiple times.

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

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

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

DESCRIPTION OF INDICES

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

Table B.9. Description of indices: elementary school parent survey

Variable	Principal survey items included	Scale/Definition
Frequency of negative comments to parent about school (index)	F1. On average, during the first two months of this school year how often... **a. did the student complain about school?** **b. was the student upset or reluctant to go to school?** **c. did the student pretend to be sick or stay home from school?**	Mean across items (a), (b), and (c) using the following scale: Not at all (1) Once a week or less (2) More than once a week (3)
Frequency of positive comments to parent about school (index)	F1. On average, during the first two months of this school year how often... **d. did the student say good things about school?** **e. did the student say he/she like his/her teacher?** **f. did the student look forward to going to school?**	Mean across items (d), (e), and (f) using the following scale: Not at all (1) Once a week or less (2) More than once a week (3)
Index of student development compared to peers	F2. How does the student compare to other children of the same age? **a. Student is independent and takes care of himself/herself** **b. Student pays attention** **c. Student learns, thinks, and solves problems**	Mean across items (a), (b), and (c) using the following scale: Much less well than other children (1) Slightly less well than other children (2) As well as other children (3) Slightly better than other children (4) Better than other children his/her age (5)
Index indicating a student is well-adjusted to school	F3. For each of the following statements, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree." **a. The student gets along with others** **b. The student likes schools** **c. The student works hard at school** **d. The student is self-confident** **e. The student is creative** **f. The student is happy** **g. The student respects adults**	Mean across items (a) through (g) using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of parental satisfaction with school	D6. Please rate each of the following features of the school the student attends/attended for the 2012-2013 school year, as "excellent," "good," "fair," or "poor." **a. Facilities, like the classrooms, library, cafeteria, or the gym** **b. Academics, the teachers, and classes** **c. Safety** **d. Discipline**	Mean across items (a) through (d) using the following scale: Poor (1) Fair (2) Good (3) Excellent (4)

Table B.9 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Index of parental perceptions of problems in student's school	D1. For each of the following issues, please tell me if you feel it [is/was] "not a problem," "a small problem," "a medium problem," or "a big problem" at this school. **a. Student destroying property** **b. Student being late for school** **c. Student missing school** **d. Fighting** **e. Bullying** **f. Stealing**	Mean across items (a) through (f) using the following scale: Not a problem (1) A small problem (2) A medium problem(3) A big problem (4)
Index of parent outreach to school	C3. Since the beginning of this school year, have you or any other adults in your household… **a. Attended an open house or a back-to-school night?** **b. Attended a meeting of a PTA, PTO, or Parent-Teacher Student Organization?** c. Gone to a parent-teacher conference or meeting with his/her teacher? **d. Attended a school or class event, such as a play, sports event, or science fair?** **e. Acted as a volunteer at the school or served on a committee?** **f. Participated in fundraising for the student's school?** g. Visited the school's website for information?	Mean across items (a), (b), (d), (e), and (f) using the following scale: No (0) Yes (1)
Index of school efforts to engage parents in school	C4. How well would you say the student's school… **a. Lets you know between report cards how the student is doing in school?** **b. Helps you understand what children at the student's age are like?** **c. Makes you aware of chances to volunteer at the school?** **d. Provides workshops, materials, or advice about how to help he student learn at home?** **e. Provides information on community services to help the student or your family?** **f. Communicates with you through email or online forums?** **g. Provides online resources?**	Mean across items (a) through (g) using the following scale: Not at all (1) Just O.K (2) Very well (3)

Table B.9 *(continued)*

Variable	Principal survey items included	Scale/Definition
Index of family engagement at home	C8. In a typical week, how often do you or any other family or household member do the following things with him/her? a. Read books to him/her? b. Tell stories to him/her? **c. Sing songs with him/her?** **d. Help him/her to do arts and crafts?** **e. Involve him/her in household chores, like cooking, cleaning, setting the table, or caring for pets?** **f. Play games or do puzzles with him/her?** **g. Talk about nature or do science projects with him/her?** **h. Build something or play with construction toys with him/her?** **i. Play a sport or exercise together?**	Mean across items (c) though (i) using the following scale: Not at all (1) Once or twice (2) 3 to 6 times (3) Every day (4)
Index of student engagement at home	C10. How often did he/she look at picture books outside of school in the past week? C11. In the past week, how often did he/she read to or pretend to read to himself/herself or to others outside of school?	Mean across C10 and C11 using the following scales: Never (1) Once or twice (2) 3 to 6 times (3) Every day (4)
Index of parent indicating school is too easy	B4. Do you think the homework (is/was) too difficult, about right, or too easy for the student? B5a. Do you think the material covered in (his/her) math class (is/was) too difficult, about right, or too easy for the student? B6a. Do you think the material covered in (his/her) English/language arts class (is/was) too difficult, about right, or too easy for the student?	Mean across all items using the following scale: Too easy (1) About right/Too difficult (0)
Index of parent indicating school is too difficult	See items B4, B5a, and B6a	Mean across all items using the following scale: Too difficult (1) Other responses (0)

Table B.10. Description of indices: middle school parent and student surveys

Variable	Survey questions	Scale/definition and included items
Index of positive school engagement	Student survey C3. For the following statements about school, do you do each one "almost always," "often," "sometimes," or "almost never?" **a. stick with a class assignment or task until it is done** **b. put in your best effort on class assignments, projects, and homework** **c. ask a teacher for help when you don't understand an assignment** **d. ask another student for help when you don't understand an assignment** **e. take part in class discussions or activities** **f. feel challenged in class** **g. receive recognition or praise for doing good school work** **h. learn from your mistakes at school** **j. complete class assignments, projects, and homework on time**	Mean across items (a) though (i) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost Always (4)
Index of negative school engagement	Student survey C3. For the following statements about school, do you do each one "almost always," "often," "sometimes," or "almost never?" **i. think of dropping out of school** **j. try to stay home from school**	Mean across items (j) and (k) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost Always (4)
Index of self-control	Student survey C4. Thinking about a typical week during the 2012-2013 school year, how often did you do each of the following things? **a. Went to all of your classes prepared** **b. Remained calm even when things happened that could upset you** **c. Paid attention in all of your classes** **d. Listened to other students speak without interrupting them** **e. Were polite to adults and other students** **f. Remembered and followed directions** **g. Controlled your temper** i. Got to work right away rather than procrastinating	Mean across items (a) though (g) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost Always (4)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of academic collaboration	Student survey B2. How much you agree or disagree with the following statements? **a. You like to work with other students** b. You learn things quickly in most school subjects c. Because reading is fun, you wouldn't want to give it up d. You are good at most school subjects **e. You learn most when you work with other students** f. English/Language Arts is one of your best subjects **g. You do your best work when you work with other students** h. Math is one of your best subjects **i. You like to help other people do well in group assignments** j. You do well in tests in most school subjects **k. It is helpful to put together everyone's ideas when you work on a project**	Mean across items (a), (e), (g), and (i) using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of effort in school	Student survey B1. I'm going to read you some statements about your schoolwork. For each please tell me if these things apply to you "almost always," "often," "sometimes," or "almost never"? a. You're sure you can understand even the most difficult material presented in textbooks or other written material b. You can learn something really difficult when you want to **c. In school you work as hard as possible** d. You're certain you can understand even the most difficult material presented by the teacher e. If you decide to not get any bad grades, you can really do it **f. In school, you keep working even if the material is difficult** g. You're certain you can do an excellent job on assignments and tests **h. You try to do your best to learn the knowledge and skills taught** **i. You work hard in school so you can get a good job** **j. If you want to learn something well, you can** k. You're certain you can master the material you are taught **l. If you don't understand something in your schoolwork, you try to find additional information to help you learn** **m. You put forth your best effort in school**	Mean across items (c), (f), (h), (i), (j), (l), and (m) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of academic confidence	Student survey B1. I'm going to read you some statements about your schoolwork. For each please tell me if these things apply to you "almost always," "often," "sometimes," or "almost never"? **a. You're sure you can understand even the most difficult material presented in textbooks or other written material** **b. You can learn something really difficult when you want to** c. In school you work as hard as possible **d. You're certain you can understand even the most difficult material presented by the teacher** **e. If you decide to not get any bad grades, you can really do it** f. In school, you keep working even if the material is difficult **g. You're certain you can do an excellent job on assignments and tests** h. You try to do your best to learn the knowledge and skills taught i. You work hard in school so you can get a good job j. If you want to learn something well, you can **k. You're certain you can master the material you are taught** l. If you don't understand something in your schoolwork, you try to find additional information to help you learn **m. You put forth your best effort in school**	Mean across items (a), (b), (d), (e), (g) and (k) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of Grit	Student survey C5. For each item, tell us whether it is very much like you, mostly like you, somewhat like you, not much like you or not at all like you. **a. New ideas and projects sometimes distract you from previous ones.** **b. Delays and obstacles don't discourage you** **c. You have been obsessed with a certain idea or project for a short time but later lost interest.** **d. You are a hard worker** **e. You often set a goal but later choose to follow a different one** **f. You have difficulty keeping your focus on projects that take more than a few months to complete** **g. You finish whatever you begin** **h. You are hard working and careful**	Mean across all items using the following two scales: For items (a), (c), (e), and (f): Very much like you (1) Mostly like you (2) Somewhat like you (3) Not much like you (4) Not like you at all (5) For items: (b), (d), (g), and (h): Very much like you (5) Mostly like you (4) Somewhat like you (3) Not much like you (2) Not like you at all (1)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of Peer Pressure for Bad Behaviors	Student survey E2a. During the 2012-2013 school year, did your friends pressure you to do any of the following things? **a. Skip class or school?** **b. Drink alcohol?** **c. Smoke cigarettes?** **d. Use marijuana or other drugs?** e. Commit a crime or do something violent?	Mean across items (a) through (d) using the following scale: Yes (1) No (0)
Index of student reported undesirable behavior	Student survey E1. During the 2012-2013 school year, how often have you: **a. Argued with your parents or guardians?** b. Smoked cigarettes? **c. Lied to your parents or guardians?** d. Stolen something from someone else? e. Taken something from a store without paying for it? **f. Given a teacher a hard time?** g. Drunk alcohol? h. Skipped, or cut, classes during school? i. Skipped, or cut, the entire school day? j. Used marijuana or other drugs? **k. Gotten in trouble at school?** **l. Lost your temper at home or at school?** m. Gotten arrested or held by police?	Mean across items (a), (c), (f), (k), and (l) using the following scale: Just once or twice during the past year (1) Every couple of months (2) About once a month (3) About once a week (4) Almost every day (5)
Index of student reported illegal action	Student survey E1. During the 2012-2013 school year, have you ever done any of the following things? a. Argued with your parents or guardians? b. Smoked cigarettes? c. Lied to your parents or guardians? **d. Stolen something from someone else?** **e. Taken something from a store without paying for it?** f. Given a teacher a hard time? **g. Drunk alcohol?** h. Skipped, or cut, classes during school? i. Skipped, or cut, the entire school day? j. Used marijuana or other drugs? k. Gotten in trouble at school? l. Lost your temper at home or at school? m. Gotten arrested or held by police?	Mean across items (d), (e), (g), and (m) using the following scale: Yes (1) No (0)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of parent report of undesirable behavior	Parent survey F6. During the 2012-2013 school year, how often did [STUDENT] do the following things? **a. Break something on purpose** **b. Punch or hit someone in anger** **c. Argue with you**	Mean across all items using the following scale: Never (1) Sometimes (2) Often (3)
Parent reported any school disciplinary problems for student	Parent survey F5. During the 2012-2013 school year, how many times was [STUDENT]… a. Sent out of class for disciplinary reasons? b. Suspended from school? c. Expelled from school?	Respondent receives a value of 1 if they provided a nonzero answer to (a), (b), or (c), and 0 otherwise.
Index of Parent-Reported Frequency of School Disciplinary Actions for Student	See item F5 above.	Mean across all items using the following scale: Not at all (0) 1 to 3 times (1) 4 to 6 times (2) 7 to 10 times (3) More than 10 times (4)
Index of good behavior, student report	Student survey E3b.During the 2012-2013 school year, how often have you: **a. Helped another student with school work** **b. Helped people in your local community, for example, help a neighbor, or do volunteer work** c. Read for fun d. Gone to the library outside of school **e. Helped your parents or guardians with chores**	Mean across items (a), (b), and (e) using the following scale: Just once or twice during the past year (1) Every couple of months (2) About once a month (3) About once a week (4) Almost every day (5)
Index of good behavior, parent report	Parent survey F7. During the 2012-2013 school year, how often did [STUDENT] do the following things? **a. Help you with chores or other tasks** **b. Stay and help teachers in (his/her) classrooms** **c. Help people in your local community, for example, help a neighbor, or do volunteer work** **d. Read for fun** **e. Go to the library outside of school**	Mean across all items using the following scale: Never (1) Sometimes (2) Often (3)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index indicating well-adjusted student	Parent survey F3. For each of the following statements, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree." **a. The student gets along with others** **b. The student likes school** **c. The student works hard at school** **d. The student is self-confident** **e. The student is creative** **f. The student is happy** **g. The student respects adults**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of Parental Concerns About Student	Parent survey F4. For each of the following statements, please tell me if it is "not a problem," "a small problem," "a medium problem," or "a big problem" with [STUDENT] in or out of school. **a. Getting into trouble** **b. Smoking, drinking alcohol or using drugs** **c. The friends (he/she) has chosen** **d. (His/Her) academic achievement**	Mean across all items using the following scale: Not a problem (1) A small problem (2) A medium problem (3) A big problem (4)
Index of student's feelings about school	Student survey A1. Now, I'm going to read you some statements on how you (feel/felt) about school. For each, statement, please tell me if you "strongly agree," "agree," "disagree," or "strongly disagree". a. You have good friends at your school **b. You are treated fairly at your school** **c. You are happy to be at your school** **d. You feel like you are part of the community in your school** **e. You feel safe in your school** **f. You are treated with respect at your school** g. You know how you are doing in school h. You have the materials and equipment you need to do your school work right **i. You get the chance to be independent at school** j. You have opportunities to choose how you learn	Mean across items (b), (c), (d), (e), (f), (i), and (j) using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of parental satisfaction with school	Parent survey D6. Please rate each of the following features of the school the student attends/attended for the 2012-2013 school year, as "excellent," "good," "fair," or "poor." **a. Facilities, like the classrooms, library, cafeteria, or the gym** **b. Academics, the teachers, and classes** **c. Safety** **d. Discipline**	Mean across all items using the following scale: Poor (1) Fair (2) Good (3) Excellent (4)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of student perceptions of schoolmates	Student survey A2. Please tell me how much you agree or disagree with these statements about the students in your classes (this/last) year at school: **a. Students usually complete their homework** **b. Students get along well with the teachers** **c. Students are interested in learning** **d. Students help one another** e. Students are well behaved	Mean across items (a), (b), (c), and (d) using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of student perceptions of teachers	Student survey A3. These next statements are about your teachers (this/last) year at school. Again please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree" with each statement. **a. They are available for help** **b. They listen to what you have to say** **c. They give corrections and suggestions for improvement** **d. They care about students** **e. They encourage you to think about your future** **f. Their classes are challenging** **g. They make you feel like your school work is important** **h. You like your teachers**	Mean across all items using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of school disciplinary environment	Student survey A4. For each of the following statements about the rules (this/last) year at your school, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree." **a. Everyone knows what the school rules are** **b. The school rules are fair** **c. The punishment for breaking school rules is the same no matter who you are** **d. If a school rule is broken, students know what the punishment will be** **e. Students receive specific positive rewards for good behavior** **f. You follow the rules at school**	Mean across all items using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of parental perceptions of problems in student's school	Parent survey D1. For each of the following issues, please tell me if you feel it [is/was] "not a problem," "a small problem," "a medium problem," or "a big problem" at this school. **a. Students destroying property** **b. Students being late for school** **c. Students missing school** **d. Fighting** **e. Bullying** **f. Stealing**	Mean across all items using the following scale: Not a problem (1) A small problem (2) A medium problem(3) A big problem (4)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of parent outreach to school	Parent survey C3. Since the beginning of this school year, have you or any other adults in your household… **a. Attended an open house or a back-to-school night?** **b. Attended a meeting of a PTA, PTO, or Parent-Teacher Student Organization?** **c. Gone to a parent-teacher conference or meeting with his/her teacher?** **d. Attended a school or class event, such as a play, sports event, or science fair?** **e. Acted as a volunteer at the school or served on a committee?** **f. Participated in fundraising for the student's school?** **g. Visited the school's website for information?**	Mean across all items using the following scale: No (0) Yes (1)
Index of school efforts to engage parents in school	Parent survey C4. How well would you say the student's school… **a. Lets you know between report cards how the student is doing in school?** **b. Helps you understand what children at the student's age are like?** **c. Makes you aware of chances to volunteer at the school?** **d. Provides workshops, materials, or advice about how to help the student learn at home?** **e. Provides information on community services to help the student or your family?** **f. Communicates with you through email or online forums?** **g. Provides online resources?**	Mean across all items using the following scale: Not at all (1) Just O.K (2) Very well (3)
Index indicating school is too easy	Parent survey B4. Do you think the homework (is/was) too difficult, about right, or too easy for the student? Parent survey B5a. Do you think the material covered in (his/her) math class (is/was) too difficult, about right, or too easy for the student? Parent survey B6a. Do you think the material covered in (his/her) English/language arts class (is/was) too difficult, about right, or too easy for the student?	Mean across all items using the following scale: Too easy (1) About right/Too difficult (0)
Index indicating school is too difficult	See items B4, B5a, and B6a above	Mean across all items using the following scale: This responses (0)

Table B.11. Description of indices: high school student survey

Variable	Principal survey items included	Scale/definition
Index of positive school engagement	C3. Please select how often you (do/did) each of the following things: **a. I (stick/stuck) with a class assignment or task until it (is/was) done** **b. I (ask/ed) a teacher for help when I (don't/didn't) understand an assignment** **c. I (take/took) part in class discussions or activities** **d. I (receive/received) recognition or praise for doing good school work** **e. I (learn/learned) from my mistakes at school** **f. I (complete/completed) class assignments, projects, and homework on time**	Mean across items (a) through (f) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of negative school engagement	C3. Please select how often you (do/did) each of the following things: **g. I (think/thought) of dropping out of school** **h. I (try/tried) to stay home from school** **i. I (am/was) absent from school**	Mean across items (g), (h), and (i) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of self control	C4. Last week in school, how many days did you do each of the following things? a. I went to all of my classes prepared b. I remained calm even when things happened that could upset me c. I paid attention in all of my classes d. I listened to other students speak without interrupting them e. I was polite to adults and other students f. I remembered and followed directions g. I controlled my temper h. I got to work right away rather than procrastinating	Mean number of days across all items
Index of academic collaboration	C2. For the following statements, select how much you agree or disagree with each. **a. I like to work with other students** b. I learn things quickly in most school subjects c. Because reading is fun, I wouldn't want to give it up d. I am good at most school subjects **e. I learn most when I work with other students** f. English/Language Arts is one of my best subjects **g. I do my best work when I work with other students** h. Math is one of my best subjects **i. I like to help other people do well in group assignments** j. I do well in tests in most school subjects **k. It is helpful to put together everyone's ideas when I work on a project** l. If I don't do well on a test, I study harder next time m. I set aside time to do my homework and study n. I try to do well on my schoolwork even when it isn't interesting to me o. Grades in high school matter for success in college p. What I learn in class is necessary for success in the future	Mean across items (a), (e), (g), (i), and (k) using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of academic motivation	C2. For the following statements, select how much you agree or disagree with each. **l. If I don't do well on a test, I study harder next time** **m. I set aside time to do my homework and study** **n. I try to do well on my schoolwork even when it isn't interesting to me** **o. Grades in high school matter for success in college** **p. What I learn in class is necessary for success in the future**	Mean across items (l) through (p) using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of academic confidence	C1. For each please select if these things apply to you "almost always," "often," "sometimes," or "almost never." **a. I (am/was) certain I (can/could) understand even the most difficult material presented in textbooks or other written material** **b. I (can/could) learn something really difficult when I (want/ed) to** c. In school I (work/ed) as hard as possible **d. I (am/was) certain I (can/could) understand even the most difficult material presented by the teacher** **e. If I (decide/d) to not get any bad grades, I (can/could) really do it** **f. I (am/was) certain I (can/could) do an excellent job on assignments and tests** g. I (work/worked) hard in school so I (can/could) get a good job h. I (work/ed) hard in school so I (can/could) go to college **i. If I (want/ed) to learn something well, I (can/could)** **j. (am/was) certain I (can/could) master the material I (am/was) taught** k. If I (don't/didn't) understand something in my schoolwork, I (try/tried) to find other resources to help me learn	Mean across items (a), (b), (d), (e), (f), (i), and (j) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of effort in school	C1. For each please select if these things apply to you "almost always," "often," "sometimes," or "almost never." **a. I (am/was) certain I (can/could) understand even the most difficult material presented in textbooks or other written material** **b. I (can/could) learn something really difficult when I (want/ed) to** c. In school I (work/ed) as hard as possible **d. I (am/was) certain I (can/could) understand even the most difficult material presented by the teacher** **e. If I (decide/d) to not get any bad grades, I (can/could) really do it** **f. I (am/was) certain I (can/could) do an excellent job on assignments and tests** g. I (work/worked) hard in school so I (can/could) get a good job h. I (work/ed) hard in school so I (can/could) go to college **i. If I (want/ed) to learn something well, I (can/could)** **j. I (am/was) certain I (can/could) master the material I (am/was) taught** k. If I (don't/didn't) understand something in my schoolwork, I (try/tried) to find other resources to help me learn	Mean across items (c), (g), (h), and (k) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of student reports of discussions about college at school	F7. (Has/Did) anyone at your school (discussed/discuss) the following with you? **a. Different admissions requirements for two-year vs. four-year colleges** **b. Different admissions requirements among four-year colleges** **c. How to decide which college to attend** **d. Your likelihood of being accepted at different types of schools** **e. What ACT/SAT scores you need to get into the colleges you want to attend** **f. Opportunities to attend out-of-state schools** **g. Your readiness for college-level coursework** **h. What kinds of study skills you will need in college or vocational/technical school** **i. How to pay for college**	Mean across all items using the following scale: Did not discuss (1) Discussed briefly (2) Discussed in depth (3)
Index of student reports of teachers/counselor assistance with planning for college	F8. How much (do/did) your high school teachers or guidance counselors… **a. Encourage you to apply to several different schools** **b. Talk to you about what college would be like** **c. Help you fill out applications for colleges or vocational/technical schools** **d. Help you find scholarships to apply for** **e. Help you decide which school to attend** **f. Help you plan how to pay for tuition and other expenses** **g. Help you with your college application essays or personal statements**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of student reports of parent assistance with planning for college	F9. How much (do/did) your parent(s)/guardian(s)… **a. Encourage you to apply to several different schools** **b. Talk to you about what college would be like** **c. Help you fill out applications for colleges or vocational/technical schools** **d. Help you find scholarships to apply for** **e. Help you decide which school to attend** **f. Help you plan how to pay for tuition and other expenses** **g. Help you with your college application essays or personal statements**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)
Index of college preparation activities	F5. At anytime during high school did you. . . **a. Attend college fairs** **b. Speak with college representatives** **c. Visit in-state college campuses** **d. Visit out-of-state college campuses** e. Sit in on a college-level course f. Participate in a college summer program or early college access program **g. Take practice ACT/SAT exams** h. Use college guidebooks (on-line or print) **i. Obtain information from college websites** j. Apply for financial aid k. Apply for a scholarship **l. Research career possibilities**	Mean across items (a), (b), (c), (d), (g), (i), and (l) using the following scale: No (0) Yes (1)
Index of student's feelings about school	B1. Here are some statements related to how you feel about your school. For each statement, please select whether you “strongly agree,” “agree,” “disagree,” or “strongly disagree.” **a. I (have/had) good friends at my high school** **b. I (am/was) treated fairly at my high school** **c. I (am/was) happy to be at my high school** **d. I (feel/felt) like I (am/was) part of the community in my high school** **e. I (feel/felt) safe in my high school** **f. I (am/was) treated with respect at my high school** **g. I (know/knew) how I (am/was) doing in high school** **h. I (have/had) the materials and equipment I need(ed) to do my school work right** **i. I (get/got) the chance to be independent at high school** **j. High school (is/was) seen as preparation for the future** **k. All students (are/were) encouraged to go to college**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of student perceptions of schoolmates	B2. Please select how much you agree or disagree with these statements about the students in your classes at your school. **a. Students usually (complete/d) their homework** **b. Students (get/got) along well with the teachers** **c. Students(are/were) interested in learning** **d. Students(help/ed) one another** **e. Students (are/were) well behaved**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of student perceptions of teachers	B3. Please select how much you agree or disagree with these statements about your teachers at your school. **a. Teachers are/were available for help** **b. Teachers listen/ed to what I have to say** **c. Teachers give/gave corrections and suggestions for improvement** **d. Teachers care/d about students** **e. Teachers make/made me feel like my school work is important** **f. Teachers work/ed hard to make sure that all students are learning** **g. Teachers work/ed hard to make sure that students stay in school** **h. I like/d my teachers**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of school disciplinary environment	B5. Please select to what extent you agree or disagree with the following statements about your school. **a. Behavioral standards and discipline policies (are/were) established and enforced consistently across the entire school** **b. The school (has/had) a zero-tolerance policy for potentially dangerous behaviors** **c. The school (has/had) a school-wide behavior code that includes specific positive rewards for students who consistently (behave/d) well** **d. The school (has/had) a school-wide behavior code that (includes/ed) specific consequences for students who break the rules** **e. I (follow/ed) the rules at school**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of helpfulness of high school in career planning	D2. How helpful has/was your school in the following: **a. Assessing your career interests and abilities** **b. Developing a career plan** **c. Providing information about occupations (e.g., salaries, working conditions, and future outlook of various occupations)** **d. Teaching job search techniques (e.g., where/how to look for jobs)** **e. Teaching resume writing** **f. Helping you find a job** **g. Teaching job interviewing skills**	Mean across all items using the following scale: Not helpful (1) Somewhat helpful (2) Helpful (3) Very helpful (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of teacher helpfulness and encouragement	D3. Think about the teachers that you (have) met with throughout high school. Please indicate how much your teachers (have done/did) the following: **a. Helped me select courses that (meet/met) my high school's graduation requirements** **b. Helped me select courses that I (need/ed) for work or admission to college** **c. Helped me decide what I (want/ed) to do after I graduate** **d. Encouraged me to take Advanced Placement (AP)/honors courses** **e. Encouraged me to continue my education after high school** **f. Talked to me about how to get a job** **g. Talked to me about colleges/schools that (are/were) suited to my interests and abilities**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)
Index of counselor helpfulness and encouragement	D4. Thinking about the counselors that you have met with throughout high school. Please indicate how much your counselors have done the following: **a. Helped me select courses that (meet/met) my high school's graduation requirements** **b. Helped me select courses that I (need/ed) for work or admission to college** **c. Helped me decide what I (want/ed) to do after I graduate** **d. Encouraged me to take Advanced Placement (AP)/honors courses** **e. Encouraged me to continue my education after high school** **f. Talked to me about how to get a job** **g. Talked to me about colleges/schools that (are/were) suited to my interests and abilities**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)

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

APPENDIX C

CUMULATIVE MIDDLE AND HIGH SCHOOL RESULTS

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

Our primary analyses of KIPP high schools employ two different approaches. The first analysis estimates the *marginal* impact of KIPP high schools among students who previously attended a KIPP middle school. This involves comparing one group of KIPP middle school students (those who had the option to attend a KIPP high school) to a different group of KIPP middle school students (those who did not have the option to attend a KIPP high school, because there was no KIPP high school available in their area at the time they completed middle school). A more detailed description of this analysis can be found in Appendix H. The second analysis focuses on students who entered KIPP for the first time in grade 9 and matches to these new entrants a set of comparison students who never attended KIPP at any level (elementary, middle, or high school). These matches between KIPP and comparison students were based on the students' demographic characteristics and "baseline" test scores in grade 7 and grade 8 (see Appendix G for more information about the matching model and other details of this second analysis).

As an additional exploratory analysis, this appendix presents the results of a third analysis of high school outcomes that examines the *combined* impact of KIPP middle schools and KIPP high schools. These estimates compare a set of KIPP high school students who came through KIPP middle schools with a matched set of comparison students who never attended KIPP at any point in their schooling. The KIPP and non-KIPP groups both attended non-KIPP elementary schools, but at that point their educational paths diverged. Since the groups attended different middle schools and high schools, the resulting impact estimates capture the cumulative effect of KIPP at both levels, and cannot distinguish between the separate effects of KIPP middle versus high schools.

The cumulative impact analysis of KIPP middle and high schools on high school outcomes uses a matched-student design. In particular, students who attended KIPP middle and high schools after attending non-KIPP elementary schools were identified, and comparison students who attended non-KIPP schools at all levels were matched to them. This matching was based primarily on their demographic characteristics and test scores in elementary school (typically grades 3 and 4). In addition, each of these KIPP students was matched to a comparison student who did not attend a KIPP middle or high school, but did remain in our data from elementary school through grade 9. In other words, students who dropped out or transferred to a different school district during middle school were not eligible for the comparison group. This requirement was added because the group of KIPP students, by definition, also remained in our data (and in KIPP schools) through grade 9, at minimum.

The sample of KIPP middle schools and linked high schools in this analysis (as well as the number of KIPP and matched comparison students in the sample for each school) is summarized in Table C.1. Overall, the analysis is based on a sample of 7,404 students and includes KIPP students who went through 25 KIPP middle schools and 14 KIPP high schools. To match the KIPP students and comparison group, we used the same propensity score model and imputation procedures as we used for the matched-student analysis of KIPP middle school impacts (described in Appendix F). The impact models also closely resemble the models used for our main analysis of high school outcomes described in Appendix G; the only difference is that for this supplemental analysis baseline test scores were defined in elementary school rather than middle school grades.

Table C.1. Cumulative middle and high school analysis sample

Middle school	Linked high school	Analytic baseline sample			Number of KIPP cohorts in data (School Years)
		Treatment (N)	Comparison (N)	Total sample size	
KIPP Academy Lynn Middle School	KIPP Academy Lynn Collegiate High School	182	182	364	4 (2006-07 to 2009-10)
KIPP WAYS Academy	KIPP Atlanta Collegiate	120	120	240	5 (2006-07 to 2010-11)
KIPP STRIVE Academy	KIPP Atlanta Collegiate	35	35	70	2 (2009-10 to 2010-11)
KIPP Austin College Prep	KIPP Austin Collegiate	308	308	616	7 (2003-04 to 2009-10)
KIPP Austin Acad. of Arts & Letters	KIPP Austin Collegiate	49	49	98	1 (2009-10)
KIPP DC: KEY Academy	KIPP DC College Preparatory	130	130	260	6 (2005-06 to 2010-11)
KIPP DC: AIM Academy	KIPP DC College Preparatory	168	168	336	6 (2005-06 to 2010-11)
KIPP DC: WILL Academy	KIPP DC College Preparatory	84	84	168	5 (2006-07 to 2010-11)
KIPP Delta College Preparatory School	KIPP Delta Collegiate High School	115	115	230	5 (2005-06 to 2009-10)
KIPP Sunshine Peak Academy	KIPP Denver Collegiate High School	254	254	508	7 (2004-05 to 2010-11)
KIPP 3D Academy	KIPP Generations Collegiate	163	163	326	5 (2006-07 to 2010-11)
KIPP Intrepid Preparatory	KIPP Generations Collegiate	69	69	138	3 (2008-09 to 2010-11)
KIPP Academy (Houston)	KIPP Houston High School	360	360	720	7 (2003-04 to 2009-10)
KIPP Sharpstown College Prep	KIPP Houston High School	130	130	260	3 (2007-08 to 2009-10)
KIPP Memphis Collegiate Middle	KIPP Memphis Collegiate High	157	157	314	4 (2008-09 to 2010-11)
KIPP Polaris Academy for Boys	KIPP Northeast College Prep	22	22	44	2 (2009-10 to 2010-11)
KIPP Voyage Academy for Girls	KIPP Northeast College Prep	35	35	70	2 (2009-10 to 2010-11)
KIPP Academy Middle School (New York)	KIPP NYC College Prep High School	162	162	324	6 (2004-05 to 2009-10)
KIPP STAR Harlem Middle School	KIPP NYC College Prep High School	147	147	294	6 (2004-05 to 2009-10)
KIPP Infinity Middle School	KIPP NYC College Prep High School	177	177	354	5 (2005-06 to 2009-10)
KIPP AMP Middle School	KIPP NYC College Prep High School	82	82	164	5 (2005-06 to 2009-10)
KIPP Gaston College Preparatory	KIPP Pride High School	315	315	630	9 (2001-02 to 2009-10)
KIPP Liberation College Prep	KIPP Sunnyside	64	64	128	5 (2006-07 to 2010-11)
KIPP Spirit College Prep	KIPP Sunnyside	139	139	278	5 (2006-07 to 2010-11)
KIPP Aspire Academy	KIPP University Prep High School	235	235	470	6 (2004-05 to 2009-10)
Total	**25 middle schools** **14 high schools**	**3,702**	**3,702**	**7,404**	

Notes: Test outcomes are drawn from administrative records collected following enrollment in high school. Treatment students attended KIPP in both middle and high school, and comparison students are matched on baseline (elementary school) characteristics.

In comparison to the other matched-student analyses of KIPP schools presented in the main text of this report, the matched student design for estimating the cumulative impact of KIPP middle and high schools relies on a stronger set of assumptions. In particular, the matched-student analysis of middle school impacts uses matching on elementary school characteristics to account for students' decision to enter or not enter a KIPP middle school (that is, to account for selection into middle school). By contrast, the matched-student analysis of cumulative KIPP middle and high school impacts uses matching on elementary school characteristics to account for both selection into a KIPP middle school *and persistence in KIPP into the high school grades*. Persistence in a KIPP middle school may be correlated with attributes that we cannot observe in our data, to the extent that factors such as parental involvement, discipline, and students' "grit" and determination may not be fully reflected in their baseline test scores or demographic characteristics.[1] If these unobserved factors are also associated with improved outcomes in high school, the impact estimates we discuss below would be upwardly biased. We believe this potential for upward bias may be particularly likely to occur for the graduation outcome, because choosing to persist in the KIPP network throughout middle school may also predict persistence to graduation in high school. As a result, we believe the findings from this supplemental analysis should be interpreted with caution.[2]

Table C.2 presents the results. We find a pattern of positive and statistically significant impacts in all the academic subjects we examined (math, ELA, science, and social studies).[3] For linked KIPP middle and high schools, the estimated impacts in math, ELA, science, and social studies are 0.34, 0.29, 0.40, and 0.27 standard deviations, respectively. Relative to the high school test scores in the matched comparison group, these impacts represent an increase from the 49th percentile (equivalent to the average percentile of the comparison group) to the 63rd percentile in math, an increase from the 53rd to the 64th percentile in ELA, an increase from the 50th to the 65th percentile in science, and an increase from the 46th to the 56th percentile in social studies relative to the distribution of students in the same district. We also estimate that this group of linked KIPP middle and high schools increased four-year graduation rates by 13 percentage points, with a graduation effect that is positive and statistically significant.

[1] In our data, on average less than half of students entering a KIPP middle school and had the option to attend a KIPP high school persisted at KIPP into grade 9.

[2] On observable characteristics, however, the treatment and comparison groups are very similar for each of the outcome samples we examined. Full baseline equivalence results for this analysis are available from the authors upon request.

[3] For more information on the data and measures used for these high school outcomes, see Appendix G.

Table C.2. Cumulative impact of KIPP middle and high school on high school achievement and graduation

	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
Impacts on High School Standardized Subject Tests (z-scores)					
Mathematics achievement	0.32	-0.03	0.34** (0.04)	11	3,122[a]
ELA achievement	0.36	0.07	0.29** (0.02)	14	4,208
Science achievement	0.39	-0.01	0.40** (0.03)	14	3,704
Social studies achievement	0.16	-0.10	0.27** (0.05)	9	1,573[a]
Impacts on high school graduation rates (percentage points)					
Graduation within 4 years after entering grade 9	78.6	65.3	13.3** (2.2)	9	2,216

Source: State and district administrative records data.

Notes: Impacts for matched student analyses were calculated by comparing the outcomes of KIPP students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. Impacts were calculated separately for each KIPP high school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. In a given high school, the included test may be a either an end-of-course exam (i.e. algebra), or an end-of-grade exam (i.e. grade 10 mathematics). Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Standard errors are shown in parentheses.

[a] The high school exams used for this analysis varied by jurisdiction. For each subject and site, we selected the exam that was observed for the largest number of students during high school, provided that the percentage of students taking the exam was similar in the treatment and matched comparison group. Each site (and each student cohort within sites) was only included in the analysis of a given test if both the treatment and matched comparison groups took the relevant exam during high school at a similar rate. In this sample, data coverage was less consistent for mathematics and social studies exams, compared to ELA and science exams.

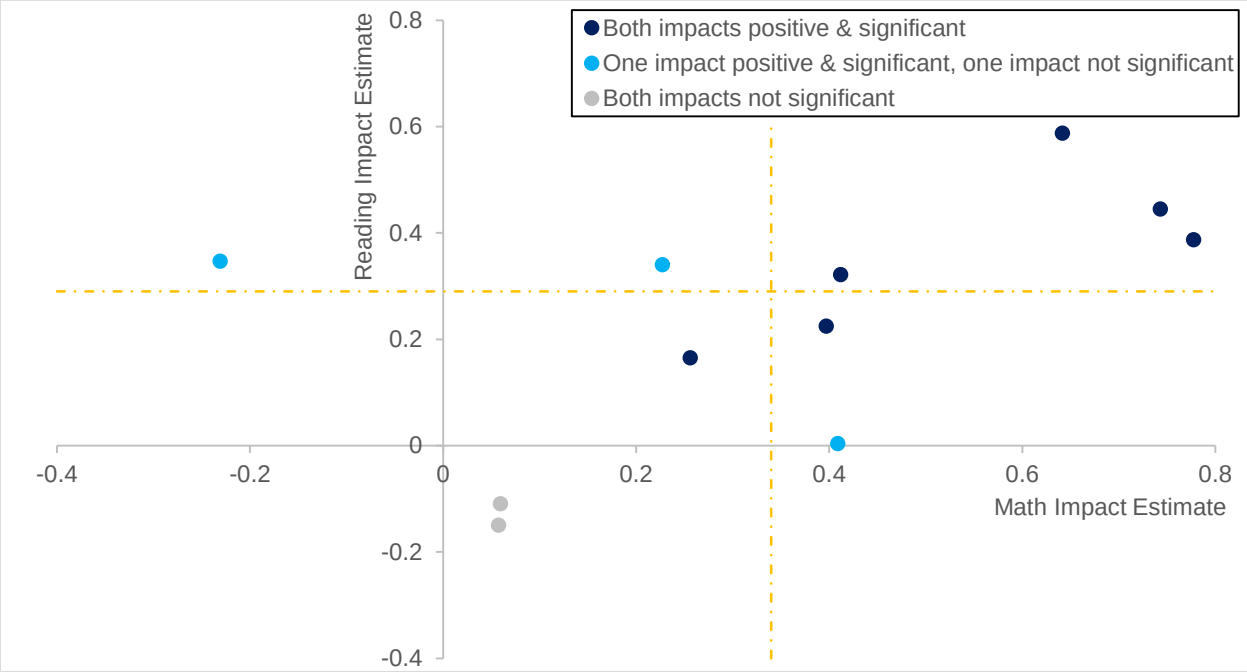
* Impact estimate is statistically significant at the 0.05 level, two-tailed test.

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

To summarize the variation in these impact estimates, Figure C.1 plots both the ELA and math impacts for the 11 sites with high school data on both outcomes. Of these, six sites have a significant and positive impact in both ELA and math, three sites have a significant and positive impact in one subject but not the other, and two sites do not have significant impacts in either subject. None of the sites has a negative and statistically significant impact on test scores in ELA or math.

While the magnitudes of these impacts are substantial, it is important to remember that the estimates represent the combined effect of KIPP middle schools and KIPP high schools. The analysis does not distinguish between the relative contributions from these two types of KIPP schools. This differs from both the analysis of new grade 9 entrants to KIPP and the matched-school analysis of the marginal impacts of KIPP high schools on students continuing from a KIPP middle school—both of those analyses examined the impacts of KIPP high schools alone. This makes it more difficult to interpret differences between the magnitudes of the impact estimates from this supplemental set of results and the study's primary two analyses of high school outcomes.

Figure C.1. Distribution of cumulative middle and high school impacts in ELA and math



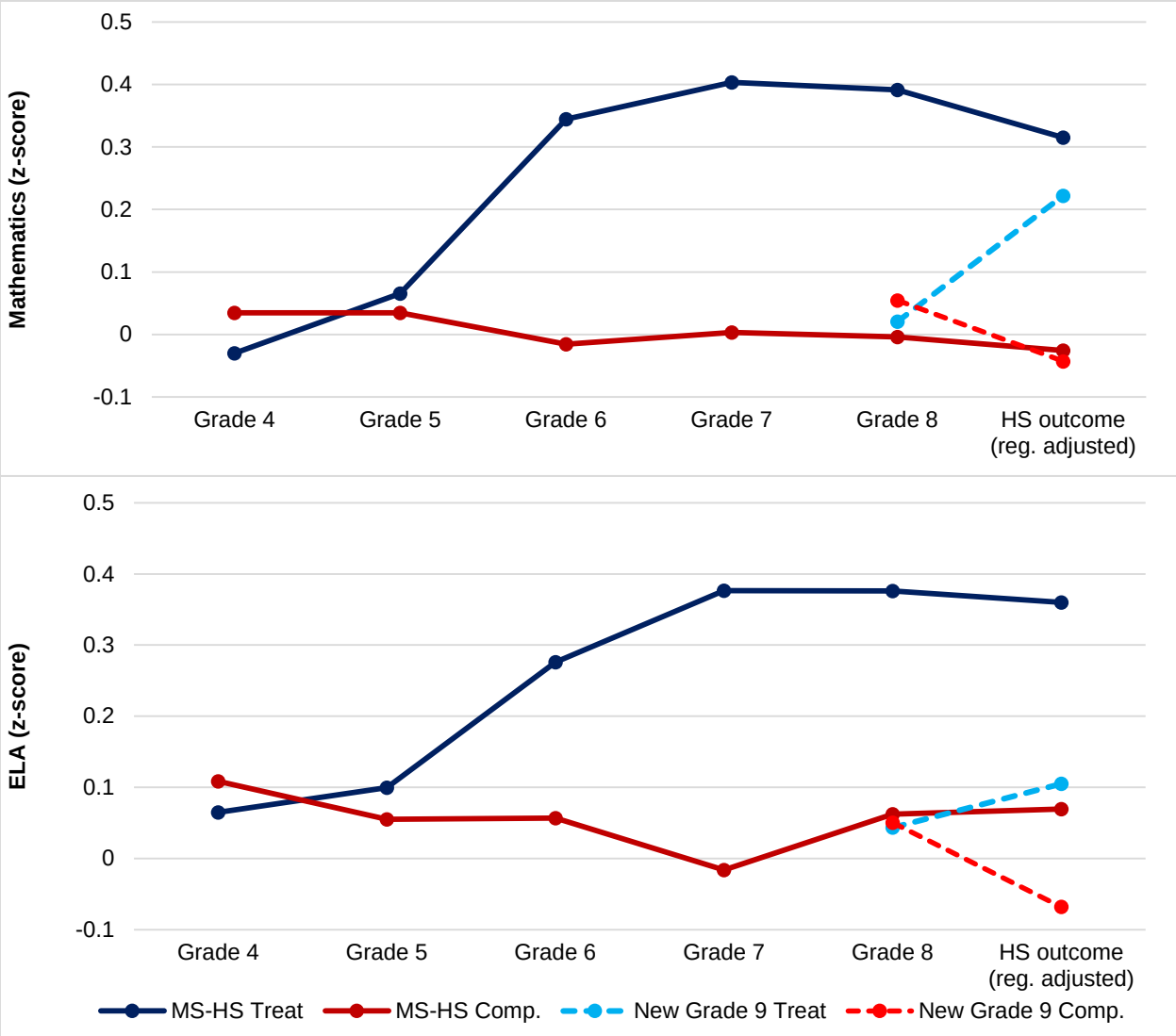
Note: Each circle represents the math and ELA impact estimate for one KIPP site (a linked set of one or more middle schools and one high school). Dark-blue circles indicate that impacts in both subjects are statistically significant and positive at the 0.05 level, two-tailed test. Light-blue circles indicate that the impact in only one of the two test subjects is statistically significant and positive. Grey indicates that both impacts are statistically indistinguishable from zero. No site has a statistically significant and negative impact in either subject. The dashed orange lines represent the average impact in math (vertical line) and ELA (horizontal line).

Compared to the primary high school impact estimates discussed in our main report, we find a pattern of larger impacts on high school test scores when we consider the effect of KIPP middle and high schools combined, relative to students who never attend a KIPP school at any level. The impacts of KIPP high schools on students who are new to the network in grade 9 are slightly smaller than these cumulative impacts, but the effects are still large and statistically significant. In the matched-school design that produced estimates of the marginal impacts of KIPP high schools on students who are continuing from a KIPP middle school, we found that those impacts were not statistically significant. These estimates from the matched-school analysis suggest that attending a KIPP high school neither helps nor harms students who also attended a KIPP middle school in terms of their high school achievement, on average.

To investigate the possible contribution of KIPP middle schools and high schools further, we compared the average achievement levels throughout middle school and high school of middle school entrants and grade 9 entrants to KIPP, alongside achievement levels for the non-KIPP comparison groups in each matched-student impact analysis (Figure C.2). The solid lines in the figure represent KIPP and non-KIPP (matched comparison group) test score outcomes, by grade, for students in the matched-student analysis of cumulative KIPP middle and high school impacts. The dotted lines represent outcomes by grade for students in the matched-student analysis of KIPP high school impacts for new entrants. In each case, matching produced similar

test scores for KIPP and comparison group students in the baseline period before entry into KIPP (grade 4 for the cumulative analysis and grade 8 for the new entrant analysis). The divergence in these estimates after the baseline period reflect possible impacts of KIPP.

Figure C.2. Average test scores by grade level, comparing the MS-HS analytic sample to the analytic sample of new grade 9 entrants to KIPP



Notes: This figure reports the average math and reading (in middle school) or ELA (in high school) z-score in each grade for four different groups of students: (1) the MS-HS analysis treatment group (solid blue); (2) the matched comparison group for the MS-HS analysis (solid red); (3) the treatment group for the analysis of students entering KIPP for the first time in grade 9 (dashed blue); and (4) the matched comparison group for the students entering KIPP in grade 9 (dashed red). The sample is restricted to students with an observed high school test score for each outcome. The HS outcome scores are regression adjusted, such that the difference between a treatment group and its associated matched comparison group equals the study's impact estimate for that analysis. The average scores shown in grade 4 through grade 8 are not regression adjusted.

The patterns in this figure suggest that middle school KIPP entrants and grade 9 new entrants to KIPP both experience large achievement impacts in their first few years after entry into the KIPP network, but for the middle school entrants the gains appear to be concentrated in middle school rather than high school years. In other words, it is possible that the new grade 9 entrants to KIPP could have attributes (such as low middle school test scores) that help make them more receptive to the "achievement benefits" of KIPP high schools than students who arrived from a KIPP middle school and have already been exposed to KIPP for several years. This pattern of results is also consistent with the matched-school design presented in the main text, in which the estimated marginal impact of KIPP high schools among those who also attended KIPP middle schools was not statistically significant.

To test whether our benchmark results are sensitive to our baseline test score imputation strategy described in Appendix F, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores. The results for this smaller sample are nearly identical to our benchmark impact estimates for the cumulative impact of KIPP middle and high schools (Table C.3). There are no statistically significant differences on any baseline measure and the KIPP impact remains positive and statistically significant for all test outcomes and high school graduation. The magnitude of each impact estimate is nearly identical to the benchmark estimate as well.

Table C.3. Baseline equivalence and impact estimates on sample with non-imputed baseline data (cumulative impact of KIPP middle and high School)

	Treatment group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	0.065	2072	0.100	1929	-0.035	0.261
Math scores (math)	-0.032	1516	0.037	1414	-0.069	0.083
Reading scores (social studies)	0.117	788	0.109	707	0.007	0.871
Math scores (science)	0.065	1946	0.133	1636	-0.067	0.078
Reading scores (4-year graduation)	0.037	1025	0.055	1008	-0.018	0.658
Math scores (4-year graduation)	0.019	1025	0.080	1008	-0.061	0.138
Free and reduced-price lunch status (4-year graduation)	0.883	1025	0.892	1008	-0.009	0.562
	Treatment group		**Comparison group**			
Outcome measure	**Adjusted mean**	**Sample size**	**Mean**	**Sample size**	**Impact estimate**	**p-value**
ELA achievement	0.381	2072	0.085	1929	0.295**	0.000
Mathematics achievement	0.341	1516	-0.001	1414	0.341**	0.000
Science achievement	0.417	788	0.001	707	0.417**	0.000
Social studies achievement	0.180	1946	-0.087	1636	0.267**	0.000
Four-year high school graduation	0.785	1025	0.651	1008	0.134**	0.000

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

APPENDIX D

DETAILED ANALYTIC METHODS: ELEMENTARY SCHOOL (LOTTERY-BASED ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP elementary schools. First we present information on the sample and the baseline equivalence of students who won an admission lottery (the treatment group) and those who did not win (the control group). Next we discuss the data and analytic methods used for the analysis. We conclude the appendix by discussing results of sensitivity analyses.

Detail on sample

Of the 23 KIPP elementary schools open in Spring 2011, 8 were sufficiently oversubscribed to be in the analysis and held admissions lotteries in Spring 2011.[4] Elementary school admissions lotteries are held primarily at the pre-kindergarten and kindergarten level, so the analysis sample is drawn from these two entry grades. Among students in the study sample schools and grades, 1,250 were admitted based on a lottery result (Table D.1).[5] The original lottery sample, as defined above, is larger than required to meet the study's targeted level of statistical power. There was a substantial imbalance in the size of the treatment and control groups at the school level for several schools. In other words, a larger number of students participating in the lotteries at these schools were offered admission and included in the treatment group than the number who were not offered admission and included in the control group, or vice versa. To conserve resources while simultaneously maximizing our ability to detect impacts, a subsample of 1,097 students at these schools was randomly selected to comprise the baseline sample for inclusion in the study's data collection (the Woodcock-Johnson test).

Table D.1. Elementary school student sample sizes (lottery-based analysis)

	Treatment	Control	Overall
Original lottery sample	499	751	1,250
Baseline sample	473	624	1,097
Analytic sample (for achievement outcomes)	284	370	654

In total, 654 students who applied to eight KIPP elementary schools were included in the final analytic sample for which we have study-administered test outcomes (284 in the treatment group and 370 in the control group).[6] These students participated in lotteries to enroll in the entry grade at a KIPP elementary school, and this entry grade varied by school: prekindergarten—age 3 (PK3) at three schools and kindergarten at the remaining five schools. Among the students in the final analytic sample, 79 percent of lottery winners ever enrolled in the KIPP school to which they applied, while 6 percent of those who did not win an admissions lottery still ended up attending KIPP at some point during the follow-up period. The 73 percentage-point difference in

[4] A ninth school was sufficiently oversubscribed to be included in the study, but ultimately dropped from our baseline and analytic samples because more than half the sample at that site lacked outcome data.

[5] Students who apply to oversubscribed schools may be guaranteed admission and thus not be eligible for the study. For example, applicants may be admitted to the school outside of the typical lottery process if they have a sibling already attending the school or if a predetermined number of seats are reserved for district residents and these reserved seats are not oversubscribed.

[6] Overall attrition from the sample of students at baseline was 40 percent (443 out of 1,097). The rates of attrition in the treatment and control groups were very similar (within 1 percentage point).

enrollment rates provides a clear contrast between lottery winners and non-winners in exposure to KIPP schools.

Further, the treatment and control groups remained enrolled in different elementary schools as of the second year following the lottery (Table D.2). For applicants at both grade levels, control group students were likely to attend either a traditional public school or a non-KIPP charter school. The rate of private school attendance was 5 percent or less for both groups.

Table D.2. Type of schools attended by KIPP elementary school applicants

	PK-3 entrants		Kindergarten entrants	
Type of school attended	Percentage of lottery winners (treatment)	Percentage of lottery losers (control)	Percentage of lottery winners (treatment)	Percentage of lottery losers (control)
KIPP charter	77	19	69	3
Non-KIPP charter	12	32	17	39
Private	1	5	2	3
Traditional public	10	44	12	54

Notes: Proportions reflect the schools students attended during the 2012–2013 school year, the second year following admissions lotteries, as identified in the parent survey. Among students included in the achievement analysis, 66 percent had nonmissing data on the school they attended. The proportions reported here reflect school attendance among those students. Type of school was determined using the National Center for Education Statistics Common Core of Data.

Properly executed randomization should ensure that there are no differences (observed or unobserved) between the treatment and control groups. In principle, we can test whether or not this is true by examining the baseline characteristics of the treatment and control groups. The baseline characteristics were assessed using data from a baseline survey administered to the parents of students participating in KIPP lotteries close to the time of random assignment.[7] Information from the baseline survey was supplemented with information from the follow-up parent survey where there was missing data from the baseline survey and where we determined the characteristic should not vary systematically with lottery outcome over time (gender, race/ethnicity, age, and an indicator for single-parent household).

We compared the baseline characteristics of the full data collection sample of students in the treatment and comparison groups and found only one significant difference in twenty characteristics (Table D.3), which is what we would have expected due to chance alone.[8] Since

[7] Overall, about 76 percent of parents of students in the elementary school analysis sample completed the baseline survey. Among these students, about 60 percent completed the baseline survey prior to the time of the lottery and the remaining 40 percent completed it after the lottery. For the group that completed the baseline survey after the lottery, there is some risk that the lottery outcome could have influenced their responses.

[8] One complication of this approach is that we do not have full information on the baseline characteristics of the student sample since they are based on surveys that were not completed by all parents. As a result, treatment and control group differences on these characteristics may reflect both actual differences in the two groups' baseline characteristics and also subsequent patterns of non-response on the baseline and follow-up parent surveys. In other words, if losing the lottery made some parents less likely to complete the surveys and the parents who chose not to complete the survey differed systematically from those who did choose to complete it, this could result in

the estimates of KIPP elementary schools' impact are based on the analytic sample of students who completed a WJ-III test, it is also important to examine whether there are any differences in the baseline characteristics of the treatment and control students for whom we have outcome data on the WJ-III. The second panel of Table D.3 compares the baseline characteristics of students for whom we have at least one test score. For three of the four outcome tests, the sample sizes for students with outcomes are similar, differing by fewer than five students, so a single set of baseline treatment-control comparisons is presented. The third panel presents baseline equivalence results for the Calculation test sample, which is substantially smaller since that test was only administered to students in second grade.

Comparisons of baseline characteristics within the analytic sample indicate that there were some baseline differences between treatment and control group students on these measures. In particular, treatment students are less likely to live in a household with income less than $15,000 and more likely to have a mother who completed college, in addition to being less likely to have a mother with less than a high school education (as in the baseline sample). In the Calculation sample, there is no significant difference in family income but the differences in mother's education is of a larger magnitude. These additional baseline differences in the analytic sample may be driven by different processes in the treatment and control group influencing which students take the WJ-III. It is also possible that the observed differences in family income and mother's education are caused in part by missing data on baseline characteristics. However, in order to account for differences in baseline characteristics that might influence students' achievement, impact models statistically control for all of the baseline characteristics listed in Table D.3.

differences in the overall composition of the treatment and comparison groups on these measures. However, we found no evidence of treatment-control differences even accounting for the possibility of both sources of bias.

Table D.3. Baseline equivalence for elementary school samples

Baseline characteristic	Lottery winners	Lottery losers	Difference	Std. Error	Nt	Nc
Baseline Sample						
Female	0.44	0.48	-0.04	0.04	397	488
Average age in years	4.15	4.17	-0.01	0.03	381	465
White, non-hispanic	0.01	0.01	0.00	0.01	421	516
Hispanic (any race)	0.39	0.40	-0.01	0.02	421	516
Black, non-hispanic	0.57	0.54	0.02	0.03	421	516
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.04	0.04	-0.01	0.01	421	516
English is main language spoken at home	0.62	0.60	0.03	0.03	384	473
Another language is main language spoken at home	0.21	0.19	0.02	0.03	384	473
English and another lang. spoken equally at home	0.17	0.21	-0.04	0.03	384	473
One adult in household	0.26	0.32	-0.06	0.03	422	513
Family income less than 15K	0.24	0.26	-0.03	0.03	363	450
Family income between 15K and 25K	0.25	0.24	0.01	0.03	363	450
Family income between 25K and 35K	0.21	0.20	0.01	0.03	363	450
Family income between 35K and 50K	0.16	0.17	0.00	0.03	363	450
Family income 50K or greater	0.14	0.13	0.01	0.02	363	450
Mother's education: less than HS	0.09	0.17	-0.08**	0.03	378	463
Mother's education: HS or GED	0.25	0.27	-0.01	0.03	378	463
Mother's education: some college	0.34	0.32	0.02	0.03	378	463
Mother's education: college	0.32	0.25	0.07	0.04	378	463
Student has access to computer with internet at home	0.77	0.76	0.00	0.03	379	465
Average number of children's books at home	37.52	39.08	-1.56	3.00	359	433
Analytic Sample (with Test Scores)						
Female	0.43	0.46	-0.03	0.05	258	317
Average age in years	4.17	4.20	-0.03	0.03	249	303
White, non-hispanic	0.01	0.01	0.01	0.01	270	337
Hispanic (any race)	0.38	0.42	-0.04	0.03	270	337
Black, non-hispanic	0.56	0.54	0.02	0.03	270	337
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.04	0.03	0.01	0.02	270	337
English is main language spoken at home	0.61	0.59	0.02	0.03	251	307
Another language is main language spoken at home	0.22	0.21	0.00	0.03	251	307
English and another lang. spoken equally at home	0.17	0.20	-0.02	0.03	251	307
One adult in household	0.26	0.29	-0.03	0.04	272	335
Family income less than 15K	0.22	0.29	-0.08*	0.04	238	295
Family income between 15K and 25K	0.24	0.24	0.00	0.04	238	295
Family income between 25K and 35K	0.21	0.17	0.04	0.04	238	295
Family income between 35K and 50K	0.17	0.18	-0.01	0.04	238	295
Family income 50K or greater	0.17	0.12	0.05	0.03	238	295
Mother's education: less than HS	0.08	0.18	-0.10**	0.03	248	302
Mother's education: HS or GED	0.22	0.28	-0.06	0.04	248	302
Mother's education: some college	0.34	0.28	0.06	0.04	248	302
Mother's education: college	0.36	0.25	0.11*	0.05	248	302
Student has access to computer with internet at home	0.77	0.76	0.01	0.04	247	303
Average number of children's books at home	38.99	40.77	-1.78	3.77	231	280

Table D.3 *(continued)*

Baseline characteristic	Lottery winners	Lottery losers	Difference	Std. Error	Nt	Nc
Analytic subsample with calculation score						
Female	0.36	0.39	-0.03	0.06	166	175
Average age in years	5.01	5.03	-0.02	0.04	158	164
White, non-hispanic	0.01	0.01	0.00	0.01	168	179
Hispanic (any race)	0.49	0.46	0.03	0.04	168	179
Black, non-hispanic	0.46	0.51	-0.04	0.04	168	179
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.02	0.01	0.01	168	179
English is main language spoken at home	0.57	0.56	0.01	0.04	159	166
Another language is main language spoken at home	0.24	0.20	0.04	0.04	159	166
English and another lang. spoken equally at home	0.19	0.23	-0.04	0.05	159	166
One adult in household	0.25	0.29	-0.04	0.05	170	179
Family income less than 15K	0.22	0.29	-0.07	0.05	152	160
Family income between 15K and 25K	0.23	0.22	0.01	0.05	152	160
Family income between 25K and 35K	0.21	0.17	0.03	0.06	152	160
Family income between 35K and 50K	0.21	0.20	0.01	0.05	152	160
Family income 50K or greater	0.12	0.11	0.02	0.03	152	160
Mother's education: less than HS	0.09	0.23	-0.14**	0.04	156	163
Mother's education: HS or GED	0.22	0.25	-0.03	0.05	156	163
Mother's education: some college	0.40	0.26	0.14**	0.05	156	163
Mother's education: college	0.30	0.27	0.03	0.06	156	163
Student has access to computer with internet at home	0.81	0.74	0.06	0.05	155	164
Average number of children's books at home	44.56	45.95	-1.39	5.33	145	154

Source: Baseline characteristics are drawn from a baseline survey of parents. Missing values for age, gender, race/ethnicity, and a single parent household indicator from the baseline survey are filled in using information from the follow-up survey, where possible.

Note: All values in this table are based on non-imputed data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in the “Lottery Winner” and “Non-Winner” columns.

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

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

Detail on achievement measures

For the elementary school sample, academic achievement was measured using scores on the WJ-III Letter-Word Identification (Test 1) and Reading Comprehension (Test 9) tests in reading and on the Calculation (Test 5) and Applied Problems (Test 10) tests in math. We administered the WJ-III tests to all students in spring 2014, the third follow-up year, regardless of entry grade or age. Typically students in the PK3 sample would be in kindergarten at this time and students in the kindergarten sample would be in second grade. The Calculation test was administered only to the sample of students who participated in lotteries to enroll in a KIPP kindergarten, since it was not age appropriate for the PK3 sample.

Students' scores on the WJ-III tests were standardized (into z-scores) using information on the performance of a nationally representative norming population. Thus, values reflect students' performance relative to the national population: positive values indicate that sampled students outperformed the average student nationally (in the norming population) and negative values indicate that sample students performed below the national average. The standardized score has been scaled so that a 1 unit change represents 1 standard deviation of the national population. For example, a z-score of 1 for a given student would indicate that the student's score was 1 standard deviation above that of the average student nationally, which would put them at about the 84th percentile.

Detail on analytic methods

Model specification

To obtain estimates of the impact of KIPP admissions for the subset of KIPP elementary schools included in the lottery-based analysis sample, we use the following model:

$$\text{(D.1)}\ y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * T_i + \gamma * X_i + \varepsilon_i$$

where *i* and *k* index students and schools, respectively, and *y* is the student-level outcome of interest. *SCHOOL* is a set of binary variables indicating the school that the student applied to and thus the lottery in which the student participated, *T* is a binary treatment status variable indicating whether the student was offered admission to the school via the lottery, and *X* is a set of demographic and other controls. The βs represent site/lottery fixed effects, which capture differences in outcomes across sites that are not related to KIPP school attendance itself. These effects may capture variation across schools in the characteristics of KIPP applicants and/or the characteristics and performance of non-KIPP schools attended by control students. Including fixed effects in the model (as opposed to random effects) implies that KIPP schools were selected purposefully for the lottery-based analysis and that the results cannot be generalized beyond the study schools. The parameter δ represents the average impact of winning a KIPP elementary school lottery; this is an intent-to-treat (ITT) estimate.

Our analysis includes the following student covariates (in X):

- gender,
- student age in years,
- race/ethnicity,

- language spoken at home,
- whether there is only one adult in the household,
- family income,
- mother's education,
- whether the student has access to a computer with internet at home, and
- the number of children's books in the home.

Weighting

The impact model incorporates sample weights to account for the fact that not all students in the lottery have the same probability of being offered admission to the KIPP school (that is, being selected into the treatment group). Some students have a higher probability of being offered admission, either based on their inclusion in a particular stratum defined by a student characteristic or because they have a sibling in the lottery. If no sample weights were used and if these student characteristics were not otherwise accounted for in the impact model, then the characteristics of students in the treatment group and control group would differ on average, potentially leading to a bias in the impact estimate. For example, since several KIPP schools use sibling preference rules in their lotteries, students with siblings will tend to be over-represented in the treatment group and students without siblings will be over-represented in the control group. If having siblings affects student performance directly or is correlated with some other student or family characteristic that is not accounted for, this could bias the impact estimate.

The creation of the sample weights is based on the procedure used in Gleason et al. (2010). In the simple case, where all students interested in attending a particular KIPP school enter the lottery and no preferences are given for siblings or other characteristics, the sample weight for a given student is based upon the probability that he or she ended up in a particular experimental group (that is, treatment or control group). This probability is used in the calculation of each student's *base weight*. In particular, the base weight assigned to treatment group members is set to the inverse of the probability of being selected into the treatment group. The base weight for control group members is set to the inverse of the probability of being selected into the control group. We then normalize this weight to account for the fact that the sample will be representative of the set of all consenting lottery participants at that site. We set this normalization factor such that the weights of each experimental group sum to one-half of the total sample size within the site. Thus, the sum of all students' weights within a site will be equal to the overall sample size in that site (that is, the number of consenting lottery participants), with the sum of weights among treatments equal to that among controls.

In sites with sibling preference rules, the basic approach to calculating sample weights is the same as in the simple case above.[9] The difference, however, is in the calculation of the probability of admission. No longer can we simply use the number of students offered admission divided by the number of lottery participants. The exact probabilities of admission depend on the

[9] An example of sibling preference rules occurs when a school enters two siblings separately in an admissions lottery. If one of the two siblings is drawn as a lottery winner and offered admission to the school, the other sibling is pulled from the lottery pool and also offered admission.

number of sets of siblings who participate in the lottery at the school as well as the number of students within each sibling set. With sibling preference rules, each sibling in the lottery has a higher probability of admissions than non-siblings, so the probabilities are adjusted to account for the number of siblings in each affected lottery.

Imputation of baseline characteristics

If there were missing values for the model's covariates, we imputed these values based on other baseline information we collected from the student so that he or she could be included in the sample and contribute to our impact estimates. Our imputation procedure, known as multiple imputation by chained equations, uses non-missing values of baseline covariates to estimate plausible values of baseline characteristics for observations with missing baseline data. In particular, this method first generates multiple datasets with estimated ("imputed") values for missing baseline characteristics. A separate impact estimate is then calculated with each of the imputed datasets. Finally, these impact estimates are combined using procedures described in Rubin (1987) that account for the variability of estimates calculated using the different imputed datasets. The standard error of each combined impact estimate is adjusted to reflect this variability. The imputation procedure and impact estimation using imputed data are conducted using standard commands in Stata and 20 imputations are used. Imputation is conducted separately by treatment group, and all baseline characteristics included as covariates in the impact model are included in the imputation model. Finally, no outcome measures are imputed, only baseline characteristics.

While we use these imputed baseline covariates in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variable were simply treated as being missing from the sample.

Additional analyses

Complier Average Causal Effect (CACE) estimate of the impact of KIPP attendance

For the subset of KIPP elementary schools in which randomized lotteries created viable treatment and control groups, we present two sets of impact estimates: (1) intent-to-treat (ITT) estimates that rely on treatment status as defined by the random lotteries to estimate the impact of being offered admission to a KIPP elementary school and (2) Complier Average Causal Effect (CACE) estimates that represent the impact of attending a KIPP elementary school.[10]

Because families and students choose whether or not to attend KIPP after winning an admissions lottery, and not all lottery winners ultimately attend KIPP, we cannot simply compare outcomes of KIPP attendees and non-attendees to get an unbiased estimate of attending a KIPP elementary school. To generate CACE estimates of the impact of attending a KIPP elementary school, we use the outcome of the lottery for each student as an instrumental variable for KIPP attendance, where attendance is defined as ever having attended a KIPP elementary school. In other words, to obtain CACE estimates we calculate the difference between the outcomes of treatment and control students, adjusting it to reflect the difference between the proportion of

[10] CACE estimates are sometimes referred to as treatment-on-the-treated, or TOT, estimates.

treatment and control students who enroll at KIPP.[11] The CACE model uses the same covariates and weights as the ITT model, and a similar imputation strategy is used, but with a single imputation rather than multiple imputation.

Specifically, we used two-stage least squares to first estimate the effect of winning an admissions lottery on KIPP attendance (Equation D2a), and in the second stage estimated the impact of KIPP attendance on outcomes (Equation D2b). In effect, the CACE approach adjusts the ITT results to account for whether students actually attended a KIPP school.

(D2a) $$attendKIPP_i = \eta + \lambda * T_i + \sum_{k=1}^{K} \theta_k * SCHOOL_{i,k} * T_i + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \gamma * X_i + \upsilon_i$$

(D2b) $$y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * \widehat{attendKIPP_i} + \gamma * X_i + \varepsilon_i$$

As shown in Table D.4 below, the CACE estimates are generally larger than the ITT estimates, as expected. These estimates indicate that, among students whom the admissions lottery caused to attend a KIPP elementary school, KIPP had impacts of 0.41 standard deviations in Calculation, 0.23 standard deviations in Letter and Word Identification, and 0.29 standard deviations in Passage Comprehension, all of which are statistically significant. The estimated impact on Applied Problems is small and not statistically significant, as is the ITT estimate for that outcome. The CACE estimate in Calculation is approximately equivalent to a student moving from the 58th percentile to the 73rd percentile; the Letter and Word Identification estimate is equivalent to moving from the 78th percentile to the 86th percentile; and the Passage Comprehension estimate is equivalent to moving from the 48th percentile to the 60th percentile. (In each of these examples, the starting percentile corresponds with the control group mean score for that outcome.)

Table D.4. ITT and CACE estimates of KIPP elementary school impact

Outcome (z-score)	ITT	Std. Error	CACE	Std. Error	$F_{instruments}$	Number of instruments	N_t	N_c
WJ Calculation	0.28**	0.11	0.41**	0.15	66.62	5	176	195
WJ Applied Problems	0.07	0.05	0.06	0.07	62.56	8	282	370
WJ Letter-Word Identification	0.25**	0.07	0.34**	0.10	61.38	8	281	370
WJ Passage Comprehension	0.22**	0.07	0.29**	0.09	64.70	8	280	368

Notes: Outcomes are measured on Woodcock-Johnson III, administered in the spring of the third follow-up year. Impacts in the first column of this table are intent-to-treat (ITT), based on regression models that pool all lottery elementary schools and that control for baseline covariates. Impacts in the third column are complier average causal effect (CACE, sometimes referred to as treatment-on-treated or TOT estimate) for each outcome. The same covariate set is used for ITT and CACE models. Standard errors are to the right of each impact estimate. The F-statistic and number of instruments are provided for each CACE estimate to document the instruments' predictive strength. Instruments used for the calculation outcome reflect that only five schools were in that sample. CACE estimates are calculated using a dataset in which missing values of baseline characteristics are imputed using a single imputation rather than multiple imputation.

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

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

[11] Control students may end up enrolling at KIPP if they are offered admission after the October cut-off date for assignment to the treatment group (for example, during the second semester), if they apply and are offered admission for the following school year, or in rare cases when they are offered admission out of order off the waitlist.

Sensitivity to weighting and imputation method

The weighting and imputation approaches used in our main elementary school impact estimates are described earlier in this appendix. This section presents evidence on the sensitivity of impact estimates to alternative weighting and imputation approaches. As described above, the normalization of our main sample weights causes each school to contribute to the overall impact estimate in proportion to its sample size. We also calculated impacts using a different normalization approach in which each school contributes equally to the overall impact estimates.[12] As shown in the third column of Table D.5, impacts on Calculation calculated using this approach is substantially smaller than the benchmark estimate and is not statistically significant, in contrast to the benchmark estimate. This difference is driven in large part by a single school with a very small sample size; when this school is omitted from the sample the main impact estimate and alternative estimate differ by less than 0.01 standard deviation and are each statistically significant. Impacts on other test scores estimated using the alternative weighting approach are similar to the main estimates in magnitude and statistical significance.

In order to test the sensitivity of impact estimates to our approach for imputing missing baseline information, we also estimated impacts using mean-imputed baseline characteristics. In this alternative approach, missing baseline characteristics were imputed as the mean among students not missing that characteristic within the same school and treatment group. As shown in the fifth column of Table D.5, there are no substantive differences in the magnitude or statistical significance of these estimates, relative to the main impacts.

Table D.5. ITT Estimates from alternative analysis approaches

Outcome (z-score)	Main ITT Estimate	Std. Error	Schools weighted equally	Std. Error	Mean imputed covariates	Std. Error	N_t	N_c
WJ Calculation	0.28**	0.11	0.18	0.10	0.29**	0.10	176	195
WJ Applied Problems	0.07	0.05	0.04	0.05	0.04	0.05	282	370
WJ Letter-Word Identification	0.25**	0.07	0.22*	0.08	0.22**	0.07	281	370
WJ Passage Comprehension	0.22**	0.07	0.21**	0.07	0.18**	0.07	280	368

Notes: Outcomes are measured on Woodcock-Johnson III, administered in the spring of the third follow-up year. Impacts in the first column of this table are intent-to-treat (ITT), based on regression models that pool all lottery elementary schools and that control for baseline covariates. Impacts in the third column are calculated using a model that weights schools equally, regardless of enrollment size. Impacts in the fifth column use mean imputed data rather than the multiple imputation approach used in the main impact estimates. The same set of baseline characteristics is controlled for in all models. Estimates calculated with schools weighted equally use a single imputed dataset rather than multiple imputed datasets. Standard errors are to the right of each impact estimate.

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

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

[12] This alternative weighting approach does not change the method by which we accounted for differential probabilities of admission within a school for siblings versus non-siblings. Rather, it only affects the weight given to each school when combining each school's impact into an overall cross-site impact.

APPENDIX E

DETAILED ANALYTIC METHODS: MIDDLE SCHOOL (LOTTERY-BASED ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP middle schools. First we present information on the sample and the baseline equivalence of students who won an admission lottery (the treatment group) and those who did not win (the control group). Next we discuss the data and analytic methods used for the analysis. We conclude the appendix by discussing results of sensitivity analyses.

Detail on sample

Of the 59 KIPP middle schools open in Spring 2011, 16 (27 percent) were sufficiently oversubscribed to include in the lottery-based analysis.[13] Middle school admissions lotteries are held primarily at the 5th and 6th grade level, so the analysis sample is drawn from these two entry grades. Among students in the study sample schools and grades, 996 were admitted based on a lottery result (Table E.1).[14] The original lottery sample, as defined above, is larger than required to meet the study's targeted level of statistical power. There was a substantial imbalance in the size of the treatment and control groups at several schools in the baseline sample. In other words, a larger number of students participating in the lotteries at these schools were offered admission and included in the treatment group than the number who were not offered admission and included in the control group, or vice versa. To conserve resources while simultaneously maximizing our ability to detect impacts, a subsample of 857 students at these schools was randomly selected to comprise the baseline sample for inclusion in the study's data collection (follow-up surveys and state test records).[15]

Table E.1. Middle school student sample sizes (lottery-based analysis)

	Treatment	Control	Overall
Original lottery sample	530	466	996
Baseline sample	436	421	857
Analytic sample (Year 1)	314	295	609
Analytic sample (Year 2)	291	274	565
Analytic sample (Year 3)	234	225	459

In each of the three outcome years for which state test data were collected, students with test scores available were included in the analysis sample. In total, 609 students were included in the

[13] One oversubscribed school was located in a jurisdiction that did not provide any state test data and is therefore not include in the achievement analysis sample. Of the remaining 43 schools, 19 also conducted lotteries for admission but either exhausted their waitlists or did not provide a sufficiently large treatment or control group for the analysis.

[14] Students who apply to oversubscribed schools may be guaranteed admission and thus not be eligible for the study. For example, applicants may be admitted to the school outside of the typical lottery process if they have a sibling already attending the school or if a predetermined number of seats are reserved for district residents and these reserved seats are not oversubscribed.

[15] The sample of students used in lottery-based estimates of KIPP middle schools' impact on student achievement differs slightly from the sample used when estimating impacts on survey measures because one jurisdiction categorically declined to provide test score data on students whose parent had signed a consent form electronically rather than manually. These students were considered to have been randomly removed from the sample after verifying that there was not a substantial difference in the rate at which lottery winners and losers were excluded on this basis.

year one analysis sample, 565 students were included in the second year's sample and 459 students were included in year three. The main lottery-based impact analysis measures the impact of admission to—rather than attendance at—a KIPP school (in other words, it is an intent-to-treat [ITT] estimate), although admission and attendance are closely related. Among lottery participants, 72 percent of treatment group students and 5 percent of control group students attended a KIPP middle school.

Properly executed randomization should ensure that there are no differences (observed or unobserved) between the treatment and control groups. In principle, we can test whether or not this is true by examining the baseline characteristics of the treatment and control groups. The baseline characteristics were assessed using data from administrative records and from a baseline survey administered to the parents of students participating in KIPP lotteries close to the time of random assignment.[16] Information from the baseline survey was supplemented with information from the follow-up parent survey where there was missing data from the baseline survey on characteristics not expected to vary systematically with lottery outcome over time (for example, gender, race/ethnicity, and date of birth).

We compared the baseline characteristics of the baseline sample of students in the treatment and control groups and found only one difference that was statistically significant at the 5 percent level among the 28 characteristics we examined, which is what we would have expected due to chance alone (Table E.2). Since the lottery-based impact analysis is based on the sample of students with test outcomes in each of three years following the lottery, it is also important to examine whether there are any differences in the baseline characteristics of the treatment and control students for whom we have outcome data in each sample (that is, in the analytic samples).

We examined baseline equivalence separately for each of these analytic samples, and the results provide evidence that the treatment and control groups were similar at baseline (Tables E.3-E.5). In each of the analytic samples—academic achievement in 2012, 2013, and 2014—only one out of 28 measured baseline characteristics shows a statistically significant difference between groups, which is again what would be expected due to chance alone. Given these findings, we are confident that the admissions lotteries were conducted correctly and that the treatment and control groups in each analysis sample are similar in terms of their background characteristics, motivation, and prior educational experiences, aside from the outcome of the lottery itself.

[16] Overall, about 72 percent of parents of students in the middle school analysis sample completed the baseline survey. Among those who completed the baseline survey, about 35 percent completed it prior to the time of the lottery and the remaining 65 percent completed it after the lottery. For the group that completed the baseline survey after the lottery, there is some risk that the lottery outcome could have influenced their responses. However, the test scores are based on administrative data and cover the year (or two years) prior to middle school entry.

Table E.2. Baseline equivalence for middle school baseline sample

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std. Error	N_t	N_c
Baseline Reading (z-score)	-0.21	-0.28	0.07	0.07	282	292
Baseline Math (z-score)	-0.12	-0.23	0.10	0.08	280	291
Pre-Baseline Reading (z-score)	-0.21	-0.26	0.05	0.08	270	271
Pre-Baseline Math (z-score)	-0.18	-0.21	0.03	0.08	270	273
Female	0.52	0.52	0.00	0.04	372	363
Age in years	10.63	10.64	-0.01	0.05	300	289
White, non-hispanic	0.02	0.04	-0.02	0.01	382	378
Hispanic or latino	0.53	0.51	0.02	0.03	382	378
Black, non-hispanic	0.43	0.42	0.01	0.03	382	378
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.03	0.00	0.01	382	378
English is main language spoken at home	0.51	0.55	-0.04	0.03	347	331
Another language is main language spoken at home	0.25	0.24	0.01	0.03	347	331
English and another lang. spoken equally at home	0.24	0.21	0.03	0.03	347	331
One adult in household	0.32	0.30	0.01	0.04	340	326
Free/Reduced Price Lunch Eligible	0.84	0.86	-0.02	0.03	358	353
Has an Individualized Education Program (IEP)	0.10	0.10	0.00	0.02	355	355
Family income less than 15K	0.25	0.20	0.05	0.04	293	270
Family income between 15K and 25K	0.25	0.27	-0.02	0.04	293	270
Family income between 25K and 35K	0.17	0.27	-0.10**	0.04	293	270
Family income between 35K and 50K	0.20	0.17	0.03	0.04	293	270
Family income 50K or greater	0.14	0.10	0.04	0.03	293	270
Mother's education: less than HS	0.21	0.21	0.00	0.03	342	325
Mother's education: HS or GED	0.24	0.26	-0.02	0.04	342	325
Mother's education: some college	0.28	0.29	-0.01	0.04	342	325
Mother's education: college	0.28	0.25	0.03	0.03	342	325
Student has access to computer with internet at home	0.78	0.84	-0.06	0.03	307	286
Parent helps student with homework 5 days per week or more	0.73	0.72	0.01	0.04	304	282
Parent discussed college with student over 2x during pre-baseline school year	0.87	0.85	0.02	0.03	304	281

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample randomly assigned by the admissions lottery, then randomly subsampled for data collection. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

Table E.3. Baseline equivalence for middle school analysis sample, year 1

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std Error	Nt	Nc
Baseline Reading (z-score)	-0.16	-0.24	0.08	0.07	268	264
Baseline Math (z-score)	-0.09	-0.18	0.09	0.08	266	264
Pre-Baseline Reading (z-score)	-0.17	-0.21	0.03	0.08	256	243
Pre-Baseline Math (z-score)	-0.14	-0.16	0.02	0.08	256	244
Female	0.52	0.52	0.00	0.04	311	293
Age in years	10.63	10.64	-0.01	0.05	248	224
White, non-hispanic	0.03	0.04	-0.01	0.01	314	295
Hispanic or latino	0.56	0.54	0.02	0.03	314	295
Black, non-hispanic	0.38	0.39	-0.01	0.03	314	295
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.03	0.00	0.01	314	295
English is main language spoken at home	0.48	0.52	-0.05	0.03	283	255
Another language is main language spoken at home	0.27	0.25	0.02	0.04	283	255
English and another lang. spoken equally at home	0.25	0.22	0.03	0.04	283	255
One adult in household	0.31	0.27	0.04	0.04	277	250
Free/Reduced Price Lunch Eligible	0.87	0.87	-0.01	0.03	302	284
Has an Individualized Education Program (IEP)	0.09	0.08	0.00	0.03	305	286
Family income less than 15K	0.26	0.21	0.04	0.04	243	210
Family income between 15K and 25K	0.24	0.28	-0.04	0.05	243	210
Family income between 25K and 35K	0.19	0.27	-0.09*	0.04	243	210
Family income between 35K and 50K	0.21	0.14	0.07	0.04	243	210
Family income 50K or greater	0.11	0.10	0.01	0.03	243	210
Mother's education: less than HS	0.22	0.24	-0.02	0.04	279	250
Mother's education: HS or GED	0.24	0.26	-0.02	0.04	279	250
Mother's education: some college	0.29	0.27	0.02	0.04	279	250
Mother's education: college	0.25	0.23	0.02	0.04	279	250
Student has access to computer with internet at home	0.77	0.83	-0.07	0.04	253	222
Parent helps student with homework 5 days per week or more	0.71	0.71	0.01	0.04	250	218
Parent discussed college with student over 2x during pre-baseline school year	0.86	0.83	0.03	0.04	250	218

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data and non-missing baseline data,. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

Table E.4. Baseline equivalence for middle school analysis sample, year 2

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std Error	Nt	Nc
Baseline Reading (z-score)	-0.13	-0.29	0.16*	0.07	246	243
Baseline Math (z-score)	-0.03	-0.17	0.14	0.08	245	242
Pre-Baseline Reading (z-score)	-0.14	-0.24	0.10	0.08	237	224
Pre-Baseline Math (z-score)	-0.08	-0.15	0.07	0.08	237	226
Female	0.53	0.51	0.02	0.04	288	274
Age in years	10.66	10.65	0.02	0.05	231	211
White, non-hispanic	0.03	0.04	-0.01	0.01	291	274
Hispanic or latino	0.56	0.54	0.02	0.03	291	274
Black, non-hispanic	0.39	0.39	0.00	0.03	291	274
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.02	0.03	-0.01	0.01	291	274
English is main language spoken at home	0.47	0.51	-0.04	0.04	262	238
Another language is main language spoken at home	0.27	0.26	0.01	0.04	262	238
English and another lang. spoken equally at home	0.25	0.22	0.03	0.04	262	238
One adult in household	0.32	0.26	0.06	0.04	257	233
Free/Reduced Price Lunch Eligible	0.87	0.88	-0.01	0.03	279	266
Has an Individualized Education Program (IEP)	0.09	0.10	-0.02	0.03	281	268
Family income less than 15K	0.27	0.21	0.06	0.04	225	198
Family income between 15K and 25K	0.23	0.29	-0.07	0.05	225	198
Family income between 25K and 35K	0.20	0.27	-0.07	0.04	225	198
Family income between 35K and 50K	0.21	0.14	0.07	0.04	225	198
Family income 50K or greater	0.10	0.09	0.01	0.03	225	198
Mother's education: less than HS	0.21	0.24	-0.04	0.04	259	233
Mother's education: HS or GED	0.26	0.25	0.01	0.04	259	233
Mother's education: some college	0.28	0.26	0.02	0.04	259	233
Mother's education: college	0.25	0.24	0.01	0.04	259	233
Student has access to computer with internet at home	0.76	0.82	-0.06	0.04	234	209
Parent helps student with homework 5 days per week or more	0.72	0.70	0.01	0.04	231	206
Parent discussed college with student over 2x during pre-baseline school year	0.85	0.84	0.01	0.04	231	206

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data and non-missing baseline data. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

Table E.5. Baseline equivalence for middle school analysis sample, year 3

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std Error	N_t	N_c
Baseline Reading (z-score)	-0.11	-0.26	0.16	0.08	196	195
Baseline Math (z-score)	-0.04	-0.14	0.10	0.09	195	194
Pre-Baseline Reading (z-score)	-0.18	-0.20	0.02	0.09	188	180
Pre-Baseline Math (z-score)	-0.12	-0.09	-0.03	0.10	188	182
Female	0.53	0.51	0.02	0.05	231	224
Age in years	10.63	10.64	-0.01	0.06	182	173
White, non-hispanic	0.03	0.05	-0.02	0.02	234	225
Hispanic or latino	0.51	0.47	0.04	0.04	234	225
Black, non-hispanic	0.44	0.44	-0.01	0.04	234	225
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.02	0.04	-0.02	0.02	234	225
English is main language spoken at home	0.50	0.56	-0.06	0.04	209	196
Another language is main language spoken at home	0.25	0.26	-0.02	0.04	209	196
English and another lang. spoken equally at home	0.26	0.18	0.07	0.04	209	196
One adult in household	0.35	0.27	0.08	0.05	206	191
Free/Reduced Price Lunch Eligible	0.86	0.86	-0.01	0.03	222	218
Has an Individualized Education Program (IEP)	0.09	0.11	-0.02	0.03	224	219
Family income less than 15K	0.28	0.20	0.08	0.05	177	164
Family income between 15K and 25K	0.24	0.27	-0.03	0.05	177	164
Family income between 25K and 35K	0.18	0.30	-0.12*	0.05	177	164
Family income between 35K and 50K	0.20	0.14	0.06	0.04	177	164
Family income 50K or greater	0.11	0.10	0.01	0.03	177	164
Mother's education: less than HS	0.20	0.23	-0.02	0.04	208	191
Mother's education: HS or GED	0.24	0.27	-0.03	0.05	208	191
Mother's education: some college	0.31	0.28	0.02	0.05	208	191
Mother's education: college	0.25	0.23	0.03	0.04	208	191
Student has access to computer with internet at home	0.78	0.82	-0.05	0.04	186	171
Parent helps student with homework 5 days per week or more	0.72	0.68	0.04	0.05	183	168
Parent discussed college with student over 2x during pre-baseline school year	0.84	0.84	0.01	0.04	183	168

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data and non-missing baseline data. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

Detail on achievement measures

Academic achievement was measured using scores on state-wide assessments drawn from administrative records collected from states and districts. Students' scores on the state tests were standardized (that is, converted into z-scores) using state-wide means and standard deviations. Thus, values reflect students' performance relative to all tested students in a given state (within grade and subject): positive values indicate that sample students outperformed the average student state-wide and negative values indicate that sample students performed below the state average. This standardization to a common scale allows us to combine outcomes for students in different states. For most students and jurisdictions we were able to collect test score outcomes in the spring of 2012, 2013, and 2014, corresponding to the first, second, and third year after random assignment.[17]

Detail on analytic methods

Model specification

To obtain estimates of the impact of KIPP admissions for the subset of KIPP middle schools included in the lottery-based analysis sample, we use the following model:

$$\text{(E.1)}\ y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * T_i + \gamma * X_i + \varepsilon_i$$

where *i* and *k* index students and schools, respectively, and *y* is the student-level outcome of interest. *SCHOOL* is a set of binary variables indicating the school that the student applied to and thus the lottery in which the student participated, *T* is a binary treatment status variable indicating whether the student was offered admission to the school via the lottery, and *X* is a set of demographic and other controls, which are listed in Table E.6 below[18]. The βs represent site/lottery fixed effects, which capture differences in outcomes across sites that are not related to KIPP school attendance itself. These effects may capture variation across schools in the characteristics of KIPP applicants and/or the characteristics and performance of non-KIPP schools attended by control students. Including fixed effects in the model (as opposed to random effects) implies that that KIPP schools were selected purposefully for the lottery-based analysis and that the results cannot be generalized beyond the study schools. The parameter δ represents the average impact of winning a KIPP middle school lottery; this is an intent-to-treat (ITT) estimate.

[17] One jurisdiction did not provide test score data for the 2013-14 school year, so the school from that jurisdiction is not included in the last year of test score analysis. Two jurisdictions did not provide their statewide means and standard deviations for the 2014 test administration. The impact estimates presented in this memo for schools in these states from 2014 use z-scores based on 2013 means and standard deviations.

[18] In addition to demographics and baseline test scores, these covariates include school-specific indicators for a handful of students who are a grade ahead of their admissions-lottery peers *before and after* the admissions lottery. These indicators control for systematic differences between the test scores in different grades. Such students are only included in the analysis sample if they have a counterpart in the opposite treatment group at their school. An additional indicator variable for students who took an end-of-course math exam in year three similarly controls for systematic differences between end-of-course and end-of-grade subject tests. As with students ahead of grade, end-of-course exam takers are only included if their school includes such students in the treatment and control group.

Table E.6. List of covariates included in impact model

Math baseline test score from 1 year prior to lottery
Math baseline test score from 2 years prior to lottery
Reading baseline test score from 1 year prior to lottery
Reading baseline test score from 2 years prior to lottery
Gender indicator variable
Age in years at baseline
Set of race/ethnicity indicator variables
Set of indicators for home language: English, Non-English, or bilingual
Indicator for single-parent household
Free or reduced price lunch status indicator variable
Special education status indicator variable
Set of family income indicator variables
Set of mother's education indicator variables
Indicator variable for access to a computer with internet at home
Indicator variable for parent helping with homework five days per week or more
Indicator variable for parent discussing college with student more than two times in the baseline school year
Set of math and reading imputation dummies indicating whether math and reading baseline test scores are imputed
Set of dummy variables indicating whether demographic variables were imputed
School-specific dummy variables indicating students who are ahead of grade before and after the admission lottery, based on the lottery entry grade
Dummy variable for lottery entry grade
Dummy variable indicating that a student took an end of course exam (math, year three only)

Note: Baseline test scores and other covariates were imputed when missing.

Weighting

The impact model incorporates sample weights to account for the fact that not all students in the lottery have the same probability of being offered admission to the KIPP school (that is, being selected into the treatment group). Some students have a higher probability of being offered admission, either based on their inclusion in a particular stratum defined by a student characteristic or because they have a sibling in the lottery. If no sample weights were used and if these student characteristics were not otherwise accounted for in the impact model, then the characteristics of students in the treatment group and control group would differ on average, potentially leading to a bias in the impact estimate. For example, since several KIPP schools use sibling preference rules in their lotteries, students with siblings will tend to be over-represented in the treatment group and students without siblings will be over-represented in the control group. If having siblings affects student performance directly or is correlated with some other student or family characteristic that is not accounted for, this could bias the impact estimate.

The creation of the sample weights is based on the procedure used in Gleason et al. (2010). In the simple case, where all students interested in attending a particular KIPP school enter the lottery and no preferences are given for siblings or other characteristics, the sample weight for a given student is based upon the probability that he or she ended up in a particular experimental group (that is, treatment or control group). This probability is used in the calculation of each student's *base weight*. In particular, the base weight assigned to treatment group members is set to the inverse of the probability of being selected into the treatment group. The base weight for control group members is set to the inverse of the probability of being selected into the control group. We then normalize this weight to account for the fact that the sample will be representative of the set of all consenting lottery participants at that site. We set this

normalization factor such that the weights of each experimental group sum to one-half of the total sample size within the site. Thus, the sum of all students' weights within a site will be equal to the overall sample size in that site (that is, the number of consenting lottery participants), with the sum of weights among treatments equal to that among controls.

In sites with sibling preference rules, the basic approach to calculating sample weights is the same as in the simple case above.[19] The difference, however, is in the calculation of the probability of admission. No longer can we simply use the number of students offered admission divided by the number of lottery participants. The exact probabilities of admission depend on the number of sets of siblings who participate in the lottery at the school as well as the number of students within each sibling set. With sibling preference rules, each sibling in the lottery has a higher probability of admissions than non-siblings, so the probabilities are adjusted to account for the number of siblings in each affected lottery.

Imputation of Baseline Characteristics

If there were missing values for the model's covariates, we imputed these values based on other baseline information we collected from the student so that he or she could be included in the sample and contribute to our impact estimates. Our imputation procedure, known as multiple imputation by chained equations, uses non-missing values of baseline covariates to estimate plausible values of baseline characteristics for observations with missing baseline data. In particular, this method first generates multiple datasets with estimated ("imputed") values for missing baseline characteristics. A separate impact estimate is then calculated with each of the imputed datasets. Finally, these impact estimates are combined using procedures described in Rubin (1987) that account for the variability of estimates calculated using the different imputed datasets. The standard error of each combined impact estimate is adjusted to reflect this variability. The imputation procedure and impact estimation using imputed data are conducted using standard commands in Stata and 20 imputations are used. Imputation is conducted separately by treatment group, and all baseline characteristics included as covariates in the impact model are included in the imputation model. Finally, no outcome measures are imputed, only baseline characteristics.

While we use these imputed baseline covariates in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variable were simply treated as being missing from the sample.

Grade Repeaters and Skippers

In our main lottery-based analysis of state test scores, we excluded students who repeated or skipped a grade after the admission lottery because they do not have the same grade progression as their peers and therefore do not have the same pretest-posttest relationship.[20] This strategy is

[19] An example of sibling preference rules occurs when a school enters two siblings separately in an admissions lottery. If one of the two siblings is drawn as a lottery winner and offered admission to the school, the other sibling is pulled from the lottery pool and also offered admission.

[20] A very small number of students were ahead of the grade progression indicated by their lottery entry grade *both before and after* the lottery. We retained these students in the sample and included their actual (not imputed) test

in contrast to the matching approach that "freezes" grade repeaters in the test score distribution. As discussed later in this appendix, we tested the sensitivity of our impacts on state test scores to this approach by calculating impacts using the alternate approach. As shown below, when we retain grade repeaters in the sample but impute their outcome score, the impact estimates do not change substantively.

Additional analyses

Complier Average Causal Effect (CACE) estimate of the impact of KIPP attendance

For the subset of KIPP middle schools in which randomized lotteries created viable treatment and control groups, we present two sets of impact estimates: (1) intent-to-treat (ITT) estimates that rely on treatment status as defined by the random lotteries to estimate the impact of being offered admission to a KIPP middle school and (2) Complier Average Causal Effect (CACE) estimates that represent the impact of attending a KIPP middle school.[21]

Because families and students choose whether or not to attend KIPP after winning an admissions lottery, and not all lottery winners ultimately attend KIPP, we cannot simply compare outcomes of KIPP attendees and non-attendees to get an unbiased estimate of attending a KIPP middle school. To generate CACE estimates of the impact of attending a KIPP middle school, we use the outcome of the lottery for each student as an instrumental variable for KIPP attendance, where attendance is defined as ever having attended a KIPP middle school. In other words, to obtain CACE estimates we calculate the difference between the outcomes of treatment and control students, adjusting it to reflect the difference between the proportion of treatment and control students who enroll at KIPP.[22] The CACE model uses the same covariates and weights as the ITT model, and a similar imputation strategy is used, but with a single imputation rather than multiple imputation. Specifically, we used two-stage least squares to first estimate the effect of winning an admissions lottery on KIPP attendance (Equation E2a), and in the second stage estimated the impact of KIPP attendance on outcomes (Equation E2b). In effect, the CACE approach adjusts the ITT results to account for whether students actually attended a KIPP school.

(E2a) $$attendKIPP_i = \eta + \lambda * T_i + \sum_{k=1}^{K} \theta_k * SCHOOL_{i,k} * T_i + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \gamma * X_i + \upsilon_i$$

(E2b) $$y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * \widehat{attendKIPP_i} + \gamma * X_i + \varepsilon_i$$

As shown in Table E.7 below, the CACE estimates are larger than the ITT estimates, as expected. These indicate that in the second year, for example, KIPP had impacts of 0.32 standard

scores in the impact model, including an indicator variable to control for any systematic differences in their test scores, relative to peers in other grades. We only retained such students in the sample if they had a counterpart in the opposite treatment group at the same school. That is, if at a given school all students in this situation were in the treatment group and none were in the control group, then they were not included in the sample.

[21] CACE estimates are sometimes referred to as treatment-on-the-treated, or TOT, estimates.

[22] Control students may end up enrolling at KIPP if they are offered admission after the October cut-off date for assignment to the treatment group (for example, during the second semester), if they apply and are offered admission for the following school year, or in rare cases when they are offered admission out of order off the waitlist. For each outcome year, attendance is defined for each student as having ever attended a KIPP school between the lottery and that school year.

deviations in Math and 0.27 standard deviations in reading among students whom the admissions lottery caused to attend a KIPP middle school. The CACE estimate in math in year two is approximately equivalent to a student moving from the 40th percentile to the 53rd percentile; the CACE estimate in reading in the same year is equivalent to a student moving from the 37th percentile to the 47th percentile. (In each of these examples, the starting percentile corresponds with the control group mean score for that outcome.)

Table E.7. ITT and CACE estimates of KIPP middle school impact

Outcome (z-score)	ITT	Std. Error	CACE	Std. Error	$F_{instruments}$	Number of instruments	N_t	N_c
Math								
Year 1	0.10*	0.05	0.12	0.06	73.3	16	313	294
Year 2	0.24**	0.06	0.32**	0.08	64.7	16	287	268
Year 3	0.18**	0.07	0.25**	0.09	56.6	14	233	222
Reading								
Year 1	0.03	0.05	0.05	0.06	75.1	16	314	294
Year 2	0.18**	0.05	0.27**	0.07	66.5	16	291	272
Year 3	0.14*	0.06	0.16	0.08	56.9	14	234	224

Source: State and district administrative records data, study-administered survey data (baseline characteristics)

Note: Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Impact estimates in the third column measure the impact of an offer of admission to a KIPP middle school (the ITT estimate) and are based on regression models that pool all lottery schools and control for baseline covariates. Impact estimates in the fifth column are complier average causal effect (CACE, sometimes referred to as treatment-on-treated or TOT estimate) for each outcome. The same covariate set is used for ITT and CACE models. Standard errors are to the right of each impact estimate. The F-statistic and number of instruments are provided for each CACE estimate to document the instruments' predictive strength. The number of instruments for year 3 estimates reflects that two grades of one school were excluded from this outcome sample because they did not provide test scores for that year. CACE estimates are calculated using a dataset in which missing values of baseline characteristics are imputed using a single imputation rather than multiple imputation.

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

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

Sensitivity to weighting

The weighting and imputation approaches used in our main middle school impact estimates are described earlier in this appendix. This section presents evidence on the sensitivity of impact estimates to alternative weighting and imputation approaches. As described above, the normalization of our main sample weights causes each school to contribute to the overall impact estimate in proportion to its sample size. We also calculated impacts using a different normalization approach in which each school contributes equally to the overall impact estimates.[23] As shown in the third column of Table E.8, impacts calculated using the alternate approach generally have a magnitude similar to the benchmark model, with the exception of the reading impact in year three. In the third year, sites with smaller sample sizes tended to have more lower (or negative impacts); because the alternative approach gives small sites more weight

[23] This alternative weighting approach does not change the method by which we accounted for differential probabilities of admission within a school for siblings versus non-siblings. Rather, it only affects the weight given to each school when combining each school's impact into an overall cross-site impact.

than the main impact estimates do, these small sites contributed to a lower overall impact estimate.

Table E.8. ITT estimates using alternative methods

Outcome (z-score)	Main impact	Std. error	Impact weighting schools equally	Std. Error	N_t	N_c	Impact including repeaters, skippers	Std. Error	N_t	N_c
Math										
Year 1	0.10	0.05	0.06	0.05	313	294	0.10*	0.04	315	300
Year 2	0.24**	0.06	0.19**	0.06	287	268	0.25**	0.06	303	278
Year 3	0.18*	0.07	0.15*	0.06	233	222	0.20**	0.06	251	236
Reading										
Year 1	0.03	0.05	0.01	0.05	314	294	0.02	0.04	316	300
Year 2	0.18**	0.05	0.17**	0.06	291	272	0.18**	0.05	306	278
Year 3	0.14*	0.06	0.03	0.06	234	224	0.13*	0.05	252	237

Source: State and district administrative records data, study-administered survey data (baseline characteristics)

Note: Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Estimates measure the impact of an offer of admission to a KIPP middle school (the ITT estimate) and are based on regression models that pool all lottery schools and control for baseline covariates. Impact estimates in the first column are those presented in the main report body, which combine school-level estimates weighting each school's estimate by its sample size. Impacts in the third column weight schools equally, regardless of their sample size. Impacts in the seventh column include students who were excluded from other impact models based on having repeated or skipped grades only following the lottery. The alternative models use a single imputed dataset to account for missing baseline data, while the main impact model uses multiple imputation, as described earlier in this appendix. All models control for the same set of baseline characteristics. Standard errors are to the right of each impact estimate.

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

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

Sensitivity to including grade repeaters and skippers

As described above, in our main lottery-based analyses of KIPP middle school impacts, we exclude students who repeated or skipped a grade following the admissions lottery. To test whether impact estimates are sensitive to the approach used to address students these students, we also calculate impacts including these students. In this sensitivity analysis, students' test scores are imputed in repeated or advanced grades: the last score before the student skipped a grade or was held back is used as the outcome measure for that student in every year after starting with the first skipped or repeated grade. As shown in Table E.8 above, the impacts calculated using this approach are generally slightly larger, and the year one math impact is statistically significant in this analysis but not in the main analysis. However, overall, the magnitudes and patterns of impacts are similar.

APPENDIX F

DETAILED ANALYTIC METHODS: MIDDLE SCHOOL (MATCHED-STUDENT ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP middle schools. First we present information on the sample and the baseline equivalence of the treatment and matched comparison group for each outcome. Next we discuss the data used in the analysis, and we conclude the appendix with a discussion of the analytic methods used for the analysis.

Detail on sample

Overall, the middle school analysis included 37 KIPP middle schools. Seven of these middle schools opened in fall 2011, during the period of KIPP network expansion facilitated by i3 scale-up grant issued to KIPP by the federal Department of Education. The total sample of treatment and matched comparison students for each included school is shown in Table F.1, along with the number of cohorts and years of outcome data included.

A key goal for the matching analysis was to include as many schools as possible so that our estimated impacts would be based on as large a portion of the KIPP network as possible. Two criteria were used to select KIPP middle schools for the analysis. First, all included schools had to be established in the 2012–13 school year or earlier to ensure that a minimum of two cohorts of students per school would be observed by spring of 2014.[24] Second, the schools had to be located in jurisdictions (states or school districts) that provided at least three consecutive years of complete, longitudinally linked student-level data for both traditional public and charter schools. These data were needed to track individual KIPP and non-KIPP students in the years prior to middle school enrollment, as well as during the middle school. Throughout this report, we use the term "baseline year" to refer to the school year that began one year prior to when a cohort of students first entered middle school at KIPP; the term "pre-baseline year" refers to the point two years before middle school entry.

The middle school sample includes a combination of schools that were included in our previous report on KIPP middle school impacts (Tuttle et al. 2013) and newer schools that could not be included in that report. In addition, some middle schools that were included in the previous report could not be included here because we could not obtain data for them. In the current study, 25 of the 37 schools were also included in the analyses for the 2013 report.[25] The 37 schools in our sample represent more than half of the KIPP middle school network as of the 2012-2013 school year. Our sample also includes 7 of the 10 KIPP middle schools that opened in fall 2011 or fall 2012.

[24] Throughout the matching analysis, a "cohort" is defined as the group of students who first enrolled in a KIPP middle school at the beginning of a given school year.

[25] For our current study we did not receive district-wide data from Philadelphia (which only provided data on the RCT sample for this study), New Orleans, Los Angeles, Indiana, Oklahoma, and the San Francisco Bay Area (all of these jurisdictions were present in the 2013 study).

Table F.1. Middle school matched-student analytic sample

State	School	Region	Analytic baseline sample			Total Number of KIPP cohorts through 2013-14	Number of KIPP cohorts in data (school years)
			Treatment (N)	Comp. (N)	Total sample size		
AR	KIPP Blytheville College Prep. School	KIPP Delta	169	169	338	4	4 (2010-11 to 2013-14)
	KIPP Delta College Preparatory School	KIPP Delta	601	601	1,202	12	9 (2005-06 to 2013-14)
CO	KIPP Montbello College Prep*	KIPP Denver	322	322	644	3	3 (2011-12 to 2013-14)
	KIPP Sunshine Peak Academy	KIPP Denver	1,134	1,134	2,268	12	11 (2004-05 to 2013-14)
DC	KIPP DC: AIM Academy	KIPP DC	707	707	1,414	9	9 (2005-06 to 2013-14)
	KIPP DC: KEY Academy	KIPP DC	884	884	1,768	13	11 (2003-04 to 2013-14)
	KIPP DC: WILL Academy	KIPP DC	632	632	1,264	8	8 (2006-07 to 2013-14)
GA	KIPP STRIVE Academy	KIPP Atlanta	396	396	792	5	5 (2009-10 to 2013-14)
	KIPP Vision Academy	KIPP Atlanta	377	377	754	4	4 (2010-11 to 2013-14)
	KIPP WAYS Academy	KIPP Atlanta	794	794	1,588	11	10 (2004-05 to 2013-14)
MA	KIPP Academy Boston Middle School*	KIPP Massachusetts	143	143	286	2	2 (2012-13 to 2013-14)
	KIPP Academy Lynn Middle School	KIPP Massachusetts	892	892	1,784	10	10 (2004-05 to 2013-14)
NY	KIPP Academy Middle School (New York)	KIPP NYC	480	480	960	19	10 (2004-05 to 2013-14)
	KIPP AMP Middle School	KIPP NYC	595	595	1,190	9	10 (2005-06 to 2013-14)
	KIPP STAR Harlem Middle School	KIPP NYC	606	606	1,212	11	10 (2004-05 to 2013-14)
	KIPP Infinity Middle School	KIPP NYC	568	568	1,136	9	10 (2005-06 to 2013-14)
	KIPP Washington Heights Middle School*	KIPP NYC	180	180	360	2	2 (2012-13 to 2013-14)
NC	KIPP Charlotte	KIPP Charlotte	335	335	670	7	7 (2007-08 to 2013-14)
	KIPP Gaston College Preparatory	KIPP Eastern NC	931	931	1,862	13	13 (2001-02 to 2013-14)
TN	KIPP Memphis Academy Middle*	KIPP Memphis	210	210	420	2	2 (2012-13 to 2013-14)
	KIPP Memphis Collegiate Middle	KIPP Memphis	1,113	1113	2,226	12	12 (2002-03 to 2013-14)

Table F.1 *(continued)*

State	School	Region	Analytic baseline sample			Total Number of KIPP cohorts through 2013-14	Number of KIPP cohorts in data (school years)
			Treatment (N)	Comp. (N)	Total sample size		
TX	KIPP Austin Academy of Arts & Letters	KIPP Austin	371	371	742	5	4 (2009-10 to 2013-14)[a]
	KIPP Austin Beacon Prep*	KIPP Austin	180	180	360	2	2 (2012-13 to 2013-14)
	KIPP Austin College Prep	KIPP Austin	816	816	1,632	12	10 (2003-04 to 2013-14)[a]
	KIPP Austin Vista Middle School*	KIPP Austin	170	170	340	2	2 (2012-13 to 2013-14)
	KIPP TRUTH Academy	KIPP Dallas-Fort Worth	616	616	1,232	11	10 (2003-04 to 2013-14)[a]
	KIPP Aspire Academy	KIPP San Antonio	834	834	1,668	11	10 (2003-04 to 2013-14)[a]
	KIPP Camino Academy	KIPP San Antonio	362	362	724	4	4 (2010-11 to 2013-14)
	KIPP 3D Academy	KIPP Houston	948	948	1,896	13	11 (2004-05 to 2013-14)
	KIPP Academy (Houston)	KIPP Houston	692	692	1,384	19	10 (2003-04 to 2013-14)[a]
	KIPP Courage College Prep*	KIPP Houston	160	160	320	2	2 (2012-13 to 2013-14)
	KIPP Intrepid Preparatory School	KIPP Houston	516	516	1,032	6	6 (2008-09 to 2013-14)
	KIPP Liberation College Prep	KIPP Houston	531	531	1,062	8	8 (2006-07 to 2013-14)
	KIPP Polaris Academy for Boys	KIPP Houston	518	518	1,036	7	7 (2007-08 to 2013-14)
	KIPP Sharpstown College Prep	KIPP Houston	454	454	908	7	6 (2007-08 to 2013-14)[a]
	KIPP Spirit College Prep	KIPP Houston	682	682	1,364	8	8 (2006-07 to 2013-14)
	KIPP Voyage Academy for Girls	KIPP Houston	393	393	786	5	5 (2009-10 to 2013-14)
	Total (schools opened prior to 2011)	**30 schools**	**18,947**	**18,947**	**37,894**		
	Total (schools opened 2011 or later)	**7 schools**	**1,365**	**1,365**	**2,730**		
	Total (all schools)	**37 schools**	**20,312**	**20,312**	**40,624**		

Notes: Test outcomes are drawn from administrative records for each of the first four years following enrollment in middle school. Treatment students are KIPP middle school students who never attended a KIPP elementary school, and are matched with comparison students based on baseline (grade 4) characteristics. Starred (*) schools opened in fall 2011 or later. Data was either provided by states or individual school districts. In each school year, data files included the following student-level variables: school of enrollment; indicators for gender, race/ethnicity, special education status, free or reduced price lunch status, and limited English proficiency status (except in TX); and test scores in reading, math, and science (except for New York City). History test scores were provided in Atlanta, Texas, and Memphis.

[a] Data does not include cohort from the 2010-2011 school year

While the study's matching procedures (described below) ensured that the treatment group (KIPP students) and comparison group (non-KIPP students) were equivalent at baseline, we did not observe outcome measures for every matched student in every outcome sample. To check that treatment and comparison students included in each of our analysis samples (that is, with valid data on a particular outcome) are equivalent on observable characteristics, we examined baseline equivalence separately for the study's middle school outcomes: math and reading exams one through four years after enrollment at KIPP, and middle school social studies and science exams.[26] We looked at nine baseline characteristics including baseline reading and math test scores; gender, race, special education, limited English proficiency, and free- or reduced price lunch status; and whether the student repeated a grade in the baseline year. The following tables (Tables F.2 through F.6) show the baseline equivalence of each outcome's analytic sample.

Table F.2. Baseline equivalence for matched middle school analysis (analytic sample: year 1 reading and math outcomes)

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 1 outcome						
Reading scores (z-score)	-0.104	-0.087	-0.017	0.011	17,668	17,562
Math scores (z-score)	-0.097	-0.083	-0.013	0.011	17,634	17,502
Student is male	0.489	0.495	-0.006	0.006	18,460	18,314
Student is black	0.512	0.515	-0.003	0.004	18,460	18,314
Student is Hispanic	0.465	0.459	0.006	0.004	18,460	18,314
Special Education	0.066	0.069	-0.003	0.003	18,460	18,314
Limited English Proficiency	0.102	0.108	-0.007	0.005	11,615	11,348
Free- or reduced-price lunch	0.889	0.882	0.007*	0.004	18,460	18,314
Grade Repeaters	0.011	0.013	-0.002	0.001	18,460	18,314
Sample for math year 1 outcome						
Reading scores (z-score)	-0.104	-0.088	-0.016	0.011	17,668	17,564
Math scores (z-score)	-0.097	-0.085	-0.012	0.011	17,646	17,538
Student is male	0.489	0.497	-0.007	0.006	18,461	18,337
Student is black	0.512	0.515	-0.003	0.004	18,461	18,337
Student is Hispanic	0.465	0.459	0.005	0.004	18,461	18,337
Special Education	0.066	0.070	-0.004	0.003	18,461	18,337
Limited English Proficiency	0.102	0.109	-0.008	0.005	11,627	11,370
Free- or reduced-price lunch	0.889	0.882	0.008*	0.004	18,461	18,337
Grade Repeaters	0.011	0.013	-0.002	0.001	18,461	18,337

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 1 year after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator.

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

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

[26] While math and reading exams are typically given annually in middle school, science and social studies are administered once, twice, or not at all, and in different grade levels depending on the jurisdiction. For science and social studies, we used the outcome test from the highest middle school grade in each jurisdiction where the test could be observed for more than one cohort of KIPP students in our sample.

Table F.3. Baseline equivalence for matched middle school analysis (analytic sample: year 2 reading and math outcomes)

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 2 outcome						
Reading scores (z-score)	-0.089	-0.089	0.000	0.014	14,210	13,816
Math scores (z-score)	-0.072	-0.082	0.010	0.015	14,156	13,782
Student is male	0.492	0.498	-0.007	0.008	14,855	14,531
Student is black	0.516	0.518	-0.002	0.005	14,855	14,531
Student is Hispanic	0.461	0.457	0.004	0.005	14,855	14,531
Special Education	0.065	0.070	-0.005	0.004	14,855	14,531
Limited English Proficiency	0.106	0.117	-0.011	0.006	9,479	8,752
Free- or reduced-price lunch	0.886	0.882	0.003	0.004	14,855	14,531
Grade Repeaters	0.011	0.012	-0.002	0.002	14,855	14,531
Sample for math year 2 outcome						
Reading scores (z-score)	-0.089	-0.092	0.003	0.014	14,193	13,798
Math scores (z-score)	-0.072	-0.085	0.012	0.015	14,149	13,782
Student is male	0.492	0.500	-0.008	0.008	14,848	14,523
Student is black	0.516	0.519	-0.003	0.005	14,848	14,523
Student is Hispanic	0.461	0.456	0.005	0.005	14,848	14,523
Special Education	0.065	0.071	-0.006	0.004	14,848	14,523
Limited English Proficiency	0.105	0.117	-0.011	0.006	9,477	8,747
Free- or reduced-price lunch	0.885	0.882	0.003	0.004	14,848	14,523
Grade Repeaters	0.011	0.012	-0.002	0.002	14,848	14,523

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 2 years after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in the “KIPP” and “Comparison” columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator. *Significantly different from zero at the .05 level, two-tailed test.

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

Table F.4. Baseline equivalence for matched middle school analysis (analytic sample: year 3 reading and math outcomes)

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 3 outcome						
Reading scores (z-score)	-0.050	-0.048	-0.002	0.014	11,424	10,959
Math scores (z-score)	-0.042	-0.055	0.013	0.015	11,373	10,916
Student is male	0.485	0.499	-0.014	0.008	11,928	11,505
Student is black	0.563	0.556	0.007	0.005	11,928	11,505
Student is Hispanic	0.411	0.413	-0.002	0.005	11,928	11,505
Special Education	0.064	0.063	0.001	0.004	11,928	11,505
Limited English Proficiency	0.097	0.102	-0.005	0.006	7,585	6,934
Free- or reduced-price lunch	0.872	0.871	0.001	0.005	11,928	11,505
Grade Repeaters	0.013	0.013	0.000	0.002	11,928	11,505
Sample for math year 3 outcome						
Reading scores (z-score)	-0.052	-0.056	0.003	0.014	11,291	10,855
Math scores (z-score)	-0.045	-0.065	0.020	0.015	11,245	10,821
Student is male	0.484	0.499	-0.015	0.008	11,796	11,388
Student is black	0.564	0.556	0.008*	0.005	11,796	11,388
Student is Hispanic	0.410	0.413	-0.002	0.005	11,796	11,388
Special Education	0.064	0.063	0.001	0.004	11,796	11,388
Limited English Proficiency	0.097	0.102	-0.005	0.006	7,575	6,912
Free- or reduced-price lunch	0.871	0.870	0.001	0.005	11,796	11,388
Grade Repeaters	0.013	0.013	0.001	0.002	11,796	11,388

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 3 years after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in the “KIPP” and “Comparison” columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator.

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

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

Table F.5. Baseline equivalence for matched middle school analysis (analytic sample: year 4 reading and math outcomes)

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 4 outcome						
Reading scores (z-score)	-0.009	-0.039	0.029	0.019	7,523	7,210
Math scores (z-score)	-0.022	-0.029	0.006	0.020	7,458	7,159
Student is male	0.482	0.496	-0.015	0.010	7,863	7,528
Student is black	0.587	0.571	0.017**	0.006	7,863	7,528
Student is Hispanic	0.392	0.399	-0.007	0.005	7,863	7,528
Special Education	0.062	0.061	0.001	0.005	7,863	7,528
Limited English Proficiency	0.072	0.079	-0.007	0.005	5,058	4,604
Free- or reduced-price lunch	0.853	0.858	-0.005	0.007	7,863	7,528
Grade Repeaters	0.013	0.014	-0.002	0.002	7,863	7,528
Sample for math year 4 outcome						
Reading scores (z-score)	-0.059	-0.084	0.025	0.020	6,977	6,822
Math scores (z-score)	-0.067	-0.086	0.018	0.021	6,912	6,776
Student is male	0.490	0.497	-0.007	0.011	7,290	7,135
Student is black	0.595	0.575	0.020**	0.006	7,290	7,135
Student is Hispanic	0.385	0.395	-0.011	0.006	7,290	7,135
Special Education	0.062	0.061	0.001	0.005	7,290	7,135
Limited English Proficiency	0.072	0.079	-0.006	0.005	5,010	4,552
Free- or reduced-price lunch	0.856	0.860	-0.004	0.007	7,290	7,135
Grade Repeaters	0.013	0.015	-0.002	0.002	7,290	7,135

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 4 years after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator. *Significantly different from zero at the .05 level, two-tailed test.

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

Table F.6. Baseline equivalence for matched middle school analysis (analytic sample: social studies and science)

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for social studies outcome						
Reading scores (z-score)	0.020	-0.019	0.040	0.022	4,892	4,938
Math scores (z-score)	0.016	-0.025	0.041	0.022	4,895	4,937
Student is male	0.477	0.491	-0.014	0.011	5,191	5,249
Student is black	0.513	0.510	0.003	0.006	5,191	5,249
Student is Hispanic	0.484	0.483	0.002	0.006	5,191	5,249
Special Education	0.038	0.043	-0.005	0.005	5,191	5,249
Limited English Proficiency	0.003	0.006	-0.003	0.003	1,670	1,460
Free- or reduced-price lunch	0.892	0.888	0.004	0.007	5,191	5,249
Grade Repeaters	0.014	0.013	0.001	0.003	5,191	5,249
Sample for science outcome						
Reading scores (z-score)	-0.068	-0.075	0.007	0.015	8,928	8,700
Math scores (z-score)	-0.055	-0.063	0.009	0.016	8,886	8,652
Student is male	0.479	0.493	-0.014	0.008	9,329	9,104
Student is black	0.529	0.528	0.002	0.005	9,329	9,104
Student is Hispanic	0.447	0.445	0.002	0.005	9,329	9,104
Special Education	0.058	0.059	-0.002	0.004	9,329	9,104
Limited English Proficiency	0.089	0.105	-0.015**	0.006	5,136	4,658
Free- or reduced-price lunch	0.875	0.878	-0.003	0.005	9,329	9,104
Grade Repeaters	0.011	0.011	0.000	0.002	9,329	9,104

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have social studies or science test outcome data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator. *Significantly different from zero at the .05 level, two-tailed test.

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

Out of the 10 outcome samples in our middle school analysis, none of them had a statistically significant difference on baseline math or reading test scores. With respect to the baseline demographic characteristics we examined, four show zero statistically significant differences between the treatment and control group and six show only one significant difference on a demographic characteristic. Where we did find a statistically significant difference on a demographic indicator, the magnitude was small: for all of the baseline demographic attributes we examined, the treatment group is within two percentage points of the control group in each outcome sample.

Detail on data

For the matched comparison group analysis, we used de-identified, longitudinally-linked student- data from jurisdictions (states or districts) hosting at least one KIPP school and able to provide student-level records at the time of data collection. The variables from jurisdictions' administrative data systems included: test scores in reading, mathematics, social studies, and science (where middle school scores represent the primary outcome and elementary school scores a key matching variable and baseline covariate); demographic characteristics, used for

matching and as baseline covariates; and schools attended and dates of enrollment, identifying students' exposure to KIPP. Within each jurisdiction, we requested data for all school years beginning with the year prior to the KIPP middle school's first year (to capture baseline data) through the 2013–14 school year. We obtained data from districts for 13 of the 37 schools in the analysis; for the other 24 schools, we obtained records from the state in which the school was located but limited our data to the district (or districts) from which the KIPP school drew students.

To make the analysis of state test scores comparable across states and districts, all raw test scores were converted to z-scores defined relative to the distribution of scores in each grade, year, subject, and jurisdiction. That is, for each jurisdiction associated with a given KIPP school, we calculated the difference between each student's raw score and the mean score in that grade, year, and subject, and then divided the difference by the standard deviation of raw scores in the jurisdiction in that grade, year, and subject. Thus, each z-score reflects a student's achievement level relative to the average student in the relevant cohort and jurisdiction (in terms of the number of standard deviations above or below the mean).[27]

For a variety of reasons, some students may not have valid data in the year when a given outcome was measured. For example, some students may transfer to a jurisdiction outside of our data catchment area, while others may transfer to local private schools or drop out of school altogether. In a small number of cases, students may simply have missing variable values in a given year or subject. We categorize these cases when students disappear from the analytic sample as out-of-district transfers. If KIPP students transfer out-of-district at a different rate than matched comparison students, it could undermine the validity of impact estimates. As noted above, we checked this by examining the baseline equivalence of the sample for each of our 10 outcome measures. There were no differences on baseline reading or math scores on any of the outcome samples and no outcome sample had more than one significant difference on a demographic characteristic. All variables assessed for baseline equivalence are also controlled for in our impact regression model.

Different analytic sample attrition might occur when students are missing one or more baseline or pre-baseline test scores. To address this we imputed missing baseline data, ensuring that all students with at least one recorded baseline test score remain in the sample. For a detailed discussion of our imputation methods, see the following section of this appendix.

Detail on analytic methods

This study relied on a matched comparison group design that used "nearest neighbor" matching to identify a similar comparison student for each treatment student entering a KIPP middle school in grade 5 or grade 6. The validity of our matched comparison group design depends on the ability to eliminate or minimize differences in key characteristics between students who enter KIPP and students in the comparison group who remain in non-KIPP public

[27] By definition, the distribution of student z-scores has a mean of 0 and standard deviation of 1 for each subject (math, reading, science, and social studies) in each of the four outcome years examined in the matching analysis.

schools.[28] Our approach achieved this in two ways. First, we used student-level data that included a rich set of student characteristics and multiple years of baseline (prior to KIPP entry) test scores. We used this information to identify a matched comparison group of students who are similar to KIPP students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in elementary school. By matching on more than one year of baseline test score data, we accounted for achievement levels at the time when students applied to KIPP schools as well as pre-KIPP trends in student achievement. After we identified the matched comparison group, the second feature of our approach estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between KIPP students and comparison students. Specifically, the impact estimates adjust for any differences between KIPP students and the matched comparison group pertaining to demographic characteristics or students' prior two years of math and reading test scores.

The combination of propensity-score matching and OLS accounts for differences in observed baseline characteristics and achievement scores between KIPP students and comparison students (in other words, the differences associated with initial selection into KIPP schools). But it remains possible that KIPP students and comparison students differ in unobserved ways that may affect later test scores. However, previous studies have suggested that applying a combination of propensity-score matching and OLS, as we did here, can succeed in replicating experimental impact estimates in certain contexts (Cook et al. 2008; Bifulco 2012; Furgeson et al. 2012; Tuttle et al. 2013; Fortson et al. 2015). We used the same analytic approach for the propensity score matching model as we implemented in our previous report on KIPP middle schools (Tuttle et al. 2013). As part of that report, we also ran a variety of sensitivity tests to check the robustness of our model to alternatives to our main specifications.[29]

There are several other threats to the validity of these impact estimates that we addressed, including students moving from KIPP middle schools to other district schools (attrition from KIPP schools), students who are retained in grade, and attrition from the sample.

Attrition from KIPP Schools. The fact that some students depart KIPP schools and return to non-KIPP schools in the surrounding district before the end of 8th grade could potentially introduce selection bias if not appropriately handled. At both KIPP and district schools, students who transfer before the end of middle school tend to be those who are not doing as well academically as those who remain (Nichols-Barrer et al. 2012). In this way, an analysis that only includes persistently enrolled KIPP students in the treatment group would positively bias the estimated impact of KIPP schools (that is, make KIPP impacts look more positive than they actually are). We addressed this problem by permanently assigning to the treatment group any student who can be found in the records as ever enrolling at KIPP in grades 5 or 6, regardless of whether the student remained in a KIPP school or transferred elsewhere before the end of middle

[28] Specifically, to produce unbiased impact estimates the design must eliminate differences in student characteristics that could explain academic achievement outcomes and thus be confounded with the treatment of KIPP attendance.

[29] Sensitivity tests included running models that: excluded pre-baseline test scores, included dummies for 4th grade schools, weighted schools by sample size, matched with replacement, used caliper matching, only included students currently enrolled in KIPP in the treatment group, only included students with non-imputed baseline data, and a variety of alternative specifications for grade repeaters. The overall average impact estimates are not sensitive to any of these changes in specification.

school.[30] In other words, a student who enrolled at KIPP in 5th grade for the 2009–10 school year but left KIPP after completing 6th grade in the 2010–11 school year is included in the treatment group for all four years he or she appears in the data (from 2009–10 to 2012–13, inclusive). By including all students observed attending a KIPP school, regardless of whether they stay through eighth grade, we avoid the problem of overstating the effect of KIPP. Instead, this approach is likely to produce a conservative estimate of KIPP's full impact on students during the years they actually attended KIPP schools.

Grade Repetition. KIPP schools retain students in grades 5 and 6 at a substantially higher rate than do conventional public schools in their local districts (Tuttle et al. 2013). This produces a missing data problem for the analysis of state test scores during middle school, as students who repeat a grade do not take the same tests as others in their original cohort. Because KIPP students and comparison students are retained at different rates, our impact estimates could also be biased if we simply excluded all of the retained students from the analysis (doing so would exclude a larger proportion of KIPP students and a smaller proportion of the comparison students). To address this, in the matching analysis of math and reading scores we used information on students' past performance to predict (impute) their outcome scores in the years after retention. For more details on this procedure, see the discussion of imputation methods later in this appendix.

Attrition from the sample. As discussed in the previous section of this appendix, we conducted a detailed battery of baseline equivalence tests to determine if there was differential sample attrition for any of the outcomes we examined. Overall, as shown in the baseline equivalence tables presented above, the pattern of sample attrition for KIPP students is similar to the pattern for students in the matched comparison group for all of the middle school outcomes included in our analysis.

The remainder of this appendix presents the additional details regarding the study's propensity score estimation model, matching procedures, imputation model for baseline test scores, and imputation model for the middle school test scores of grade repeaters.

Propensity Score Matching Procedures

The matching procedure consists of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of KIPP students in the sample; and (3) testing the balance of baseline characteristics between our KIPP sample and matched comparison group.

For the first step, we separated the students in each district-level data set[31] into cohorts—grade-by-year groups for each typical KIPP entry grade (5th and 6th) in each year observed in

[30] In some locations, our analysis may miss some students who exit very soon after arriving at KIPP. Some of the schools included in our study have day-to-day enrollment records, but others are not so finely grained, creating the possibility of losing students who transfer out before designated student count dates, after which they appear in our administrative records data for surrounding schools.

[31] For the purpose of estimating the matching model, we grouped KIPP middle schools together into a single district-wide file when multiple KIPP schools were located in the same district or metropolitan area. After matching

the data. For each cohort of students at a given KIPP school, the pool of eligible comparison students was limited to those in the same district and grade as the KIPP students the year before they first enrolled in a KIPP middle school; comparison students were restricted to those never enrolled in KIPP at any time during middle school. We then performed an iterative propensity score estimation procedure on a combined data set of all cohorts. The dependent variable in this propensity score model is an indicator of whether the student enrolled in a KIPP school in either grade 5 or grade 6.[32] Covariates in the model were selected using an iterative process that identifies the baseline demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. Table F.7 provides an exhaustive list of potential covariates for inclusion in each model.

At a minimum, we required the logistic model to include one year of baseline test scores in both math and reading. The other covariates were iteratively included and tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

Table F.7. List of potential covariates for inclusion in the propensity score estimation model

Observed and imputed (when missing) math and reading baseline test scores from one year prior (always included)
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from one year prior
Observed and imputed (when missing) math and reading baseline test scores from two years prior
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from two years prior
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from one or two years prior are imputed
Dummy variables indicating whether student repeated a grade one or two years prior
Demographic variables (gender, race/ethnicity, special education status, free or reduced price lunch status, and limited English proficiency status, where available)
Interactions of baseline test scores from one year prior and all available demographic variables
Interactions of gender and race/ethnicity variables
Interactions of special education status and race/ethnicity variables
Interactions of free and reduced price lunch status and race/ethnicity variables
Interactions of limited English proficiency status and race/ethnicity variables

was complete, we estimated impacts using the study's impact model (discussed below) separately for each KIPP middle school in the sample.

[32] We did not distinguish between students who enrolled for part of middle school or for the entire duration of middle school. We also did not distinguish between students who enrolled in a single KIPP school and those who enrolled in multiple KIPP schools; before matching, all KIPP students in our data were grouped by the first recorded KIPP school they attended in our data.

Next, we calculated propensity scores for KIPP entry. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (without replacement) of comparison group students to treatment group students, separately by cohort, from within the region of common support. In other words, for each KIPP student with a propensity score that fell within the range of propensity scores found among non-KIPP students, we identified the non-KIPP district student whose propensity score was closest to that of the KIPP student.

We then tested the balance of the KIPP group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, special education status, free and reduced price lunch status, and limited English proficiency status). For the matched comparison group sample associated with each KIPP school, we required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no more than one significant difference on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this covariate for the comparison group are not significantly different from the treatment group at the five percent level.[33] If the first round of matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that KIPP school, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[34] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

Impact model and covariates

To obtain impact estimates using this matched sample, we estimated an ordinary least squares (OLS) regression model that considered all math and reading test score data from grades 5–8 to measure students' outcome test scores and incorporated baseline (4th grade) demographic controls including indicators for gender, race/ethnicity, free/reduced-price lunch status, special education status, grade retention in a baseline year, and limited English proficiency status; cohort (year by entry grade); outcome test grade level; and two years of baseline mathematics and reading test scores (3rd and 4th grade for cohorts entering KIPP in grade 5; 4th and 5th grade for cohorts entering KIPP in grade 6). See Table F.8 for a full list of these covariates. The basic form of the model for each school is defined in equation F1:

$$\text{(F1)}\ y_{it} = a + X_i\beta + \delta_1 T1_{it} + \delta_2 T2_{it} + \delta_3 T3_{it} + \delta_4 T4_{it} + grade_dummies + cohort_dummies + \varepsilon_{it}$$

[33] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table F.2 through Table F.6, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups for any of the outcomes in this study.

[34] If balance was not achieved in the first round of matching for a given school, under our protocol we would remove the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20). In addition, to address estimation problems in the logistic regression we would consider removing terms identifying exceedingly rare attributes in the treatment group.

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and two years of baseline test scores) of student i; $T1_{it}$ through $T4_{it}$ are binary variables for treatment status in up to four years,[35] indicating whether student i had first enrolled at KIPP one, two, three, or four years previously, as of school year t. For example $T3_{it}$ would be equal to 1 for student i at time t if the student had first enrolled at KIPP at time (t-3), regardless of whether the student was still enrolled at KIPP at time t; otherwise, $T3$ would be equal to 0. The model also include a set of dummy indicator variables for each middle school grade and student cohort in the sample. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ_1, δ_2, δ_3, δ_4, and β are parameters or vectors of parameters to be estimated. The estimated coefficient on the treatment indicator, δ_n represents the cumulative impact of n years of KIPP treatment. Robust standard errors were clustered at the student level since individual students could contribute up to four observations to the analysis sample.

We used the model to separately estimate the impact of each KIPP middle school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP middle schools in the sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

Table F.8. List of covariates included in OLS model

Math baseline test score from 1 year prior
Math baseline test score from 2 years prior
Reading baseline test score from 1 year prior
Reading baseline test score from 2 years prior
Gender indicator variable
Set of race/ethnicity indicator variables
Special education status indicator variable
Free or reduced price lunch status indicator variable
Limited English proficiency status indicator variable
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from 1 and 2 years prior are imputed
Set of dummy variables indicating if a student is missing data for demographic variables
Dummy variables indicating whether student repeated grades in either of the two baseline years
Dummy variables for grades 5-8
Dummy variables for each student cohort in the sample

Note: Baseline test scores were imputed when missing. In one jurisdiction data was not available on limited English proficiency status.

[35] Due to a combination of data availability and the year when the KIPP school opened, at six KIPP schools treatment students in the sample received no more than two years of KIPP treatment; at an additional one school, students received no more than three years of treatment.

Imputation for missing baseline data

Our benchmark analyses used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores; imputation was completed separately by treatment status. This imputation process involved estimating the following model:

(F2a) $$Yp_math_{it} = a + X_i\beta + \sum_r \varphi_r Yr_math_{it} + \sum_{q=3}^{8} \gamma_q Yq_reading_{it} + \varepsilon_{it}$$

(F2b) $$Yp_reading_{it} = a + X_i\beta + \sum_r \varphi_r Yr_reading_{it} + \sum_{q=3}^{8} \gamma_q Yq_math_{it} + \varepsilon_{it}$$

where Yp_math_{it} is a single grade *p* math baseline test score for student *i* at time *t; $Yp_reading_{it}$* is a single grade *p* reading baseline test score for student *i* at time *t; X_i* is a vector of demographic characteristics (gender, race/ethnicity, special education status, free or reduced price lunch status and limited English proficiency status, where available) of student *i*; Yr_math_{it} and $Yr_reading_{it}$ are all available grades 3–8, excluding grade *p*, math and reading baseline or outcome test scores for student *i* at time *t*; and Yq_math_{it} and $Yq_reading_{it}$ are all available grades 3–8 math and reading baseline or outcome test scores for student *i* at time *t.* Note that the treatment dummies are not part of the imputation model because imputation is performed separately for the treatment group and then the comparison group.

We first estimated equations (F2a) and (F2b) for baseline test scores one and two years prior to KIPP entry using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores (in other words, values of the right hand side variables in equations F2a and F2b) and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student. We only imputed missing baseline test scores for students who have at least one non-missing baseline test score in either math or reading.

Finally, to obtain the imputed baseline test scores used in our benchmark model, we added a stochastic component to the predicted values of Yp_math_{it} and $Yp_reading_{it}$ obtained from estimating equations (F2a) and (F2b) above. For each student, the stochastic component is randomly selected from the set of all residuals estimated in equations (F2a) and (F2b) for the full sample. The stochastic component is included to ensure that the variance of the imputed baseline test scores is the same as that of the observed values.

While we use these imputed baseline and pre-baseline test scores in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variables were simply treated as being missing from the sample.

To test whether our results are sensitive to this imputation strategy, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores. The results for this smaller sample are nearly identical to our benchmark

impact estimates for the matched middle school analysis (Table F.9). There are no statistically significant differences on any baseline measure and the KIPP impact for each test outcome except year 1 reading remains positive and statistically significant at the 0.01 level, two-tailed test. The magnitude of each impact estimate is very similar to the benchmark estimate as well.

Table F.9. Baseline equivalence and impact estimates on sample with non-imputed baseline data (matched-student middle school impact analysis)

	Treatment group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (reading year 1)	-0.102	17,518	-0.084	17,397	-0.018	0.101
Reading scores (reading year 2)	-0.087	14,079	-0.085	13,679	-0.002	0.897
Reading scores (reading year 3)	-0.049	11,318	-0.045	10,837	-0.004	0.761
Reading scores (reading year 4)	-0.007	7,430	-0.035	7,121	0.027	0.149
Math scores (math year 1)	-0.095	17,525	-0.082	17,413	-0.013	0.244
Math scores (math year 2)	-0.070	14,064	-0.082	13,672	0.012	0.413
Math scores (math year 3)	-0.043	11,185	-0.060	10,741	0.017	0.259
Math scores (math year 4)	-0.065	6,887	-0.083	6,737	0.018	0.386
Reading scores (history)	0.022	4,866	-0.013	4,896	0.034	0.112
Math scores (science)	-0.053	8,826	-0.059	8,587	0.006	0.704

	Treatment group		Comparison group			
Outcome measure	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading year 1	-0.105	17,518	-0.111	17,397	0.005	0.426
Reading year 2	-0.008	14,079	-0.113	13,679	0.105**	0.000
Reading year 3	0.062	11,318	-0.092	10,837	0.154**	0.000
Reading year 4	0.076	7,430	-0.085	7,121	0.161**	0.000
Math year 1	-0.051	17,525	-0.109	17,413	0.057**	0.000
Math year 2	0.091	14,064	-0.141	13,672	0.232**	0.000
Math year 3	0.170	11,185	-0.121	10,741	0.291**	0.000
Math year 4	0.136	6,887	-0.131	6,737	0.268**	0.000
History	0.107	4,866	-0.131	4,896	0.238**	0.000
Science	0.084	8,826	-0.166	8,587	0.249**	0.000

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline and outcome tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Reported impacts are an average of equally weighted impact estimates for each KIPP middle school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. Impacts represent the cumulative effect of KIPP after the noted number of years after admission for math and reading scores, not the marginal effect of an additional year. The grade level of middle school exams used for history and science outcomes varied by jurisdiction. We selected the highest middle school grade level where science or social studies was observed for more than one cohort of KIPP students. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Imputation for students repeating a grade

We also impute the math and reading state test scores of students who repeat a grade if they were retained in one of the study's four outcome years. For example, if a student in the treatment group entered KIPP in grade 5 and then repeated grade 6, they would still be in grade 6 (and would take the grade 6 state assessment) at the end of the third follow-up year. Members of their cohort who remained on track would have taken the grade 7 state assessment. Because the grade repeater's grade 6 assessment score would not be comparable to grade 7 scores, we treat this student's year 3 follow-up score as missing and impute its value. To do so, we use the following approach in the math and reading analyses: for each grade repeater, in the year of repetition and subsequent years, we impute the student's z-score on the cohort-appropriate (rather than grade-appropriate) test by setting his or her score equal to the student's standardized score in the last year prior to grade repetition. In this example, we would use the standardized score of the grade repeater on the grade 6 assessment in the second follow-up year (the score from the first time the student took that assessment). In effect, this imputation procedure assumes students maintain the same percentile rank relative to their cohort in the year of grade retention and in all subsequent years. In other words, we assume that each retained student does neither better or worse in relative terms than before retention. If KIPP in fact has a positive impact on retained students, this would cause us to underestimate KIPP's impact. Conversely, if KIPP has a negative impact, this would cause us to overestimate the impact.

This imputation procedure was not possible for the matching-based analysis of science and social studies test scores—these are often administered only once during middle school (usually in grade 8). For these two subjects, the outcome scores for each student were drawn from the highest available middle school grade where the test could be observed for more than one cohort of KIPP students in our sample, regardless of whether students were retained in prior years.

To test the sensitivity of our results to the method used for retained students, in our prior analyses of KIPP schools we estimated KIPP impacts using several alternative approaches to analyzing the test scores of retained students (Tuttle et al. 2013). For example, we conducted a sensitivity test that assigned the test score of a student to the fifth percentile of the jurisdiction's analysis sample in the grades they would have attended under a "normal" grade progression.[36] Even using this conservative approach, in that analysis the KIPP impact estimates remained positive and statistically significant in both reading and math. However, as we might expect, the magnitude of each positive impact was somewhat smaller than under our benchmark approach (the estimates from this sensitivity test were between 0.02 and 0.06 standard deviations smaller in both math and reading).

[36] On average, students who repeat a grade tend to have test scores that are higher than the fifth percentile in the year before they were retained. For example, in two large urban school districts in our sample the average prior scores of grade repeaters were respectively at the 23rd and 15th percentile in math and the 25th and 19th percentile in reading.

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

APPENDIX G

DETAILED ANALYTIC METHODS: HIGH SCHOOL (MATCHED-STUDENT ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP high schools for students who entered the KIPP network for the first time in grade 9 (our matched- student analysis for high school outcomes). First we present information on the sample and the baseline equivalence of the treatment and matched comparison group for each outcome. Next we discuss the data used in the analysis, and we conclude the appendix with a discussion of the analytic methods used for the analysis.

Detail on sample

Overall, the analysis included 14 different KIPP high schools, nine of which opened in fall 2010 or earlier, meaning we have at least four years of high school data for those schools. Four high schools in our sample opened in fall 2011 and one opened in fall 2013. The 14 high schools represent 9 different states and include 70 percent of the high schools in the KIPP network as of the 2013-14 school year. The total sample of treatment and matched comparison students for each included school is shown in table G.1, along with the number of student cohorts represented in the data.

At the high school level, students served by KIPP have similar characteristics to students at KIPP elementary and middle schools, but differ from other students in the school districts where KIPP high schools in our sample operate. KIPP high school students are more likely to be female, black, and Hispanic than those attending non-KIPP public high schools (Figure G.1).[37] For example, as at the elementary and middle school levels, most KIPP high school students are black (52 percent) or Hispanic (44 percent), compared with 42 and 33 percent among the non-KIPP high school population. KIPP students are also significantly more likely than students at nearby non-KIPP high schools to qualify for free or reduced-price lunch, a proxy for having low family income (84 versus 68 percent). They are slightly (but statistically significantly) less likely to have special education needs than non-KIPP high school students (9 versus 11 percent), while being slightly more likely to be limited English proficiency students (7 versus 6 percent). Finally, KIPP high school students have significantly higher baseline (grade 8) math and reading test scores than non-KIPP students. In particular, on state tests in grade 8, KIPP high school students scored at the 57th percentile in reading and the 60th percentile in math, compared with the 50th percentile in each subject for the non-KIPP high school students. As noted in chapter II, about two-thirds of KIPP high school students attended a KIPP middle school in grade 8, so these grade 8 achievement scores may reflect the influence of KIPP in that year and previous years.

[37] We also compared the baseline (grade 8) characteristics of KIPP high school students to the characteristics of all students who attended a "feeder middle school"—a non-KIPP school where at least one student within the school went on to attend a KIPP high school. The results of that analysis were similar to those presented in Figure III.3.

Table G.1. High school matched-student analytic sample

School	Region	Analytic baseline sample			Number of cohorts (years of school data)
		KIPP (N)	Non-KIPP (N)	Total sample size	
KIPP Denver Collegiate High School	Denver, CO	272	272	544	5 (2009-10 to 2013-14)
KIPP DC: College Preparatory	Washington DC	68	68	136	5 (2009-10 to 2013-14)
KIPP Atlanta Collegiate	Atlanta, GA	113	113	226	3 (2011-12 to 2013-14)
KIPP NYC College Prep High School	New York City	78	78	156	4 (2010-11 to 2013-14)
KIPP Memphis Collegiate High	Memphis, TN	149	149	298	3 (2011-12 to 2013-14)
KIPP Pride High School	Eastern NC	110	110	220	8 (2005-06 to 2013-14)[a]
KIPP Delta Collegiate High School	Arkansas Delta	42	42	84	5 (2008-09 to 2013-14) [b]
KIPP Academy Lynn Collegiate High	Massachusetts	58	58	116	3 (2011-12 to 2013-14)
KIPP Austin Collegiate	Austin, TX	124	124	248	5 (2008-09 to 2013-14)[b]
KIPP University Prep High School	San Antonio, TX	86	86	172	5 (2009-10 to 2013-14)
KIPP Houston High School	Houston, TX	125	125	250	9 (2005-05 to 2013-14)[c]
KIPP Northeast College Preparatory	Houston, TX	23	23	46	1 (2013-14)
KIPP Sunnyside High School	Houston, TX	76	76	152	4 (2010-11 to 2013-14)
KIPP Generations Collegiate	Houston, TX	56	56	112	3 (2011-12 to 2013-14)
Total	**14 schools**	**1,380**	**1,380**	**2,760**	

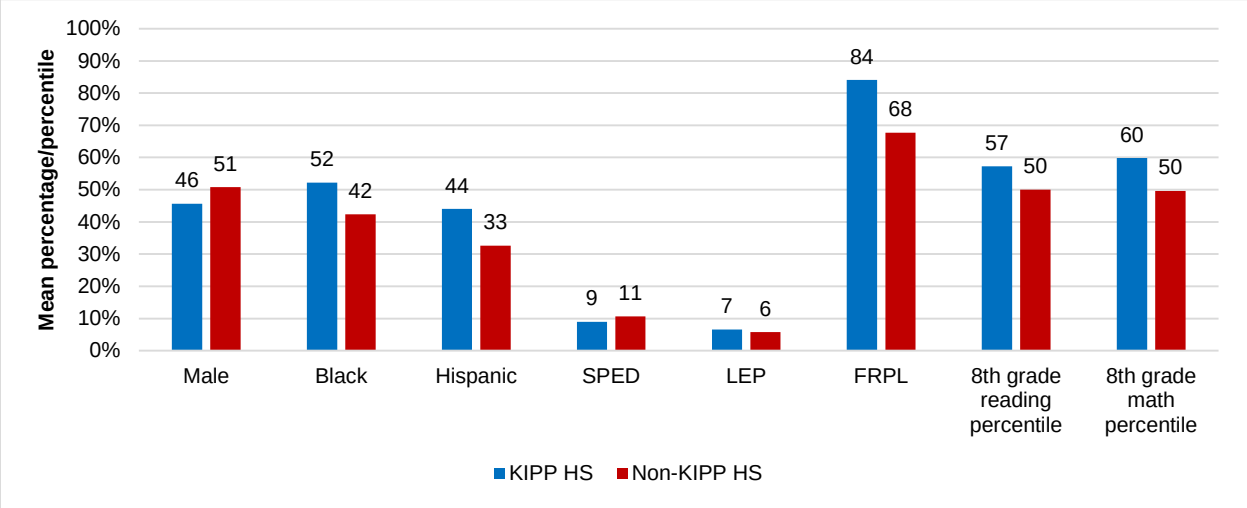
Notes: Test outcomes are drawn from administrative records for each of the first four years following enrollment high school. Treatment students are KIPP high school students who never attended a KIPP middle school, and comparison students are matched based on baseline (grade 8) characteristics.

[a] Data does not include cohort from the 2008-2009 school year

[b] Data does not include cohort from the 2010-2011 school year

[c] Data does not include cohort from the 2011-2012 school year

Figure G.1. Characteristics of KIPP versus non-KIPP high school students



Notes: Sample includes both new entrants and continuing students at the 14 high schools represented in the matched-student sample. Values are proportions unless otherwise indicated. All differences are statistically significant at the 0.05 level, two-tailed test. Each KIPP high school is given equal weight to calculate the overall average and statistical significance.

While the study's matching procedures (described below) ensured that the treatment group (KIPP students) and comparison group (non-KIPP students) were equivalent at baseline, we did not observe outcome measures for every matched student in every outcome sample. To check that treatment and comparison students included in each of our analysis samples are equivalent on observable characteristics, we examined baseline equivalence separately for the study's high school outcomes: math, ELA, science, and social studies exams; and graduating high school within four years. We looked at nine baseline characteristics including baseline reading and math test scores, gender, race, special education, limited English proficiency, free- or reduced price lunch status, and whether the student repeated 8th grade. The following table shows the baseline equivalence of each outcome's analytic sample.

Table G.2. Baseline equivalence for the matched-student high school impact analysis, by outcome sample

Baseline Characteristic	KIPP	Non-KIPP	Difference	Std. Error	# KIPP	# Comparison
Sample for ELA outcome						
Reading scores (z-score)	-0.047	-0.072	0.024	0.052	910	888
Math scores (z-score)	-0.096	-0.086	-0.011	0.049	889	864
Student is male	0.454	0.503	-0.049	0.030	943	918
Student is black	0.530	0.519	0.011	0.021	943	918
Student is Hispanic	0.438	0.436	0.001	0.022	943	918
Special Education	0.061	0.062	-0.001	0.014	943	918
Limited English Proficiency	0.082	0.091	-0.009	0.016	689	649
Free- or reduced-price lunch	0.836	0.847	-0.012	0.019	943	918
Grade Repeaters	0.012	0.016	-0.004	0.006	943	918
Sample for math outcome						
Reading scores (z-score)	0.002	0.001	0.001	0.052	738	703
Math scores (z-score)	-0.084	-0.060	-0.024	0.053	727	692
Student is male	0.446	0.495	-0.050	0.032	763	726
Student is black	0.547	0.540	0.007	0.018	763	726
Student is Hispanic	0.414	0.411	0.002	0.019	763	726
Special Education	0.060	0.071	-0.011	0.015	763	726
Limited English Proficiency	0.083	0.086	-0.003	0.014	566	508
Free- or reduced-price lunch	0.826	0.816	0.010	0.023	763	726
Grade Repeaters	0.015	0.015	0.000	0.008	763	726
Sample for science outcome						
Reading scores (z-score)	-0.081	-0.123	0.042	0.057	680	667
Math scores (z-score)	-0.110	-0.109	0.000	0.049	657	643
Student is male	0.472	0.469	0.004	0.032	697	686
Student is black	0.446	0.447	-0.001	0.024	697	686
Student is Hispanic	0.515	0.496	0.019	0.025	697	686
Special Education	0.051	0.064	-0.013	0.015	697	686
Limited English Proficiency	0.120	0.115	0.005	0.021	443	415
Free- or reduced-price lunch	0.830	0.838	-0.008	0.022	697	686
Grade Repeaters	0.010	0.007	0.003	0.005	697	686
Sample for social studies outcome						
Reading scores (z-score)	0.020	0.012	0.008	0.070	315	308
Math scores (z-score)	-0.039	-0.008	-0.031	0.071	304	297
Student is male	0.480	0.478	0.002	0.043	325	318
Student is black	0.528	0.518	0.010	0.027	325	318
Student is Hispanic	0.448	0.440	0.008	0.027	325	318
Special Education	0.033	0.045	-0.012	0.016	325	318
Limited English Proficiency	0.000	0.000	0.000	0.000	123	100
Free- or reduced-price lunch	0.796	0.793	0.003	0.034	325	318
Grade Repeaters	0.008	0.025	-0.017	0.010	325	318
Sample for 4-year graduation outcome						
Reading scores (z-score)	-0.014	-0.125	0.111	0.067	426	428
Math scores (z-score)	-0.028	-0.155	0.127	0.074	426	426
Student is male	0.485	0.508	-0.023	0.041	440	444
Student is black	0.439	0.420	0.019	0.022	440	444
Student is Hispanic	0.515	0.519	-0.004	0.022	440	444
Special Education	0.050	0.065	-0.015	0.018	440	444
Limited English Proficiency	0.129	0.187	-0.059**	0.022	232	237
Free- or reduced-price lunch	0.789	0.806	-0.018	0.030	440	444
Grade Repeaters	0.007	0.022	-0.015	0.012	440	444

Source: Test outcomes are drawn from state or district administrative records.

Notes: All values in this table are based on non-imputed data and the sample for which we have the outcome test score. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the

"Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each high school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator.

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

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

None of the outcome samples show a statistically significant difference between the treatment and comparison group with respect to baseline test scores in reading and math. In addition, for four of the five outcome samples there were no significant differences for any of the demographic characteristics we examined. The only exception was the four-year graduation outcome, where the treatment group includes fewer limited English proficiency students compared to the control group (by 5.9 percentage points).

Detail on data

For the matched-student comparison group analysis, we used de-identified, longitudinally-linked student-level data from jurisdictions (states or districts) hosting at least one KIPP school and able to provide student-level records at the time of data collection. The variables from jurisdictions' administrative data systems included: test scores in ELA, mathematics, social studies, and science (where high school scores represent the primary outcome and middle school scores represent a key matching variable and baseline covariate); demographic characteristics, used for matching and as baseline covariates; and schools attended and dates of enrollment, identifying students' exposure to KIPP. Within each jurisdiction, we requested data for all school years beginning with two years prior to the KIPP high school's first year (to capture baseline data) through the 2013–14 school year. Of the 14 high schools in our sample, we obtained data from districts for four of the schools. For the other ten we obtained records from the state in which the high school was located and then limited the data to the district (or districts) from which the KIPP school drew students.

Importantly, the test outcomes at the high school level are different than for statewide grade-specific exams administered during middle school. A few states include statewide assessment tests for high school students in particular grades (as in middle school) but other states have end-of-course exams that students take after they complete specific courses, such as algebra I or biology. For end-of-course exams, students may complete the exam in a variety of different grades and years, depending on their course progression and grade progression pattern. For our analysis, outcome scores from these exams were limited to the test from the first year each student completed a given end-of-course exam (that is, we disregarded scores from retests if a student was retained and took the same exam in multiple years). To make the analysis of state high school test scores comparable across states and districts, all raw test scores were converted to z-scores defined relative to the distribution of scores in each year, subject, and jurisdiction. That is, for each jurisdiction associated with a given KIPP school, we calculated the difference between each student's raw score and the mean score recorded in that year across all high school grades in that subject, and then divided the difference by the standard deviation of raw scores in the jurisdiction in that year and subject. Thus, each z-score reflects a student's achievement level relative to the average student in the relevant cohort and jurisdiction (in terms of the number of standard deviations above or below the mean).[38]

For some jurisdictions and subjects, there were multiple different end-of-course exams available for a given academic subject. In these cases, we selected the exam where the test-taking pattern among KIPP high school students was most similar to the pattern in the matched

[38] By definition, the distribution of student z-scores in high school has a mean of 0 and standard deviation of 1 for each subject (math, ELA, science, and social studies).

comparison group in terms of the number of test takers and the timing of the exam. Table G.3 summarizes the outcome exams used in each jurisdiction.

The high school analysis also includes an outcome indicator for graduating within four years after entering 9th grade. There are two important limitations to the study's graduation outcome. First, our propensity score matching approach relies on the assumption that the pre-KIPP characteristics observed in our data (middle school test scores and demographic attributes) fully capture attributes that are associated both with selection into KIPP and the outcomes of interest. In the case of graduation, this assumption may be somewhat questionable because some attributes that are likely to be correlated with graduation, such as parental involvement or students' commitment to achieving long-term goals, may not be fully captured in our data. A second important limitation is that the administrative data used in this study only identifies graduation for the students who remain in the data through grade 12. In other words, the graduation indicator we use cannot distinguish between dropouts (who did not graduate) and students who transferred to private school or to a different school district (who may or may not have graduated high school). Both of these groups of students are classified as "non-graduates" in the analysis. This data limitation is a potential source of bias in the graduation impact estimates. If KIPP students transfer to other districts or private schools after grade 9 at a greater rate than comparison students, our impact estimates would be more negative than the true impact of KIPP. Conversely, if the comparison group transfers out of district more often during high school than the treatment group, our impact estimates would be more positive than the true impact of KIPP. The baseline equivalence results presented above for the four-year graduation indicator provides evidence that differential attrition between KIPP and non-KIPP students does not bias the impact estimate. (Only the Limited English Proficiency indicator showed a statistically significant difference between the two groups).

There could also be a missing data issue for the test outcomes. For a variety of reasons, some students may not have valid data in the year when a given outcome was measured. For example, some students may transfer to a jurisdiction outside of our data catchment area, while others may transfer to local private schools or drop out of school altogether. In a small number of cases, students may simply have missing variable values in a given year or subject. We categorize these cases when students disappear from the analytic sample as out-of-district transfers. If KIPP students transfer out-of-district at a different rate than matched comparison students, it could undermine the validity of impact estimates. To check this, we examined baseline equivalence of the sample for each of our four academic outcome measures (math, ELA, science, and social studies). As noted above, there were no statistically significant differences on any of the baseline characteristics for each test outcome sample. All variables assessed for baseline equivalence are also controlled for in our impact regression model.

Table G.3. High school test outcomes and data source by school

High school	Data entity	Math	ELA	Science	Social studies
KIPP Denver Collegiate High School	CO Dept. of Education	CSAP grade 9 (2010-11), TCAP grade 9 (2012-14)	CSAP grade 9 (2010-11), TCAP grade 9 (2012-14)	CSAP grade 10 (2010-11), TCAP grade 10 (2012-13), ACT science (2014)	n.a.
KIPP DC: College Preparatory	OSSE (Washington D.C.)	CAS 10th grade	CAS 10th grade	n.a.	n.a.
KIPP Atlanta Collegiate	Atlanta Public Schools	Geometry	Lit Comp	Biology	U.S. History
KIPP NYC College Prep High School	NYC Dept. of Education	Integrated Algebra Regents	ELA Regents	Living Environment Regents	n.a.
KIPP Memphis Collegiate High	Shelby County Schools	Algebra II	English I	Biology	U.S. History
KIPP Pride High School	NC Dept. of Public Instruction/NCERDC	Algebra I	English I (1999-2012), English II (2013-2014)	Biology	U.S. History
KIPP Delta Collegiate High School	AR Dept. of Education	Geometry	Literacy	n.a.	n.a.
KIPP Academy Lynn Collegiate High	MA Dept. of Education	MCAS grade 10	MCAS grade 10	MCAS Biology grade 10	n.a.
KIPP Austin Collegiate	Texas Education Agency	TAKS grade 11	TAKS grade 11	TAKS grade 11	TAKS grade 11
KIPP University Prep High School	Texas Education Agency	TAKS grade 11	TAKS grade 11	TAKS grade 11	TAKS grade 11
KIPP Houston High School	Texas Education Agency	TAKS grade 11	TAKS grade 11	TAKS grade 11	TAKS grade 11
KIPP Northeast College Preparatory	Texas Education Agency	n.a.	English I	Biology	n.a.
KIPP Sunnyside High School	Texas Education Agency	TAKS 9 (2011); Algebra I (2012-14)	TAKS 9 (2011); Reading I (2012-13); English I (2014)	Biology	World Geography
KIPP Generations Collegiate	Texas Education Agency	n.a.	TAKS 9 (2011); Reading I (2012-14)	Biology	World Geography

Notes: Abbreviations for data entities correspond to the following: OSSE = Office of the State Superintendent of Education; NCERDC = North Carolina Education Research Data Center. Abbreviations for exams correspond to the following: TAKS = Texas Assessment of Knowledge and Skills; TCAP = Transitional Colordado Assessment Program; CSAP = Colorado Student Assessment Program; CAS = Comprehensive Assessment System; MCAS = Massachusetts Comprehensive Assessment System

Different analytic sample attrition might occur when students are missing one or more baseline or pre-baseline test scores. To address this we imputed missing baseline data, ensuring that all students with at least one recorded baseline test score remain in the sample. For a detailed discussion of our imputation methods, see the discussion of imputation methods below.

Detail on analytic methods

This study relied on a matched comparison group design that used "nearest neighbor" matching to identify a similar comparison student for each treatment student entering a KIPP high school in grade 9. The validity of our matched comparison group design depends on the ability to eliminate or minimize differences in key characteristics between students who enter KIPP and students in the comparison group who remain in non-KIPP public schools.[39] Our approach achieved this in two ways. First, we used student-level data that included a rich set of student characteristics and multiple years of baseline (prior to KIPP entry) test scores. We used this information to identify a matched comparison group of students who are similar to KIPP students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in middle school. By matching on more than one year of baseline test score data, we accounted for achievement levels at the time when students applied to KIPP high schools as well as pre-KIPP trends in student achievement. After we identified the matched comparison group, the second feature of our approach estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between KIPP students and comparison students. Specifically, the impact estimates adjust for any differences between KIPP students and the matched comparison group pertaining to demographic characteristics or students' prior two years of math and reading test scores.

The combination of propensity-score matching and OLS accounts for differences in observed baseline characteristics and achievement scores between KIPP students and comparison students (in other words, the differences associated with initial selection into KIPP schools). But it remains possible that KIPP students and comparison students differ in unobserved ways that may affect later test scores. However, previous studies have suggested that applying a combination of propensity-score matching and OLS, as we did here, can succeed in replicating experimental impact estimates in certain contexts (Cook et al. 2008; Bifulco 2012; Furgeson et al. 2012; Tuttle et al. 2013; Fortson et al. 2015).

There are several other threats to the validity of these impact estimates that we addressed, including students moving from KIPP middle schools to other district schools and attrition from the sample.

Attrition from KIPP Schools. The fact that some students depart KIPP schools and return to non-KIPP schools in the surrounding district before the end of 12th grade could potentially introduce selection bias if not appropriately handled. If lower-performing student tend to exit KIPP high schools before graduation, an analysis that only includes the persistently enrolled KIPP students in the treatment group would positively bias the estimated impact of KIPP schools (that is, make KIPP impacts look more positive than they actually are). We addressed this

[39] Specifically, to produce unbiased impact estimates the design must eliminate differences in student characteristics that could explain academic achievement outcomes and thus be confounded with the treatment of KIPP attendance.

problem by permanently assigning to the treatment group any student who can be found in the records as enrolling at KIPP for the first time in grade 9, regardless of whether the student remained in a KIPP high school or transferred elsewhere before the end of high school.[40] In other words, a student who enrolled at KIPP in 9th grade for the 2009–10 school year but left KIPP after completing 10th grade in the 2010–11 school year is included in the treatment group for all four years he or she appears in the data (from 2009–10 to 2012–13, inclusive). By including all students observed as entering a KIPP school in grade 9, regardless of whether they stay through the end of grade 12, we avoid the problem of overstating the effect of KIPP. Instead, this approach is likely to produce a conservative estimate of KIPP's full impact on students during the years they actually attended KIPP schools.

Attrition from the sample. As discussed in the previous section of this appendix, we conducted a detailed battery of baseline equivalence tests to determine if there was differential sample attrition for any of the outcomes we examined. Overall, as shown in the baseline equivalence tables presented above, the pattern of sample attrition for KIPP students is similar to the pattern for students in the matched comparison group for all of the high school outcomes included in our analysis.

The remainder of this appendix presents the additional details regarding the study's propensity score estimation model, matching procedures, and imputation model for baseline test scores.

Propensity Score Matching Procedures

The matching procedure consists of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of KIPP students in the sample; and (3) testing the balance of baseline characteristics between our KIPP sample and matched comparison group.

For the first step, we separated the students in each district-level data set into cohorts—grade-by-year groups for the KIPP high school entry grade (grade 9) in each year observed in the data. For each cohort of students at a given KIPP school, the pool of eligible comparison students was limited to those in the same district and grade as the KIPP students the year before they first enrolled in a KIPP high school; comparison students were restricted to those never enrolled in KIPP at any time during elementary or middle school. We then performed an iterative propensity score estimation procedure on a combined data set of all cohorts. The dependent variable in this propensity score model is an indicator of whether the student enrolled in a KIPP school in grade 9.[41] Covariates in the model were selected using an iterative process that identifies the baseline

[40] In some locations, our analysis may miss some students who exit very soon after arriving at KIPP. Some of the schools included in our study have day-to-day enrollment records, but others are not so finely grained, creating the possibility of losing students who transfer out before designated student count dates, after which they appear in our administrative records data for surrounding schools.

[41] We did not distinguish between students who enrolled for part of high school or for the entire duration of high school. We also did not distinguish between students who enrolled in a single KIPP high school and those who enrolled in multiple KIPP schools; before matching, all KIPP students in our data were grouped by the first recorded KIPP high school they attended in our data.

demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. Table G.4 provides an exhaustive list of potential covariates for inclusion in each model.

At a minimum, we required the logistic model to include one year of baseline test scores in both math and reading. The other covariates were iteratively included and tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

Table G.4. List of potential covariates for inclusion in the propensity score estimation model

Observed and imputed (when missing) math and reading baseline test scores from one year prior (always included)
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from one year prior
Observed and imputed (when missing) math and reading baseline test scores from two years prior
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from two years prior
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from one or two years prior are imputed
Dummy variables indicating whether student repeated a grade one or two years prior
Demographic variables (gender, race/ethnicity, special education status, free or reduced price lunch status, and limited English proficiency status, where available)
Interactions of baseline test scores from one year prior and all available demographic variables
Interactions of gender and race/ethnicity variables
Interactions of special education status and race/ethnicity variables
Interactions of free and reduced price lunch status and race/ethnicity variables
Interactions of limited English proficiency status and race/ethnicity variables

Next, we calculated propensity scores for KIPP entry. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (without replacement) of comparison group students to treatment group students, separately by cohort, from within the region of common support. In other words, for each KIPP student with a propensity score that fell within the range of propensity scores found among non-KIPP students, we identified the non-KIPP district student whose propensity score was closest to that of the KIPP student.

We then tested the balance of the KIPP group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, special education status, free and reduced price lunch status, and limited English proficiency status). For the matched comparison group sample associated with each KIPP school, we required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no more than one significant difference on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this

covariate for the comparison group are not significantly different from the treatment group at the five percent level.[42] If the first round of matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that KIPP school, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[43] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

Impact model and covariates

To obtain impact estimates using this matched sample, we estimated an ordinary least squares (OLS) regression model for each examined outcome (test scores in math, ELA, science, and social studies, as well as high school graduation). The model incorporated baseline (8th grade) demographic controls including indicators for gender, race/ethnicity, free/reduced-price lunch status, special education status, grade retention in a baseline year, and limited English proficiency status; cohort (year by entry grade); and two years of baseline mathematics and reading test scores (7th and 8th grade). See Table G.5 for a full list of these covariates. The basic form of the model for each school is defined in equation G1:

$$\text{(G1)} \qquad y_{it} = a + X_i\beta + \delta treat_{it} + cohort_dummies + \varepsilon_{it}$$

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and two years of baseline test scores) of student i; $treat_{it}$ is a binary variables for treatment status indicating whether student i had entered KIPP in grade 9. The model also include a set of dummy indicator variables for each student cohort in the sample. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ and β are parameters or vectors of parameters to be estimated. The estimated coefficient on the treatment indicator, δ represents the impact of the KIPP high school on the observed outcome. Robust standard errors were clustered at the student level.

We used the model to separately estimate the impact of each KIPP high school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP high schools in the sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

[42] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table G.2, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups for any of the outcomes in this study.

[43] If balance was not achieved in the first round of matching for a given school, under our protocol we would remove the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20). In addition, to address estimation problems in the logistic regression we would occasionally remove terms identifying exceedingly rare attributes in the treatment group.

Table G.5. List of covariates included in OLS model

Math baseline test score from 1 year prior
Math baseline test score from 2 years prior
Reading baseline test score from 1 year prior
Reading baseline test score from 2 years prior
Gender indicator variable
Set of race/ethnicity indicator variables
Special education status indicator variable
Free or reduced price lunch status indicator variable
Limited English proficiency status indicator variable
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from 1 and 2 years prior are imputed
Set of dummy variables indicating if a student is missing data for demographic variables
Dummy variables indicating whether student repeated grades in either of the two baseline years
Dummy variables for each student cohort in the sample

Note: Baseline test scores were imputed when missing. In one jurisdiction data was not available on limited English proficiency status.

Imputation for missing baseline data and retained students

Here we explain in greater detail how our analysis handled missing data when students were missing baseline or pre-baseline test score data in grade 7 or 8. Our analyses used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores; imputation was completed separately by treatment status. This imputation process involved estimating the following model:

(G2a) $$Yp_math_{it} = a + X_i\beta + \sum_r \varphi_r Yr_math_{it} + \sum_{q=3}^{8} \gamma_q Yq_reading_{it} + \varepsilon_{it}$$

(G2b) $$Yp_reading_{it} = a + X_i\beta + \sum_r \varphi_r Yr_reading_{it} + \sum_{q=3}^{8} \gamma_q Yq_math_{it} + \varepsilon_{it}$$

where Yp_math_{it} is a single grade *p* math baseline test score for student *i* at time *t;* $Yp_reading_{it}$ is a single grade *p* reading baseline test score for student *i* at time *t;* X_i is a vector of demographic characteristics (gender, race/ethnicity, special education status, free or reduced price lunch status and limited English proficiency status, where available) of student *i*; Yr_math_{it} and $Yr_reading_{it}$ are all available grades 3–8, excluding grade *p*, math and reading baseline or outcome test scores for student *i* at time *t*; and Yq_math_{it} and $Yq_reading_{it}$ are all available grades 3–8 math and reading baseline or outcome test scores for student *i* at time *t.* Note that the treatment dummies are not part of the imputation model because imputation is performed separately for the treatment group and then the comparison group.

We first estimated equations (G2a) and (G2b) for baseline test scores one and two years prior to KIPP entry using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores (in other words, values of the right hand side variables in equations G2a and G2b) and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student. We only imputed

missing baseline test scores for students who have at least one non-missing baseline test score in either math or reading.

Finally, to obtain the imputed baseline test scores used in our benchmark model, we added a stochastic component to the predicted values of Yp_math_{it} and $Yp_reading_{it}$ obtained from estimating equations (G2a) and (G2b) above. For each student, the stochastic component is randomly selected from the set of all residuals estimated in equations (G2a) and (G2b) for the full sample. The stochastic component is included to ensure that the variance of the imputed baseline test scores is the same as that of the observed values.

While we use these imputed baseline and pre-baseline test scores in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variables were simply treated as being missing from the sample.

To test whether our results are sensitive to this imputation strategy, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores. The results for this smaller sample are nearly identical to our benchmark impact estimates for the matched-student high school impact analysis (Table G.6). There are no statistically significant differences on any baseline measure and the KIPP impact in ELA, math, and science remains positive and statistically significant while impacts in social studies and high school graduation are positive but not statistically significant. The magnitude of each impact estimate is nearly identical to the benchmark estimate as well.

Table G.6. Baseline equivalence and impact estimates on sample with non-imputed baseline data (matched-student high school impact analysis)

Baseline measure (analyzed outcome for this sample)	Treatment group		Comparison group			
	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	-0.060	887	-0.089	861	0.028	0.591
Math scores (math)	-0.084	725	-0.059	691	-0.026	0.631
Reading scores (social studies)	0.000	304	0.001	297	-0.002	0.982
Math scores (science)	-0.111	656	-0.109	643	-0.002	0.973
Reading scores (4-year graduation)	-0.014	426	-0.125	426	0.111	0.099
Math scores (4-year graduation)	-0.028	426	-0.155	426	0.127	0.088
Free and reduced-price lunch status (4-year graduation)	0.806	426	0.809	426	-0.003	0.924
	Treatment group		**Comparison group**			
Outcome measure	**Adjusted mean**	**Sample size**	**Mean**	**Sample size**	**Impact estimate**	**p-value**
ELA achievement	0.111	887	-0.065	861	0.175**	0.000
Mathematics achievement	0.236	725	-0.036	691	0.273**	0.000
Science achievement	0.105	656	-0.223	643	0.328**	0.000
Social studies achievement	-0.134	304	-0.149	297	0.015	0.795
Four-year high school graduation	0.707	426	0.672	426	0.035	0.355

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

APPENDIX H

DETAILED ANALYTIC METHODS: HIGH SCHOOL (MATCHED-SCHOOL ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP high schools on student achievement for continuing students. Specifically, these results are estimates of the additional benefit of attending a KIPP high school among students who attended a KIPP middle school on achievement in reading, language, and math as measured by TerraNova test scores, as well as impacts on nonacademic outcomes. First we present information on the sample and baseline equivalence of the treatment and matched comparison group for each outcome. Next we discuss the data used in the analysis, and we conclude with a discussion of the analytic methods used for the analysis.

Analysis sample

Since the majority of students attending KIPP high schools also attended a KIPP middle school, our challenge was to identify a credible and rigorous comparison group for these students. We developed two different but complementary approaches, each with an associated sample of schools and students, to compare outcomes for KIPP middle school students who had an option to attend a KIPP high school with those who did not have an option to attend a KIPP high school. Those comparison students attended a wide variety of other high schools, including private, magnet, boarding, traditional public, or other charter high schools (discussed in more detail below). Our results are analogous to "intent-to-treat" findings, since not all students with the option to attend a KIPP high school do so. However, rates of enrollment in KIPP high schools among KIPP middle-school graduates in our sample are generally high where the option is available, at 70 percent overall and ranging from 59 to 83 percent across feeder KIPP middle schools. Effects on students actually enrolling in KIPP high schools would be larger than the "intent-to-treat" impacts on students with the opportunity to enroll.

For the first model (Same KIPP Middle School, Adjacent Cohorts), we focus on a set of KIPP high schools in their first year of operation. The treatment group in this model includes 8th-grade KIPP middle school students who had the option to attend the local KIPP high school in the first year of its operation. The comparison group includes the previous cohort of 8th-grade KIPP students from the same middle school who did not have the option to attend the local KIPP high school because it had not yet opened. Because sample sizes are too small to restrict the comparison group, we do not employ student-level matching in this model. Rather, we compare later outcomes of an entire class of KIPP middle school graduates in one cohort with the later outcomes of the entire classroom of KIPP middle school graduates from the previous cohort. There are five high schools (served by six feeder middle schools) included in this analysis, yielding an analytic sample of 467 students (229 treatment, 238 comparison).

For the second model (Same Cohort, Matched Middle Schools), the treatment group includes 8th-grade KIPP middle school students in 2008-09 who had the option to attend the local KIPP high school. The comparison group includes 8th-grade KIPP students from different middle schools in 2008-09 in regions with no KIPP high school open at the time. To define a sample that was equivalent at baseline (grade 8), we first identified a set of comparison KIPP middle schools that most resembled the feeder KIPP middle schools on the basis of average school-level characteristics (race/ethnicity, baseline achievement on a nationally-normed test, and baseline test instrument—either the SAT-10 or the MAP). Then, within these matched sets of schools, we conducted student-level propensity score matching to identify the individual comparison student who was the closest match to each treatment student on the basis of gender,

race/ethnicity, baseline achievement in reading and math, and whether the student was old for his/her grade at baseline. There are five high schools included in this analysis, served by six feeder middle schools. Together with students from five matched KIPP middle schools without a high school option, the analytic sample comprises 550 students (275 treatment, 275 comparison).

Across the two models, we estimate impacts for eight unique KIPP high schools and 933 students (464 treatment, 469 control). Two schools (University College Prep in San Antonio and DC College Prep) are included in both models, which allows us to compare estimates across models. A key difference between the models is that the Same KIPP Middle School, Adjacent Cohorts model examines only schools serving their first cohort of students, whereas the Same Cohort, Matched Middle Schools model includes a combination of new and more established high schools. The feeder middle school and studied cohort for each included school in the combined sample is shown in table H.1.

Table H.1. High school matched-school achievement analysis, combined sample

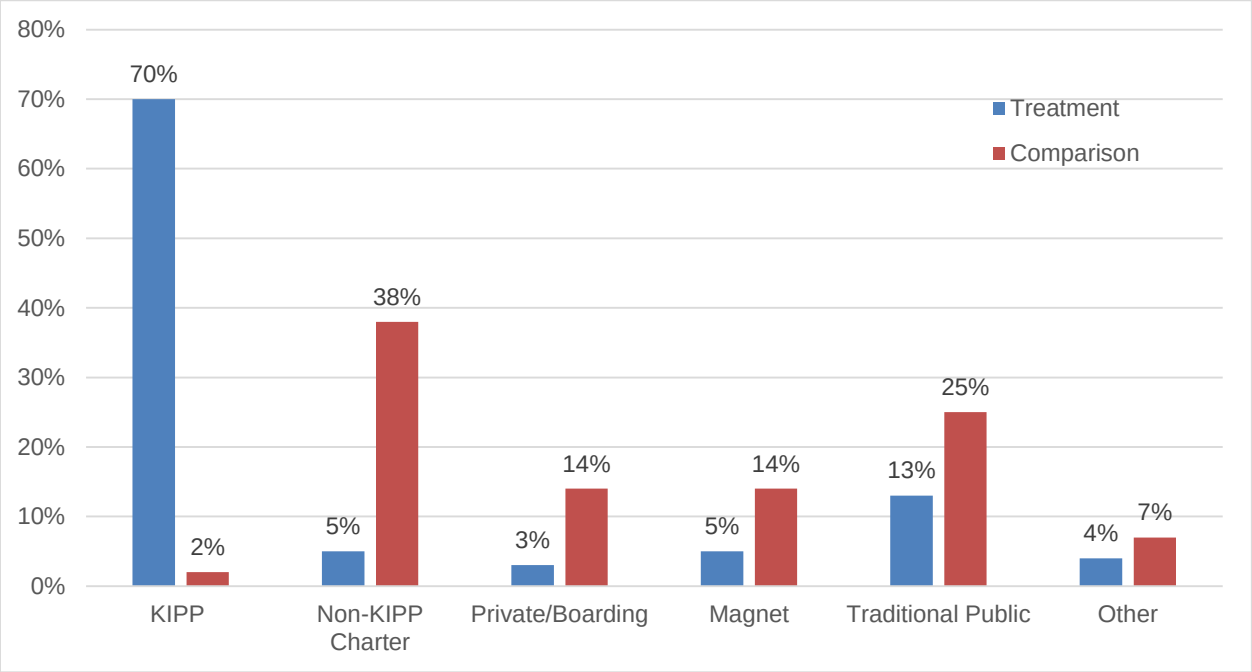
KIPP High School	Region	Year opened (cohort studied)	Same KIPP middle school, adjacent cohorts: KIPP feeder MS	Same cohort, matched middle school:	
				KIPP feeder MS	KIPP matched MS
Austin Collegiate	KIPP Austin	2008 (1st)	Austin College Prep	--	--
DC College Prep	KIPP DC	2009 (1st)	KEY Academy	KEY Academy, AIM Academy	Ascend, Philadelphia CS
Dubois Collegiate	KIPP Philadelphia	2010 (1st)	Philadelphia Charter	--	--
King Collegiate	KIPP Bay Area	2007 (3rd)	--	Summit Academy	Adelante, LA Prep, San Francisco Bay
Newark Collegiate	KIPP New Jersey	2007 (3rd)	--	TEAM Academy	Ascend, Philadelphia CS
NYC College Prep	KIPP NYC	2009 (1st)	KIPP Academy Middle (NYC), STAR	--	--
San Jose Collegiate	KIPP Bay Area	2008 (2nd)	--	Heartwood	Adelante, LA Prep, San Francisco Bay
University College Prep	KIPP San Antonio	2009 (1st)	Aspire	Aspire	Adelante, LA Prep, San Francisco Bay

Notes: Treatment students are KIPP middle school students who attended schools with an available KIPP high school, and are matched with comparison students at schools that did not have a KIPP high school option available.

As with the other grade levels, it is important to understand differences in the types of schools attended by the treatment and comparison group students to help provide context for the two groups' experiences in school and thus any impacts we observe (Figure H.1). For example, if all comparison group students attended traditional public schools, then the impact estimate would capture the effects of KIPP high schools versus these traditional public high schools. If most comparison students attended private or boarding schools, by contrast, the interpretation of the impact estimate would be different.

Using data from our matched-school sample of high school students, we investigate the types of high schools attended by KIPP middle school students, both when they have the option to attend a KIPP high school and when they do not. In the treatment group—those that have the option of a KIPP high school—the majority of students (70 percent) attend the KIPP high school. Another 13 percent attend traditional public schools and the rest are fairly evenly distributed between non-KIPP charter high schools, magnet schools, and private or boarding schools. The distribution of high school types attended is much different in the absence of a KIPP high school option, where the largest group of comparison students attend a non-KIPP charter school (38 percent) and another 14 percent attend magnet schools, another option involving choice. A much larger percentage attend private schools than do so when a KIPP high school is an option (14 percent versus 3 percent), but less than twice as many attend a traditional public school (25 percent versus 13 percent). These patterns suggest that KIPP high schools enroll students who would have otherwise enrolled in a wide variety of school types.

Figure H.1. Type of high schools attended by KIPP middle school students



Notes: Sample refers to the students in the matched-school analysis of KIPP high schools. Proportions reflect the schools students reported attending during grade 11, based on KIPP alumni records and a study-administered survey.

To check that each of our outcome samples are equivalent on observable characteristics, we examined baseline equivalence separately for each of the study's matched samples. We looked at six baseline, or 8th grade, characteristics including reading and math test scores; gender, race, special education, limited English proficiency, and free- or reduced price lunch status; and whether the student repeated a grade in the baseline year. The following tables (Tables H.2 through H.4) show the baseline equivalence of each analytic sample.

Table H.2. Baseline equivalence for matched high school achievement analysis (combined sample)

Baseline characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (z-score)	0.366	0.254	0.112	0.118	464	469
Math scores (z-score)	0.743	0.647	0.056	0.248	464	469
Student is male	0.450	0.443	0.008	0.879	464	469
Student is black	0.502	0.503	-0.001	0.982	464	469
Student is Hispanic	0.425	0.433	-0.008	0.855	464	469
Student is old for grade	0.207	0.213	-0.006	0.859	464	469

Note: All values in this table are based on the sample for which we have test outcome data. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution.

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

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

Table H.3. Baseline equivalence for matched high school achievement analysis (same middle school, adjacent cohorts sample)

Baseline characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (z-score)	0.425	0.357	0.068	0.409	208	213
Math scores (z-score)	0.772	0.800	-0.028	0.718	208	213
Student is male	0.447	0.413	0.034	0.483	208	213
Student is black	0.519	0.521	-0.002	0.969	208	213
Student is Hispanic	0.476	0.408	0.068	0.164	208	213
Student is old for grade	0.250	0.254	-0.004	0.934	208	213

Note: All values in this table are based on the sample for which we have test outcome data. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution.

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

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

Table H.4. Baseline equivalence for matched high school achievement analysis (same cohort, matched middle schools sample)

Baseline characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (z-score)	0.131	0.140	-0.009	0.926	256	256
Math scores (z-score)	0.568	0.561	0.007	0.955	256	256
Student is male	0.453	0.465	-0.012	0.858	256	256
Student is black	0.488	0.488	0.000	1.000	256	256
Student is Hispanic	0.383	0.453	-0.070	0.279	256	256
Student is old for grade	0.172	0.180	-0.008	0.871	256	256

Note: All values in this table are based on the sample for which we have test outcome data. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution.

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

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

For the combined analysis sample, among the students with non-missing outcome data, there were no statistically significant differences between the treatment and control group on any baseline characteristics. Assessing baseline equivalence separately for the two different models, the treatment and comparison groups in the Same Cohort, Matched Middle School model are equivalent, which is consistent with the model's design due to the matching process. In the Same KIPP Middle School, Adjacent Cohorts model, some differences between groups are closer to being significant (but none that are significant at the 5% level), reinforcing the importance of controlling for these characteristics in the impact model.

Data

Baseline test score data is from KIPP-administered math and reading tests from grade 8. For the Same Cohort, Matched Middle School analysis, the KIPP middle schools either administered the Measures of Academic Progress (MAP) or the Stanford Achievement Test (SAT-10), and the schools were divided into two groups by baseline test score type for this analysis. For the Same Middle School, Adjacent Cohorts analysis the KIPP-administered baseline test was used unless the KIPP middle school changed tests between the two cohorts included in our analysis. In the cases where different tests were used across the two cohorts for a particular KIPP middle school, data on the state-administered math and reading tests for grade 8 were used for the baseline test scores.

The primary outcome measure for the achievement analysis is the study-administered TerraNova assessment (Form G, Level 21/22) in reading, language, and math, administered in 2011, 2012, or 2013 (depending on the cohort and treatment group) in treatment and comparison students' third year after grade 8 (typically grade 11). We measure students' performance on the TerraNova assessment with z-scores that were standardized to capture student achievement relative to that of a nationally representative norming population.

We analyzed the impact of KIPP high schools on student behavior and attitudes covering three broad areas: student motivation and engagement, education goals and college preparation,

and school experiences and satisfaction. Data was collected from student survey conducted between March and July in 2012, 2013, or 2014 (depending on the cohort and treatment group), the spring of the students' fourth year after completing 8th grade at a KIPP middle school (grade 12 for most students).

The student survey included multiple items capturing the same underlying construct we wished to measure in many cases, so we created indices that summarize students' responses on related data items. (See Appendix B for more detail.)

Analytic methods

The Same KIPP Middle School, Adjacent Cohorts analysis examines students at KIPP high schools in their first year of operation, with a treatment group of 8th-grade KIPP middle school students who had the option to attend the local KIPP high school in the first year of its operation, and a comparison group composed of students in the previous cohort of 8th-grade KIPP students from the same middle school who did not have the option to attend the local KIPP high school because it had not yet opened.

The Same Cohort, Matched Middle School analysis relied on a matched comparison group design that used "nearest neighbor" matching to identify a similar comparison student for each treatment student. The validity of our matched comparison group design depends on our ability to eliminate or minimize differences in key characteristics between students at KIPP middle schools with a KIPP high school option and KIPP middle school students in the comparison group who lacked such an option.[44] We used student-level data that included student characteristics and baseline test scores to identify a matched comparison group of students who are similar to treatment students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in middle school. By matching on baseline test scores, we accounted for achievement levels for students prior to entering high school.

After we identified the matched comparison groups for both analyses, we estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between treatment students and comparison students. Specifically, the impact estimates adjust for any differences between treatment students and the comparison group pertaining to demographic characteristics or students' prior math and reading test scores.

The combination of matching, either with adjacent cohorts or propensity-score matching, and OLS accounts for differences in observed baseline characteristics and achievement scores between treatment students and comparison students (in other words, the differences associated with having access to a KIPP high school). But it remains possible that treatment students and comparison students differ in unobserved ways that may affect later test scores. However, previous studies have suggested that quasi-experimental methods that are similar in some respects to the approach we use here can succeed in replicating experimental impact estimates in

[44] Specifically, to produce unbiased impact estimates the design must eliminate differences in student characteristics that could explain academic achievement outcomes and thus be confounded with the estimated treatment impact.

certain contexts (Cook et al. 2008; Bifulco 2012; Fortson et al. 2012; Furgeson et al. 2012; Tuttle et al. 2013).

The remainder of this appendix presents the additional details regarding the study's propensity score estimation model, matching procedures, and imputation model for baseline data.

Propensity Score Matching Procedures (Same Cohort, Matched Middle Schools Analysis)

The matching procedure for the Same Cohort, Matched Middle Schools analysis consists of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of treatment students in the sample; and (3) testing the balance of baseline characteristics between our treatment group and matched comparison group.

For the first step, we selected a group of comparison schools by matching potential comparison KIPP middle schools to the feeder KIPP middle schools, using aggregate student-level characteristics including race, ethnicity, and average achievement in grade 8, as well as state (where possible). We then separated the feeder KIPP middle schools into two groups based on the type of test administered in grade 8 (baseline) at that school: either the MAP or SAT-10. For each group, the pool of eligible comparison schools (and thus students) was limited to those administering the same baseline tests as the treatment students. We then performed an iterative propensity score estimation procedure on the data sets for each test group. The dependent variable in this propensity score model is an indicator of whether the student had access to a KIPP high school. Covariates in the model were selected using an iterative process that identifies the baseline demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. Table H.5 provides an exhaustive list of potential covariates for inclusion in each model.

At a minimum, we required the logistic model to include baseline test scores in both math and reading. The other covariates were iteratively included and tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

Table H.5. List of potential covariates for inclusion in the propensity score estimation model

Observed and imputed (when missing) math and reading baseline test scores from one year prior (always included)
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from one year prior
Set of math and reading imputation dummies indicating whether math and reading baseline test scores are imputed
Demographic variables (gender, race/ethnicity, and whether the student was old-for-grade)
Interactions of baseline test scores and all available demographic variables
Interactions of gender and race/ethnicity variables
Interactions of old-for-grade indicator and race/ethnicity variables

Next, we calculated propensity scores for having the option to attend a KIPP high school. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (with replacement) of comparison group students to treatment group students, separately for each baseline test type (that is, only matching treatment and comparison group members who took the same type of baseline test), from within the region of common support. In other words, for each treatment student with a propensity score that fell within the range of propensity scores found among potential matched comparison students, we identified the comparison student whose propensity score was closest to that of the treatment student.

We then tested the balance of the treatment group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, and old-for-grade). We required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no significant differences on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this covariate for the comparison group are not significantly different from the treatment group at the five percent level.[45] If the first round of matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that baseline test score group, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[46] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

[45] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table H.2-H.4, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups for any of the outcomes in this study.

[46] If balance was not achieved in the first round of matching for a given school, under our protocol we would remove the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20).

Imputation for missing baseline data for propensity score models

Here we explain in greater detail how our propensity score models handled missing data when students were missing baseline test score data.[47] Our propensity score matching procedure used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores; imputation was completed separately by site. Variables in the imputation model included baseline math and reading test scores, demographic characteristics (gender, race/ethnicity, and whether the student was old for grade). Treatment status not part of the imputation model because imputation is performed separately for the treatment group and then the comparison group.

We first estimated the imputation model using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student.

While we use these imputed baseline test scores in our propensity score matching model, none of the imputed values were included in our benchmark impact analysis or tests of baseline equivalence discussed earlier in this appendix. In these cases, students missing data on a baseline test score were simply treated as being missing from the sample.

Impact model and covariates

To obtain impact estimates for the two analysis samples, as well as for both samples combined, we estimated an ordinary least squares (OLS) regression model for each examined outcome (test scores in reading, language, and math, as well as the survey-based outcomes discussed in Appendix B). The models were run on the pooled sample of schools, separately for the two different analysis samples, as well as in a pooled model combining both analysis samples. The model incorporated baseline (8th grade) demographic controls including indicators for gender, race/ethnicity, and whether the student was old for grade[48]; and baseline mathematics and reading test scores (8th grade). See Table H.6 for a full list of these covariates. The basic form of the model is defined in equation H1:

$$\text{(H1)} \qquad y_{it} = \alpha + X_i\beta + \delta treat_{it} + indicator_variables + \varepsilon_{it}$$

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and baseline test scores) of student i; $treat_{it}$ is a binary variables for treatment status indicating whether student i had access to a KIPP high school. The model also include a set of dummy indicator variables, indicating either the KIPP middle school attended (for the Same KIPP Middle School, Adjacent Cohorts model only), or the

[47] Imputed baseline test scores were used for the propensity score matching procedure. We tested whether the impact results changed when imputed baseline test scores were included as covariates, and the results were the same under both specifications.

[48] The cutoff for when a student was considered old for grade varied as was calculated based on the admission cutoff date for each location.

baseline test score type (for the Same Cohort, Matched Middle School group), and interactions of those indicators with the baseline test score variables. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ and β are parameters or vectors of parameters to be estimated. The estimated coefficient on the treatment indicator, δ represents the impact of access to a KIPP high school on the observed outcome. Robust standard errors were clustered at the student level.

We used the model to separately estimate the impact for each KIPP high school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP high schools in the sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

Table H.6. List of covariates included in OLS model

Math baseline test score
Reading baseline test score
Gender indicator variable
Set of race/ethnicity indicator variables
Old-for-grade indicator variable
Set of math and reading dummies indicating which type of test was used for the baseline test score, and interactions of those indicators with baseline math and reading test scores

APPENDIX I

DETAILED TABLES FOR WHAT WORKS CLEARINGHOUSE REVIEW

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

APPENDIX I.1 LOTTERY-BASED RANDOMIZED CONTROLLED TRIAL IMPACTS

In this section, we provide detailed tables to facilitate the review of the lottery-based academic achievement impacts following WWC standards and procedures. We do not include tables for exploratory outcomes on student behavior and attitudes.

Lottery-based impacts of KIPP elementary schools

In Table I.1, we provide the number of students randomly assigned to the intervention and comparison groups and the number of students with data on each outcome. In Table I.2 we present our benchmark, intent-to-treat (ITT) impact estimates for the elementary school sample. Due to the nature of the study design, we are unable to demonstrate the baseline equivalence of the study groups on mathematics or reading achievement.

Table I.1. Elementary school sample sizes (full randomized sample and analytic samples)

	Full randomized sample		Analytic sample	
Sample definition	Intervention	Comparison	Intervention	Comparison
Woodcock-Johnson III Calculation	290	334	176	195
Woodcock-Johnson III Applied Problems	473	624	282	370
Woodcock-Johnson III Letter-Word Identification	473	624	281	370
Woodcock-Johnson III Passage Comprehension	473	624	280	368

Notes: At schools where the sample randomly assigned by lottery had a treatment-control group size ratio greater than 2:1 (or less than 1:2), a random subsample was drawn from the students randomly assigned by admissions lottery to receive an admission offer or not. This random subsample is reflected in the "full randomized sample" columns. Analytic sample columns reflect the number of students with non-missing outcome scores for each outcome. The full randomized sample and analytic sample for the Calculation test are smaller because that test was administered only to second graders, and only five of the eight study school samples were comprised of students who entered in kindergarten and were thus in second grade in the third year, when the test was administered. (Sample members in the remaining three schools entered admissions lotteries for age 3 pre-kindergarten, or PK3, and were in kindergarten in the third year.)

Table I.2. Impact of offer of admission to KIPP elementary school (ITT)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Woodcock-Johnson III Calculation	Mathematics achievement	0.482	176	0.200	195	0.282**	0.009
Woodcock-Johnson III Applied Problems	Mathematics achievement	0.044	282	-0.030	370	0.074	0.156
Woodcock-Johnson III Letter-Word Identification	Alphabetics	1.005	281	0.757	370	0.248**	<0.001
Woodcock-Johnson III Passage Comprehension	Reading comprehension	0.185	280	-0.038	368	0.223**	0.001

Notes: Test scores are from Woodcock-Johnson subtests administered by the study in spring of the third year following admissions lotteries. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using information from the Woodcock-Johnson national norming sample. Impacts and group means are presented in z-score units. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact, which is estimated controlling for baseline characteristics. Impacts on achievement are estimated using the sample with non-missing outcome data for each measure, but missing baseline characteristics information is imputed using multiple imputation by chained equations. Imputation is conducted separately by treatment group, using standard commands in Stata to generate 20 imputed datasets, calculate and combine impact estimates, and adjust robust standard errors accounting for the number of imputations and the variability of estimated impacts across imputations. The imputation model includes all baseline characteristics that the impact model controls for. The student is the unit of assignment and unit of analysis.

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

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

Lottery-based impacts of KIPP middle schools

In Table I.3, we provide the number of students randomly assigned to the intervention and comparison groups and the number of students with outcome data for each outcome. In Table I.4 we present our benchmark, intent-to-treat (ITT) impact estimates for the middle school sample.

Tables I.3-I.4 are based on samples including imputed data. In Tables I.5 and I.6, we present information on the middle school samples using only the sample of students with non-imputed data. The analysis with non-imputed data is supplemental—our preferred estimates are those presented in table I.4.

Table I.3. Middle school sample sizes (full randomized sample and analytic samples)

	Full randomized sample		Analytic sample	
Sample definition	Intervention	Comparison	Intervention	Comparison
Math, Year 1	436	421	313	294
Math, Year 2	436	421	287	268
Math, Year 3	386	369	233	222
Reading, Year 1	436	421	314	294
Reading, Year 2	436	421	291	272
Reading, Year 3	386	369	234	224

Notes: At schools where the sample randomly assigned by lottery had a treatment-control group size ratio greater than 2:1 (or less than 1:2), a random subsample was drawn from the students randomly assigned by admissions lottery to receive an admission offer or not. This random subsample is reflected in the "full randomized sample" columns. Analytic sample columns reflect the number of students with non-missing outcome scores for each outcome. Randomized sample sizes and analysis sample sizes are smaller for year three because one of the sample schools did not provide any outcome data in the third year. Since admissions lotteries randomly assign students to intervention or comparison conditions at the student level, this school is removed from both the full randomized and analytic samples for year three.

Table I.4. Impact of offer of admission to KIPP middle school (ITT)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Math, Year 1	Mathematics achievement	-0.120	313	-0.217	294	0.097*	0.046
Math, Year 2	Mathematics achievement	-0.002	287	-0.245	268	0.244**	<0.001
Math, Year 3	Mathematics achievement	0.010	233	-0.165	222	0.176**	0.008
Reading, Year 1	General literacy achievement	-0.233	314	-0.260	294	0.026	0.584
Reading, Year 2	General literacy achievement	-0.156	291	-0.337	272	0.181**	<0.001
Reading, Year 3	General literacy achievement	-0.134	234	-0.277	224	0.143*	0.014

Notes: Test scores are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using statewide means and standard deviations provided in assessment technical documentation. Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Impacts and group means are presented in z-score units. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact, which is estimated controlling for baseline characteristics. Impacts on achievement are estimated using the sample with non-missing outcome data for each measure, but missing baseline characteristics information is imputed using multiple imputation by chained equations. Imputation is conducted separately by treatment group, using standard commands in Stata to generate 20 imputed datasets, calculate and combine impact estimates, and adjust robust standard errors accounting for the number of imputations and the variability of estimates across imputations. The imputation model includes all baseline characteristics that the impact model controls for. The student is the unit of assignment and unit of analysis.

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

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

Table I.5. Baseline equivalence for lottery-based middle school impacts (sample with non-imputed data)

	Intervention group		Comparison group			
Baseline measure, outcome year	Mean	Sample size	Mean	Sample size	Difference	p-value
Math Pretest, Year 1	-0.091	266	-0.183	263	0.092	0.239
Math Pretest, Year 2	-0.032	242	-0.190	237	0.158*	0.050
Math Pretest, Year 3	-0.043	194	-0.154	192	0.111	0.201
Reading Pretest, Year 1	-0.165	268	-0.243	263	0.078	0.281
Reading Pretest, Year 2	-0.132	246	-0.293	242	0.161*	0.031
Reading Pretest, Year 3	-0.112	196	-0.266	194	0.154	0.067

Notes: Test scores are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using statewide means and standard deviations provided in assessment technical documentation. Differences between treatment and control groups, and group means, are presented in z-score units. The comparison group mean is the unadjusted mean pretest z-score in the comparison group, while the treatment group mean is equal to the comparison group mean plus the difference between groups. The model used to estimate the difference between groups includes school and grade indicators so that intervention-comparison differences are measured between students in the same school and grade. Baseline equivalence is assessed separately for the analysis sample for each outcome year and subject, so that the sample used to assess baseline equivalence for a given outcome year and subject is identical to the sample used to estimate the impact for that outcome year and subject. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in two "Mean" columns.

* 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 I.6. Impact of KIPP middle schools (lottery-based estimates for sample with non-imputed data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Math, Year 1	Mathematics achievement	-0.091	266	-0.212	263	0.121*	0.016
Math, Year 2	Mathematics achievement	-0.003	242	-0.260	237	0.257**	<0.001
Math, Year 3	Mathematics achievement	-0.046	194	-0.161	192	0.115	0.081
Reading, Year 1	General literacy achievement	-0.249	268	-0.261	263	0.012	0.822
Reading, Year 2	General literacy achievement	-0.148	246	-0.351	242	0.204**	<0.001
Reading, Year 3	General literacy achievement	-0.163	196	-0.290	194	0.127	0.053

Notes: Test scores are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using statewide means and standard deviations provided in assessment technical documentation. Impacts and group means are presented in z-score units. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact, which is estimated controlling for a pre-test (taken in the spring before the lottery) in the same subject as the outcome. Impacts on achievement in each outcome year after the lottery are estimated using the sample of students with a non-missing outcome score for that subject and year, and a non-missing pre-test score from the baseline year in that subject. All regressions use robust standards errors. The student is the unit of assignment and unit of analysis.

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

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

APPENDIX I.2 MATCHED STUDENT IMPACTS USING NON-IMPUTED SAMPLES

In this section, we provide detailed tables to facilitate the review of the matched-student academic achievement and attainment impacts following WWC standards and procedures.

Matched-student impacts of KIPP middle schools (supplemental analysis)

In this section we present supplemental information on the quasi-experimental matched-student estimates of the impact of KIPP middle schools. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated only on the sample of students with non-imputed data. In Table I.7, we present baseline equivalence information for the sample of students in the matched-student analysis with non-missing baseline and outcome data. In Table I.8, we present estimates of the impact of KIPP middle schools using the same sample. In Table I.9, we present baseline equivalence results for the sub-sample of KIPP middle schools that opened in fall 2011 or later (new middle schools). In Table I.10, we present estimates of the impact of KIPP middle schools using the same sample of new middle schools.

Table I.7. Baseline equivalence for matched-student middle school impacts (sample with non-imputed data)

	Intervention group		Comparison group			
Baseline measure (outcome sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (reading year 1)	-0.102	17,518	-0.084	17,397	-0.018	0.101
Reading scores (reading year 2)	-0.087	14,079	-0.085	13,679	-0.002	0.897
Reading scores (reading year 3)	-0.049	11,318	-0.045	10,837	-0.004	0.761
Reading scores (reading year 4)	-0.007	7,430	-0.035	7,121	0.027	0.149
Math scores (math year 1)	-0.095	17,525	-0.082	17,413	-0.013	0.244
Math scores (math year 2)	-0.070	14,064	-0.082	13,672	0.012	0.413
Math scores (math year 3)	-0.043	11,185	-0.060	10,741	0.017	0.259
Math scores (math year 4)	-0.065	6,887	-0.083	6,737	0.018	0.386
Reading scores (history)	0.022	4,866	-0.013	4,896	0.034	0.112
Math scores (science)	-0.053	8,826	-0.059	8,587	0.006	0.704

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Because of rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in two “Mean” columns. Data shown in this table do not include imputed values for any baseline variables.

Table I.8. Impact of KIPP middle schools (matched-student estimates for sample with non-imputed data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading year 1	General literacy achievement	-0.105	17,518	-0.111	17,397	0.005	0.426
Reading year 2	General literacy achievement	-0.008	14,079	-0.113	13,679	0.105**	<0.001
Reading year 3	General literacy achievement	0.062	11,318	-0.092	10,837	0.154**	<0.001
Reading year 4	General literacy achievement	0.076	7430	-0.085	7,121	0.161**	<0.001
Math year 1	Mathematics achievement	-0.051	17,525	-0.109	17,413	0.057**	<0.001
Math year 2	Mathematics achievement	0.091	14,064	-0.141	13,672	0.232**	<0.001
Math year 3	Mathematics achievement	0.170	11,185	-0.121	10,741	0.291**	<0.001
Math year 4	Mathematics achievement	0.136	6,887	-0.131	6,737	0.268**	<0.001
History		0.107	4,866	-0.131	4,896	0.238**	<0.001
Science	Science achievement	0.084	8,826	-0.166	8,587	0.249**	<0.001

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. Reported impacts are an average of equally weighted impact estimates for each KIPP middle school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. Impacts represent the cumulative effect of KIPP after the noted number of years after admission for math and reading scores, not the marginal effect of an additional year. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. The grade level of middle school exams used for history and science outcomes varied by jurisdiction. We selected the highest middle school grade level where science or social studies was observed for more than one cohort of KIPP students. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Table I.9. Baseline equivalence for matched-student middle school impacts (sample with non-imputed data) for new KIPP middle schools

	Intervention group		Comparison group			
Baseline measure (outcome sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (reading year 1)	-0.323	1,195	-0.238	1,165	-0.085*	0.021
Reading scores (reading year 2)	-0.301	586	-0.242	532	-0.059	0.289
Math scores (math year 1)	-0.272	1,196	-0.195	1,170	-0.078*	0.036
Math scores (math year 2)	-0.225	587	-0.223	536	-0.002	0.973

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in two "Mean" columns. Data shown in this table do not include imputed values for any baseline variables.

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

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

Table I.10. Impact of new KIPP middle schools (matched-student estimates for sample with non-imputed data)

	Intervention group		Comparison group			
Outcome measure	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading year 1	-0.220	1,195	-0.269	1,165	0.049*	0.028
Reading year 2	-0.123	586	-0.243	532	0.120**	0.000
Math year 1	-0.188	1,196	-0.228	1,170	0.039	0.069
Math year 2	-0.063	587	-0.289	536	0.226**	0.000

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. Reported impacts are an average of equally weighted impact estimates for each KIPP middle school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. Impacts represent the cumulative effect of KIPP after the noted number of years after admission for math and reading scores, not the marginal effect of an additional year. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Matched-student impacts of KIPP high School on new entrants in Grade 9 (supplemental analysis)

In this section we present supplemental information on the quasi-experimental matched-student estimates of the impact of KIPP high schools for new entrants in grade 9. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated only on the sample of students with non-imputed data. In Table I.11 we present baseline equivalence information for the sample of students in the matched-student analysis with non-missing baseline and outcome data. In Table I.12, we present estimates of the impact of KIPP high schools using the same sample.

Table I.11. Baseline equivalence for matched-student high school impacts (sample with non-imputed baseline data)

	Intervention group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	-0.060	887	-0.089	861	0.028	0.591
Math scores (math)	-0.084	725	-0.059	691	-0.026	0.631
Reading scores (social studies)	0.000	304	0.001	297	-0.002	0.982
Math scores (science)	-0.111	656	-0.109	643	-0.002	0.973
Reading scores (4-year graduation)	-0.014	426	-0.125	426	0.111	0.099
Math scores (4-year graduation)	-0.028	426	-0.155	426	0.127	0.088
Free and reduced-price lunch status (4-year graduation)	0.806	426	0.809	426	-0.003	0.924

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Because of rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in two “Mean” columns. Data shown in this table do not include imputed values for any baseline variables.

Table I.12. Impact of KIPP high schools for new entrants in grade 9 (matched-student estimates for sample with non-imputed baseline data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
ELA achievement	General literacy achievement	0.111	887	-0.065	861	0.175**	<0.001
Mathematics achievement	Mathematics achievement	0.236	725	-0.036	691	0.273**	<0.001
Science achievement	Science achievement	0.105	656	-0.223	643	0.328**	<0.001
Social studies achievement	Social studies achievement	-0.134	304	-0.149	297	0.015	0.795
Four-year high school graduation	Completing school	0.707	426	0.672	426	0.035	0.355

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Matched-student cumulative impacts of KIPP middle and high schools (supplemental analysis)

In this section we present supplemental information on the quasi-experimental matched-student estimates of the cumulative impact of KIPP middle and high schools. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated only on the sample of students with non-imputed data. In Table I.13 we present baseline equivalence information for the sample of students in the matched-student analysis with non-missing baseline and outcome data. In Table I.14, we present estimates of the cumulative impact of KIPP middle and high schools using the same sample.

Table I.13. Baseline equivalence for matched-student cumulative middle and high school impacts (sample with non-imputed baseline data)

	Intervention group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	0.065	2,072	0.100	1,929	-0.035	0.261
Math scores (math)	-0.032	1,516	0.037	1,414	-0.069	0.083
Reading scores (social studies)	0.117	788	0.109	707	0.007	0.871
Math scores (science)	0.065	1,946	0.133	1,636	-0.067	0.078
Reading scores (4-year graduation)	0.037	1,025	0.055	1,008	-0.018	0.658
Math scores (4-year graduation)	0.019	1,025	0.080	1,008	-0.061	0.138
Free and reduced-price lunch status (4-year graduation)	0.883	1,025	0.892	1,008	-0.009	0.562

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Because of rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in two “Mean” columns. Data shown in this table do not include imputed values for any baseline variables.

Table I.14. Cumulative impact of KIPP middle and high schools (matched-student estimates for sample with non-imputed baseline data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
ELA achievement	General literacy achievement	0.381	2,072	0.085	1,929	0.295**	<0.001
Mathematics achievement	Mathematics achievement	0.341	1,516	-0.001	1,414	0.341**	<0.001
Science achievement	Science achievement	0.417	788	0.001	707	0.417**	<0.001
Social studies achievement	Social studies achievement	0.180	1,946	-0.087	1,636	0.267**	<0.001
Four-year high school graduation	Completing school	0.785	1,025	0.651	1,008	0.134**	<0.001

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

APPENDIX I.3 MATCHED-SCHOOL IMPACTS

In this section, we provide detailed tables to facilitate the review of the matched-school academic achievement impacts following WWC standards and procedures. We do not provide tables for exploratory outcomes on student behavior and attitudes. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated separately for the sample of schools in our adjacent cohorts and matched-middle schools models. Both models compare outcomes for KIPP middle school students who had an option to attend a KIPP high school with those who did not have an option to attend a KIPP high school.

Matched-school impacts of opportunity to attend a KIPP high school (adjacent cohorts model)

The first model (Same KIPP Middle School, Adjacent Cohorts), focuses on a set of KIPP high schools in their first year of operation. The treatment group in this model includes 8th-grade KIPP middle school students who had the option to attend the local KIPP high school in the first year of its operation. The comparison group includes the previous cohort of 8th-grade KIPP students from the same middle school who did not have the option to attend the local KIPP high school because it had not yet opened. Because sample sizes are too small to restrict the comparison group, we do not employ student-level matching in this model. In Table I.15 we present baseline equivalence information for the sample of students in the matched-school analysis for the adjacent cohorts sample with non-missing baseline and outcome data. In Table I.16, we present estimates of the impact of an opportunity to attend a KIPP high school using the same sample.

Table I.15. Baseline equivalence for matched-school impacts of opportunity to attend a KIPP high school (adjacent cohorts sample)

Baseline Characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (reading achievement, language achievement)	0.425	0.357	0.068	0.409	208	213
Math scores (mathematics achievement)	0.772	0.800	-0.028	0.718	208	213

Note: Baseline tests are either the SAT-10 or the MAP, depending on what was offered in that jurisdiction. They are presented in z-score units. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution. The relevant outcome sample for each baseline characteristic is noted in parentheses. All values in this table are based on the sample for which we have test outcome data. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

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

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

Table I.16. Impact of opportunity to attend a KIPP high school (adjacent cohort estimates)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading achievement	Reading comprehension	0.220	208	0.230	213	-0.010	0.844
Language achievement	General literacy achievement	0.081	208	0.081	213	0.000	0.995
Mathematics achievement	Mathematics achievement	0.026	208	-0.056	213	0.082	0.217

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are study-administered TerraNova assessments (Form G, Level 21/22). Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. The student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Matched-school impacts of opportunity to attend a KIPP high school (matched middle schools model)

In the second model (Same Cohort, Matched Middle Schools), the treatment group includes 8th-grade KIPP middle school students in 2008-09 who had the option to attend the local KIPP high school. The comparison group includes 8th-grade KIPP students from different middle schools in 2008-09 in regions with no KIPP high school open at the time. To define a sample that was equivalent at baseline (grade 8), we first identified a set of comparison KIPP middle schools that most resembled the feeder KIPP middle schools on the basis of average school-level characteristics (race/ethnicity, baseline achievement on a nationally-normed test, and baseline test instrument—either the SAT-10 or the MAP). Then, within these matched sets of schools, we conducted student-level propensity score matching to identify the individual comparison student who was the closest match to each treatment student on the basis of gender, race/ethnicity, baseline achievement in reading and math, and whether the student was old for his/her grade. In Table I.17 we present baseline equivalence information for the sample of students in the matched-school analysis for the matched middle schools sample with non-missing baseline and outcome data. In Table I.18, we present estimates of the impact of an opportunity to attend a KIPP high school using the same sample.

Table I.17. Baseline equivalence for matched-school impact of opportunity to attend a KIPP high school (matched middle schools sample)

Baseline Characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (reading achievement, language achievement)	0.131	0.140	-0.009	0.926	256	256
Math scores (mathematics achievement)	0.568	0.561	0.007	0.955	256	256

Note: Baseline tests are either the SAT-10 or the MAP, depending on what was offered in that jurisdiction. They are presented in z-score units. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution. The relevant outcome sample for each baseline characteristic is noted in parentheses. All values in this table are based on the sample for which we have test outcome data. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

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

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

Table I.18. Impact of opportunity to attend a KIPP high school (matched-middle school estimates)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading achievement	Reading comprehension	0.255	256	0.095	256	0.160*	0.026
Language achievement	General literacy achievement	0.069	256	-0.054	256	0.123	0.152
Mathematics achievement	Mathematics achievement	0.071	256	-0.065	256	0.136	0.099

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are study-administered TerraNova assessments (Form G, Level 21/22). Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. The student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

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

www.mathematica-mpr.com

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

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

MATHEMATICA
Policy Research

Mathematica® is a registered trademark
of Mathematica Policy Research, Inc.

MATHEMATICA Policy Research

REPORT

FINAL REPORT

Pre-Kindergarten Impacts Over Time: An Analysis of KIPP Charter Schools

August 22, 2017

Virginia Knechtel

Thomas Coen

Pia Caronongan

Nickie Fung

Lisbeth Goble

Submitted to:

Laura and John Arnold Foundation

2800 Post Oak Blvd., Suite 225

Houston, TX 77056-6169

Project Officer: Meredith McPhail, Research Integrity Manager

Submitted by:

Mathematica Policy Research

1100 1st Street, NE

12th Floor

Washington, DC 20002-4221

Telephone: (202) 484-9220

Facsimile: (202) 863-1763

Project Director: Virginia Knechtel

Reference Number: 50247.510

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

ACKNOWLEDGMENTS

This report was commissioned by the Laura and John Arnold Foundation. We thank Meredith McPhail, our project officer, and Stuart Buck and Josh McGee for their input on the evaluation plan and support of this work.

This study was made possible by the efforts of many individuals. We sincerely appreciate the willingness of students and their parents to participate in the evaluation. We are grateful to staff at the KIPP Foundation, particularly Danielle Eisenberg, for supporting the research and providing administrative data to supplement the report. We also deeply appreciate the assistance of staff at the individual KIPP schools that were the focus of the report, who provided important contextual information about the operation of KIPP pre-kindergarten during the study period.

This report would not have been possible without contributions from many other people at Mathematica. Primary data collection efforts were supported by Brittany Vas, Kevin Manbodh, Kim Mook, Tammy Chen, and Cailean Geary; staff at Mathematica's Princeton Survey Operations Center, led by Ellen Siegel, Irene Crawley, and Michael Smith; and Information Services staff, led by John Carsley. Cailean Geary also provided assistance with project management. Sarah Fisher assisted with programming for the report. Juha Sohlberg led the review of the programming code. Phil Gleason and Gregory Chojnacki provided important feedback on our analytic approach. Christina Clark Tuttle provided critical technical review and feedback. Jill Miller led the production of the report, which was edited by Molly and Jim Cameron.

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

CONTENTS

EXECUTIVE SUMMARY xi

I BACKGROUND 1

A. The KIPP network of schools 1

B. KIPP has consistently demonstrated positive impacts in previous research 1

C. Pre-kindergarten at KIPP 2

D. Pre-kindergarten produces positive impacts, but they tend to fade out 3

E. The current study: Does KIPP pre-K produce more lasting impacts? 3

II RESEARCH APPROACH AND STUDY DESIGN 5

A. Study samples 7

B. Outcome measures and data collection 9

C. A note of caution 11

III IMPACTS OF KIPP PRE-K AND EARLY ELEMENTARY SCHOOL 13

A. After five years, KIPP pre-K and KIPP early elementary has positive and statistically significant impacts on reading and math achievement (Research Question 1) 13

B. KIPP pre-K and KIPP early elementary may also have a positive impact on students' executive function (Research Question 1) 14

C. KIPP pre-K may provide an additional benefit for reading achievement above and beyond KIPP elementary school (Research Question 2) 16

D. The KIPP impact on Letter-Word Identification scores persists over time, but impacts on Passage Comprehension largely dissipate by grade 2 (Research Question 3) 17

IV KIPP PRE-KINDERGARTEN EXPERIENCE 21

A. Schools attended by students in the sample 21

B. Characteristics of KIPP pre-K during the study period 24

1. The structure of the schools supported alignment across school levels 25

2. KIPP pre-K programs were heavily focused on academics—particularly emphasizing foundational reading and math skills—during the study period 26

3. Curriculum and assessments were mostly teacher developed and contributed to alignment in instruction across grades 26

4. KIPP pre-K was designed to establish values and build a behavioral foundation for later success at KIPP 26

5. Supports for children and families varied across schools, but all schools heavily emphasized building relationships with students and their families 27

6. The training provided to staff varied considerably by school, but most teachers were relatively new to teaching and the pre-K grades 28

V IMPLICATIONS FOR POLICY AND PRACTICE 29

A. Factors that might contribute to the persistence of impacts of KIPP pre-K 29

B. Topics for future research 30

REFERENCES 33

APPENDIX: DETAILED ANALYTIC METHODS 37

TABLES

II.1 Characteristics of schools in the samples 7

II.2 Student sample sizes 7

II.3 Summary of outcome measures and testing periods 10

IV.1 Sample enrollment at KIPP 22

IV.2 Type of early childhood program attended by KIPP pre-K applicants 23

IV.3 Type of elementary school attended by KIPP pre-K applicants 23

IV.4 Features of KIPP pre-K programs (2012–2013 school year) 24

Appendix Tables:

1 Baseline equivalence of grade 2 analytic sample (pre-K cohort) 41

2 Baseline equivalence of grade 2 analytic sample (kindergarten cohort) 42

3 Baseline equivalence of longitudinal analytic sample (pre-K cohort) 43

4 Average test scores among students based on testing group, by experimental condition 44

5 Five-year impacts of offer of admission to KIPP elementary school in pre-K 48

6 Isolated effect of KIPP pre-K (exploratory analysis) 49

7 Reading impacts over time for KIPP pre-K cohort 49

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

FIGURES

II.1 Study design 6

II.2 Characteristics of students in the samples 9

III.1 KIPP academic impacts after five years 14

III.2 KIPP executive function impacts after five years 15

III.3 Isolated impact of KIPP pre-K in grade 2 16

III.4 Changes in Letter-Word Identification scores over time 18

III.5 Changes in Passage Comprehension scores over time 19

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

EXECUTIVE SUMMARY

Despite the well-documented short-term positive impacts of early childhood education participation—and specifically pre-kindergarten (pre-K)—on multiple outcome domains, numerous studies have also shown that the observed effects of pre-K decrease ("fade out") or disappear altogether over time (Lipsey et al. 2015; Puma et al. 2012). Some experts argue that consistently high quality pre-K, aligned with later educational experiences, might produce more lasting impacts on student achievement than the typical pre-K currently available.

KIPP, a national network of public charter schools, provides a possible model for high quality pre-K aligned with an elementary school educational program. In previous research, KIPP has consistently demonstrated positive impacts on student achievement, including at the elementary school level (Tuttle et al. 2015). As of fall 2016, 27 KIPP elementary schools served students in pre-K. KIPP pre-K exhibits several features experts suggest might lead it to produce more lasting impacts than more traditional pre-K programs. Specifically, because KIPP pre-K students tend to continue their education in a KIPP elementary school—typically at the same school or campus as their pre-K experience—there is an increased likelihood that their later educational experiences will be aligned with their pre-K experiences. With increased alignment, it is more likely that the knowledge and skills acquired from later experiences will build on those developed in pre-K, thus leading to larger and more persistent impacts. Additionally, KIPP pre-K meets several of the criteria widely perceived to represent a high quality pre-K experience.

In this report, we build on a previous study of KIPP elementary schools to estimate the impact of an offer of admission to a KIPP pre-K program and explore whether any impacts persist as students advance beyond kindergarten. We summarize our key findings below.

- **After five years, KIPP pre-K combined with KIPP early elementary school has positive and statistically significant impacts on reading and math achievement.** We capitalized on randomized lotteries for entry to three KIPP pre-K programs to produce experimental estimates of the impact of an offer of admission to a KIPP pre-K. We found statistically significant or substantively important impacts on three of four measures of reading and math achievement, ranging in size from 0.31 to 0.43 standard deviation units. These impacts were educationally meaningful; for example, the Letter-Word Identification (reading skills) impact is approximately equivalent to a student moving from the 66th to the 80th percentile.
- **KIPP pre-K combined with KIPP early elementary school may also have a positive impact on students' executive function.** We used the same experimental design to measure the impact of KIPP pre-K on students' executive function—the "skills that help [to] plan, focus attention, switch gears, and juggle multiple tasks"—five years after admissions lotteries (Center on the Developing Child at Harvard University 2011). Although most impacts were not statistically significant, we found some suggestive evidence that an offer of admission to a KIPP pre-K may enhance some executive function skills, which are widely believed to be related to students' long-term academic success. Specifically, an offer of admission to KIPP pre-K had a positive or substantively important impact on students' working memory and ability to follow simple instructions.
- **KIPP pre-K may provide an additional benefit for reading achievement above and beyond KIPP elementary school.** To isolate the impact of KIPP pre-K, we produced experimental estimates of the impact of an offer of admission to a KIPP elementary school

in kindergarten (a school without pre-K) and compared them to the impacts of an offer of admission to a KIPP elementary school in pre-K. We found that the magnitude of the impacts in reading were larger for the KIPP schools that did offer pre-K than for those that did not do so, although the differences were not statistically significant. There were no differences in math. These results are exploratory, but they provide preliminary evidence that earlier and longer exposure to KIPP improves reading outcomes.

- **The KIPP impact on reading skills persists over time, but impacts on reading comprehension largely dissipate by grade 2.** We restricted our sample to the set of pre-K students with test scores in both kindergarten (three years after they participated in lotteries for admission to a KIPP pre-K) and grade 2 (five years after the same lotteries). We compared the size of the impacts at the two follow-up points and found that students who won an offer of admission to a KIPP pre-K program continued to outperform their peers on the Letter-Word Identification test in grade 2, but their peers had mostly caught up on the Passage Comprehension (reading comprehension) test.

We also identified six key features of the KIPP pre-K programs in our sample, based on interviews with KIPP staff at study schools. These features may provide helpful context about what could be driving the differences in impacts between KIPP pre-K and non-KIPP programs.

1. **The structure of the schools supported alignment across school levels.** Specifically, shared leadership over and/or co-location of the pre-K and elementary grades may have created opportunities for continuity and alignment across grades, and allowed elementary-grade staff to build off students' pre-K experiences at KIPP.
2. **KIPP pre-K programs were heavily focused on academics—particularly emphasizing foundational reading and math skills—during the study period**. Staff ranked reading and math knowledge and skills among their highest priorities during the study period, and employed varied instructional strategies in their classrooms.
3. **Curriculum and assessments were mostly teacher developed and contributed to alignment in instruction across grades.** Staff designed their own materials to instill the knowledge and skills required to be successful in later grades. They also helped to develop assessments used to measure progress toward this objective.
4. **KIPP pre-K was designed to establish values and build a behavioral foundation for later success at KIPP.** These values and behavioral expectations were taught explicitly and reinforced through relationships developed at the school.
5. **Supports for children and families varied across schools, but all schools heavily emphasized building relationships with students and their families.** Two programs provided robust child and family services during the study period. All schools in our sample placed a heavy emphasis on building strong relationships with students and their families.
6. **The training provided to staff varied considerably by school, but most teachers were relatively new to teaching and the pre-K grades.** The teachers in these programs were new to teaching and came from a variety of backgrounds. Administrators or instructional coaches in two programs regularly observed teachers and provided coaching or feedback.

Our findings support the growing consensus about effective pre-K programs and factors that help sustain their benefits and shed light on features that may merit replication and future study.

I. BACKGROUND

A. The KIPP network of schools

KIPP is a national network of public charter schools comprising 200 elementary, middle, and high schools in the 2016–2017 school year and serving 80,000 students.[1] KIPP schools serve a predominantly low-income and minority population; 88 percent of KIPP students are eligible for free or reduced-price lunches, and 96 percent are African American or Latino.[2]

KIPP schools emphasize rigorous academics and character instruction, with the ultimate goal of preparing students to succeed in college and beyond. The KIPP Approach is distinguished by five key principles that evolved from the Five Pillars, a set of operating principles that have historically guided KIPP schools (Text Box):[3]

The KIPP Approach

- **High expectations**: A culture of support and achievement and personalized learning based on a student's needs, skills, and interests.
- **Focus on character**: A belief that KIPP students need both a strong academic foundation and well-developed character strengths to succeed in college and the world beyond.
- **Highly effective teachers & leaders:** An emphasis on empowering educators to lead school teams and investment in training to help them grow as professionals.
- **Safe, structured, & nurturing environments:** Schools that are safe, structured, and nurturing environment so that KIPP students thrive and maximize their learning.
- **KIPP through college:** Counselors that support students as they prepare for college and career, and navigate social, academic, and financial challenges while in college.

All 200 2016–2017 KIPP schools are public charter schools, and nearly all have been charter schools since they opened. Thus, KIPP schools have greater autonomy in setting their own policies than do most traditional public schools but are accountable to their authorizers for achieving satisfactory performance.

B. KIPP has consistently demonstrated positive impacts in previous research

Mathematica's 2015 study of KIPP's Investing in Innovation (i3) scale-up grant, which used both experimental and quasi-experimental methods, found positive and statistically significant impacts of KIPP on student achievement across the elementary, middle, and high school grade levels, although the positive impacts for high schools were limited to those students who entered

[1] KIPP: Results. Are We Serving the Children Who Need Us? Available at http://www.kipp.org/results/national/#question-1:-are-we-serving-the-children-who-need-us. Accessed April 19, 2017.

[2] KIPP: Results. Are We Serving the Children Who Need Us? Available at http://www.kipp.org/results/national/#question-1:-are-we-serving-the-children-who-need-us. Accessed April 19, 2017.

[3] This description is adapted from the KIPP's description of the approach, available at www.kipp.org/approach.

KIPP for the first time in high school. The large positive impacts on test scores were consistent with previous studies of KIPP (Angrist et al. 2010; Furgeson et al. 2012; Gleason et al. 2014; Lake et al. 2012; Tuttle et al. 2013; Woodworth et al. 2008). The study included the first rigorous estimates of the effects of KIPP elementary schools. The elementary school analysis exploited lotteries at oversubscribed KIPP schools—those with more applicants than available seats—to produce estimates of the impact of KIPP elementary schools.

Mathematica found that the KIPP elementary schools in the study sample produced positive and statistically significant impacts on three measures of students' reading and mathematics skills after three years. On tests administered three years after entry, being offered admission to a KIPP elementary school led to an increase of 0.25 standard deviation units on the Letter-Word Identification test and 0.22 on the Passage Comprehension test in reading. In math, an offer of admission led to an increase of 0.28 standard deviation units on the Calculation test.

Evidence of KIPP's impact on other outcomes is less clear. The KIPP i3 study found that KIPP elementary and middle schools had positive impacts on school satisfaction, particularly among parents, and that KIPP high schools had positive impacts on several aspects of college preparation (Tuttle et al. 2015). At all three grade levels, KIPP had few significant impacts on measures of motivation and engagement related to student self-control, academic motivation, academic confidence, grit, school engagement, or effort in school. The study also found that KIPP had no impacts on student behavior at the elementary and middle school levels—the only levels at which Mathematica examined these outcomes.

C. Pre-kindergarten at KIPP

The first KIPP schools, which opened in 1994, were middle schools. In partnership with Doris and Don Fisher, founders of Gap Inc., the KIPP co-founders established the KIPP Foundation to support the expansion of the KIPP network in 2000. In 2004, the KIPP network began serving elementary grades. As of fall 2016, the KIPP network comprised 80 elementary schools, including 27 that served students in pre-kindergarten (pre-K).[4]

The decision to serve the pre-K grades at KIPP is a local one contingent on available resources. According to its regional staff, the push to serve students in the pre-K grades stemmed from a belief that an earlier start at KIPP better prepared students for academic success and facilitated KIPP's ultimate goal of supporting students to enroll and be successful in college. They reported that they viewed early exposure to literacy, language, and school behavioral or cultural expectations as particularly valuable to students who might otherwise enter school behind their peers—those who spoke English as a second language, were from homes not rich in print or language, and for whom pre-K was their first experience of spending time outside of the home. However, the availability of funding for these grade levels was also critical to allow KIPP schools in these cities to serve students in the pre-K grades.

[4] This number includes nine schools providing transitional kindergarten (California) and one school that serves children ages 6 months to 6 years, which is a partnership between KIPP Columbus and the YMCA of Central Ohio.

D. Pre-kindergarten produces positive impacts, but they tend to fade out

Early childhood education has been promoted at the state and local levels as a promising approach to increasing student achievement and school readiness (Flowers 2016). Participation in high quality early childhood education has been linked to improved outcomes across multiple developmental domains (Yoshikawa et al. 2013). Evidence of benefits for children from low-income families is particularly strong (Schanzenbach and Cascio 2013). Positive results have been found in evaluations of model programs (Elango et al. 2016) as well as state pre-K programs. Children who attended state pre-K have been found to have higher scores in math, receptive vocabulary, and early literacy (Gormley et al. 2008; Weiland and Yoshikawa 2013; Wong et al. 2008). In addition, one study found positive impacts on executive function, or "skills that help us plan, focus attention, switch gears, and juggle multiple tasks" (Center on the Developing Child at Harvard University 2011; Weiland and Yoshikawa 2013).

Despite the well-documented short-term impacts of early childhood education participation, and pre-K in particular, numerous studies have also shown that the observed effects of pre-K decrease ("fade out") or disappear altogether over time. For example, a randomized control trial of Tennessee's state pre-K program showed positive impacts at the beginning of kindergarten, but those impacts started to fade by the end of the kindergarten year (Lipsey et al. 2015). Evidence of fade out also has been documented elsewhere; for example, an experimental study of Head Start found positive impacts at the end of the Head Start year; by grade 3, however, there were no detectable differences (Puma et al. 2012). Given the extensive resources required to provide early childhood education on a large scale and the lack of evidence to date that the effects from pre-K are sustained over time, some observers have questioned whether the investment in these experiences is worthwhile.

E. The current study: Does KIPP pre-K produce more lasting impacts?

Some experts argue that consistently high quality pre-K, aligned with later educational experiences, might produce more lasting impacts on student achievement than the typical pre-K currently available. The theory is that the effects of pre-K programs may be better sustained if they are of consistently high quality—not simply in program features, such as staff qualifications and teacher-child ratios, but also in other characteristics, such as responsive teacher-child interactions and targeted developmentally appropriate learning activities (Yoshikawa et al. 2013). Another potential strategy is to better align instructional approaches and goals used in pre-K and the early elementary grades (Stipek et al. 2017; U.S. Department of Education 2016b).

KIPP pre-K presents an opportunity to extend the knowledge base about whether high quality and better-aligned pre-K may produce more lasting academic outcomes. KIPP meets several of the criteria widely perceived to represent a high quality pre-K experience, including staffing by well-educated teachers, low teacher-child ratios, and the use of developmentally appropriate learning activities. In addition, because KIPP pre-K students tend to continue their education in a KIPP elementary school—typically at the same school or campus as their pre-K experience—there is an increased likelihood that their later educational experiences will be aligned with their pre-K experiences. With increased alignment, it is more likely that the knowledge and skills acquired from later educational experiences will build on those developed in pre-K, thus leading to larger and more persistent impacts.

In this study, we build on the KIPP i3 Evaluation design to produce suggestive evidence about the magnitude of KIPP pre-K impacts and whether they persist over time. The evaluation addresses three research questions related to the impacts of KIPP pre-K and their persistence:

1. What is the cumulative impact of KIPP pre-K and KIPP elementary school on student outcomes measured in grade 2? (Research Question 1)
2. In grade 2, is the cumulative impact of KIPP, including pre-K, larger than the cumulative impact of KIPP without pre-K? (Research Question 2)
3. How does the size of any impacts of KIPP change over time for students who attended KIPP pre-K? (Research Question 3)

In the next chapter, we provide an overview of the research designs, data, and samples of schools and students that we employed to answer each research question. In Chapter III, we discuss findings related to these three questions. In Chapter IV, we contextualize our findings, using data from qualitative interviews with KIPP staff to describe the characteristics of KIPP pre-K. In Chapter V, we present our conclusions and discuss the implications of our findings for policy and practice.

II. RESEARCH APPROACH AND STUDY DESIGN

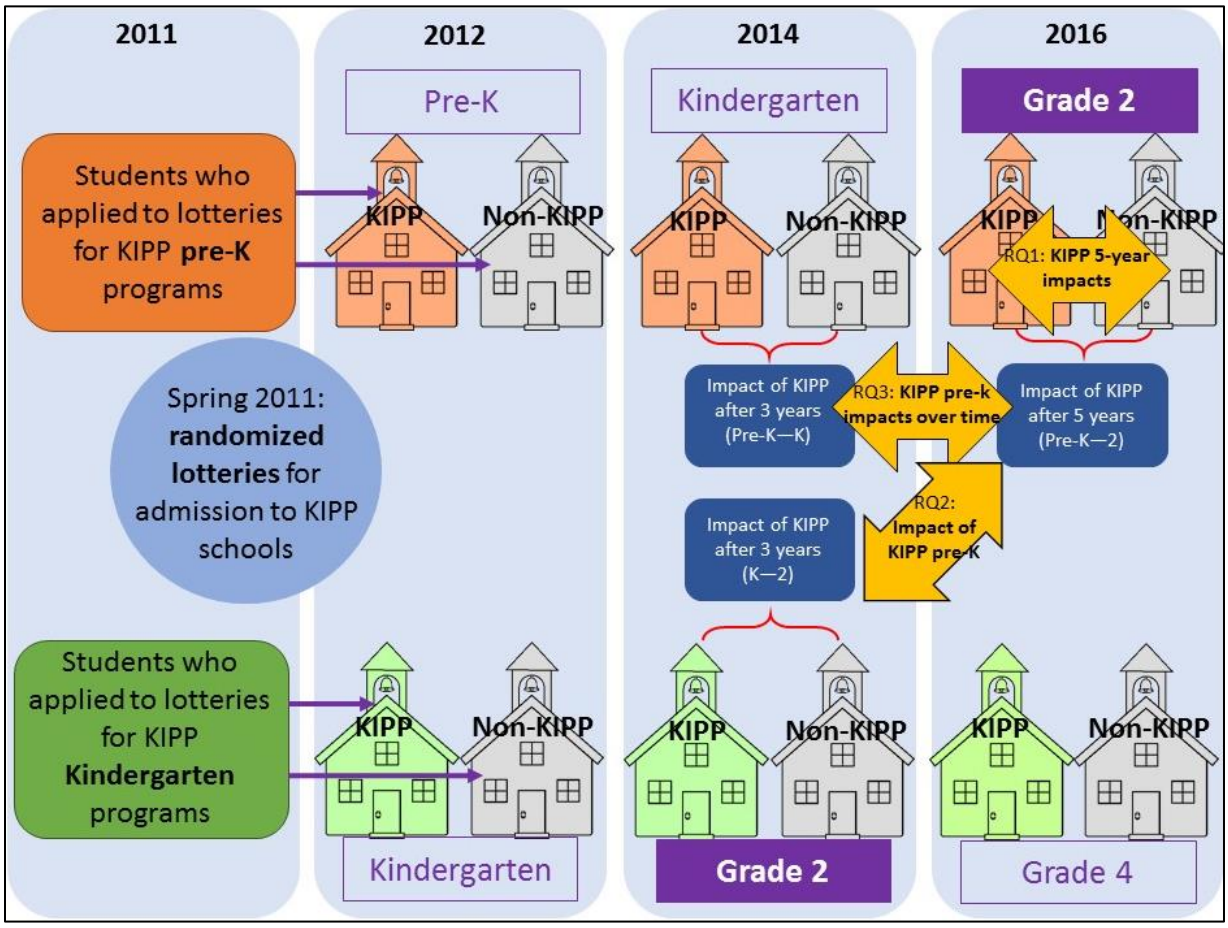
To address the research questions outlined in Chapter I, we built on the elementary school study from the KIPP i3 Evaluation. The i3 elementary study exploited randomized lotteries at oversubscribed KIPP schools—those with more applicants than available seats—to produce estimates of the impact of KIPP elementary schools. Schools conducted the lotteries in the spring and summer of 2011 for students enrolling at age 3 in three elementary schools offering pre-K and for students enrolling in five elementary schools starting in kindergarten for the 2011–2012 school year. The treatment group comprised students who participated in a lottery at either grade and won an offer of admission to a KIPP elementary school; the control group consisted of students who also participated in the lottery but did not receive an offer of admission. For the i3 evaluation, we compared the average impacts for students in the treatment and comparison groups after three years, combining students from both entry-grade levels in the estimates.

A well-executed randomized study design ensures that there are no systematic baseline differences between the treatment and control groups in both observable differences (such as academic achievement; family characteristics; and age, gender, and race) and unobservable characteristics (such as student motivation and perseverance). At the time of admission to KIPP, treatment and control group students were distinguishable only by the luck of their lottery draws; thus, any subsequent differences in their outcomes could be attributed to the impact of having the opportunity to attend a KIPP school.

Figure II.1 illustrates the approach we used to address each research question for the current study. The top panel follows the students who applied at age 3 for admission to a KIPP school that offered pre-K classes (“the pre-K cohort”); the bottom panel follows those who applied for admission beginning in kindergarten to a KIPP school that did not offer pre-K classes (“the kindergarten cohort”). For both cohorts, students were assigned either to the treatment condition (represented by orange [pre-K] or green [kindergarten] schools in the figure) or the control condition (represented by the grey schools in both panels). Moving from left to right, the figure shows the grade in which the typical sample student would be enrolled during the spring and summer of 2014 (when we collected outcome data for the i3 evaluation), and the spring and summer of 2016 (when we followed the pre-K cohort from the i3 evaluation and collected additional data for this study). The blue boxes represent the experimental impact estimates we calculated for the study; the yellow arrows highlight the contrast of interest for each research question.

To address Research Question 1, we focused on students in the pre-K cohort from the i3 evaluation (the top panel in Figure II.1). We compared grade 2 outcomes for the treatment students—those who won admission to (and predominantly attended) a KIPP pre-K—to grade 2 outcomes for the control students—those not offered admission to a KIPP pre-K. Because this comparison uses a random process to determine who can receive the KIPP pre-K treatment, it is the most rigorous way to measure its longer-term effect on student achievement relative to children’s other early education options.

Figure II.1. Study design



To isolate the effect of KIPP pre-K (Research Question 2), we tested whether the five-year impact of KIPP for the pre-K cohort (represented by the far right-hand dark blue box in the top panel of Figure II.1) is different from the three-year impact of KIPP for the kindergarten cohort (represented by the dark blue box in the bottom panel). Because both cohorts experienced the impacts of KIPP in kindergarten through grade 2, but only the pre-K cohort experienced the impacts of KIPP in pre-K, any difference represents an estimate of the additional benefit of attending KIPP pre-K over and above that of attending KIPP from grades kindergarten through 2. Unlike the estimates for Research Question 1, these impact estimates are not causal; however, they still provide preliminary evidence as to the marginal benefit of attending KIPP pre-K.

Finally, to explore how the impacts of admission to a KIPP school change over time (Research Question 3), we compared the size of the impacts for the pre-K cohort at kindergarten to those for the same cohort at grade 2 on the same academic tests (represented by the two blue boxes in the top panel of Figure II.1). Comparing impacts longitudinally for the same group of students over time provides us with insight into whether the size of any impacts from KIPP pre-K through kindergarten appears to increase, decrease, or remain steady as the children in both the treatment and comparison groups proceed to later grades.

A. Study samples

The sample for Research Question 1 included three oversubscribed KIPP elementary schools that also provided pre-K—schools that had substantially more applicants in an entry grade than seats to serve students—in two KIPP cities. Across the three schools, we randomized 473 students into treatment and control groups using the admissions lottery. For Research Question 2, we compared impacts for the same pre-K sample from Research Question 1 to impacts from our sample of kindergarteners. The kindergarten sample included five oversubscribed KIPP schools in four KIPP cities. Across the five schools, we randomly assigned 624 students to the treatment and control groups. Table II.1 summarizes the geographic location, entry grade, and year opened for each school in our study sample for each research question.

Table II.1. Characteristics of schools in the samples

School	City	Entry grade	Year opened	Research question		
				1	2	3
KIPP SHARP	Houston	Pre-K3	2008	X	X	X
KIPP SHINE Prep	Houston	Pre-K3	2004	X	X	X
KIPP LEAP Academy	Washington, DC	Pre-K3	2007	X	X	X
KIPP Academy Elementary	New York City	Kindergarten	2009		X	
KIPP Infinity Elementary	New York City	Kindergarten	2010		X	
SPARK Academy	Newark	Kindergarten	2009		X	
KIPP Philadelphia Elementary Academy	Philadelphia	Kindergarten	2010		X	
KIPP Raíces Academy	Los Angeles	Kindergarten	2008		X	

Our analytic sample varied by cohort and for each outcome year. For example, for the five-year estimates of the impact of admission to a KIPP school in pre-K (Research Question 1), we had outcome data from 52 percent of our original treatment sample (96 students) and 51 percent of our original control sample (147 students). For Research Question 2, the analytic sample included 243 students from the pre-K cohort and 386 students from the kindergarten cohort with valid test scores from grade 2. Table II.2 displays sample sizes for each cohort and outcome year by treatment and control group.

Table II.2. Student sample sizes

	Pre-K cohort			Kindergarten cohort		
	Treatment	Control	Total	Treatment	Control	Total
Baseline sample	183	290	473	290	334	624
Analytic sample (kindergarten follow-up)	104	164	268	n.a.	n.a.	n.a.
Analytic sample (grade 2 follow-up)	96	147	243	180	206	386
Analytic sample (longitudinal analysis)[a]	78	121	199	n.a.	n.a.	n.a.

[a]The longitudinal sample contains students with outcome data at both kindergarten and grade 2. We used this sample to answer Research Question 3.

n.a. = not applicable.

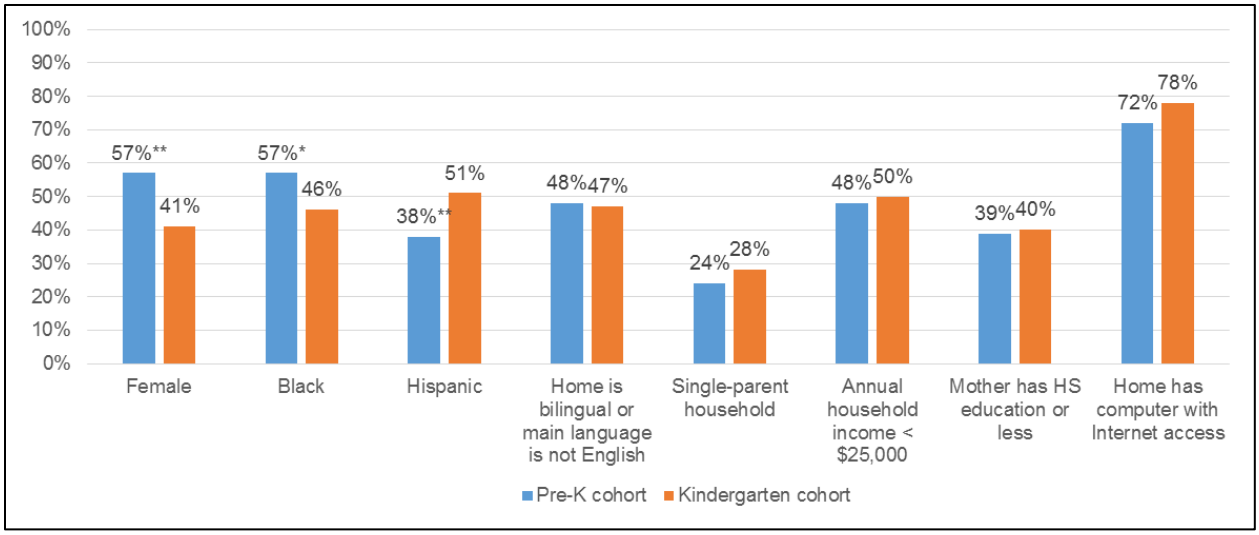
To examine changes in KIPP impacts over time (Research Question 3), we restricted the analytic sample to students who had test outcome data for both kindergarten and grade 2. Across the three pre-K schools, this sample included 199 students—78 in the treatment group and 121 in the control group. This sample poses some additional analytical challenges because the students we observed at both time periods could have been systematically different than students tested only in kindergarten or only in grade 2. Still, the academic impacts trend line for stayers provides useful information about how impacts change over time for the same group of students. (The appendix provides more information on sample composition and the differences between these groups of students.)

Across all analyses, we included lottery winners in the treatment group and lottery losers in the control group, regardless of whether they ultimately enrolled in a KIPP school. As a result, the lottery-based design produces estimates of the impact of an *offer of admission* to a KIPP school (typically referred to as an intent-to-treat, or ITT, estimate) rather than the impact of *attending* a KIPP school. Most lottery winners do attend a KIPP school, however, and most of those not offered admission never do. According to attendance records, among students in our pre-K analytic sample, 81 percent of lottery winters (treatment group) enrolled in a KIPP school between the 2011–2012 and 2015–2016 school years (ever enrolled), and 50 percent were still enrolled in a KIPP school in the 2015–2016 school year—five years after the lottery. A total of 19 percent of lottery non-winners (control group) ever enrolled in a KIPP school; 13 percent remained enrolled in a KIPP school five years later.[5] These differences in enrollment rates show a clear difference between the treatment and control groups in exposure to KIPP schools. We did not include in the analysis all students enrolling in a study school in the entry grade because some were admitted outside of the lottery (for example, if a student had a sibling already enrolled); across all eight schools in both cohorts, 61 percent of open slots were filled via the lottery.

We examined the average characteristics of the students in our analytic sample at the time of the lottery by using information from a baseline survey of the parents of students applying to the KIPP schools (Figure II.2). The two cohorts are similar on several characteristics, such as mother's education, household income, and language spoken at home. However, there are pronounced gender and racial differences across the cohorts. The racial differences are likely due to the regional locations of the schools. (The appendix provides additional detail on sample members' characteristics, types of schools attended, and the baseline equivalence of our samples.)

[5] We collected attendance data for the 2011–2012, 2012–2013, and 2013–2014 school years as part of the KIPP i3 Evaluation from the individual KIPP elementary schools included in the study. We collected attendance data for the 2014–2015 and 2015–2016 school years from the KIPP Foundation, and included data from all KIPP early childhood and elementary schools in the jurisdictions that encompassed our pre-K sample schools. The KIPP Foundation records indicate that some of the students in our sample (both treatment and control) ultimately enrolled in other KIPP schools not included in the study. More detail on the attendance patterns at KIPP schools is provided in the appendix.

Figure II.2. Characteristics of students in the samples



Notes: We obtained these data from the baseline survey of parents of applicants to KIPP elementary schools in spring 2011. The sample includes data obtained for 223 students from the pre-K cohort and 360 students from the kindergarten cohort who also have grade 2 outcome data, inclusive of students assigned to both the treatment and control groups.

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

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

B. Outcome measures and data collection

To measure academic achievement and executive function, we administered a series of tests to students in our sample. To measure academic achievement, we administered four tests of the Woodcock-Johnson III (WJ-III) Tests of Achievement.[6] We also administered two tests of executive function to the pre-K cohort five years after random assignment, when most students were in grade 2: (1) the Hearts & Flowers assessment, which measures cognitive flexibility; and (2) the WJ-IV Verbal Attention test, which measures working memory. Table II.3 describes each study test and details which tests were administered to each cohort and at what point in time.

We standardized the students' WJ-III and WJ-IV scores into z-scores, using information on the performance of a nationally representative norming population. Thus, each student's score represents his or her achievement level relative to the national average for students at that grade level: scores greater than zero represent above-average achievement in the domain being tested; scores less than zero represent below-average achievement. Because no national norming data are available for the Hearts & Flowers assessment, these outcomes were converted into z-scores using sample means and standard deviations, so scores represent students' performance relative to other students in our sample.

[6] We selected the WJ-III because, relative to other tests for this age range, it (1) posed a low testing burden on young students in the amount of time it takes to administer and (2) has a reliability for students ages 6 to 9 of greater than 0.90 for the reading tests and greater than 0.80 for the math tests (McGrew et al. 2007).

Table II.3 Summary of outcome measures and testing periods

Study administered test	Description	Outcome grade (cohort)		
		K (Pre-K)	2 (Pre-K)	2 (K)
Academic outcomes				
WJ-III Letter-Word Identification	Measures reading skills. Children name letters and read words of increasing difficulty.	X	X	X
WJ-III Passage Comprehension	Measures reading comprehension. Children silently read and complete sentences based on understanding of a sentence or passage.	X	X	X
WJ-III Calculation[a]	Measures students' skill in analyzing and solving practical math problems. Children answer questions and solve word problems to demonstrate understanding of math concepts and vocabulary.		X	X
WJ-III Applied Problems[b]	Measures students' ability to perform mathematical computations. Children solve algorithms and equations of increasing difficulty.		X	
Executive function outcomes				
WJ-IV Tests of Cognitive Abilities Verbal Attention	Measures working memory. Students listen to an intermingled series of animal names and digits presented orally, and respond to questions about each of the sequences presented (for example, "What animal came before the 5?").		X	
Hearts & Flowers	Measures working memory, inhibitory control, and cognitive flexibility. Stimuli (a heart or flower) appear on the right or left side of the screen. There are three parts to this task—congruent, incongruent, and mixed conditions. In the congruent condition (hearts), only a heart appears, and students press on the same side as the heart, requiring students to follow a simple rule. In the incongruent condition (flowers), only a flower appears, and students press on the side opposite the flower, requiring students to exercise inhibitory control. In the mixed condition (hearts and flowers), congruent and incongruent trials appear randomly, requiring subjects to switch flexibly between the two rules (cognitive flexibility). Students also use working memory to recall the rules and implement them. For each part of the task, the score captures whether a student answered correctly and their average reaction time.		X	

Note: "Outcome grade" refers to the grade level of a majority of the sample at that point in time.

[a]We did not administer the Calculation test to kindergarten students because the test is not age appropriate.

[b]We administered the Applied Problems test to students in both cohorts for the KIPP i3 Evaluation, but an error in test administration limited variation on the assessment, making it less likely we would detect impacts of KIPP elementary schools. As a result, in this report we do not use data from the Applied Problems test administered in 2014.

To understand the pre-K experience of the students in the sample, we also conducted semi-structured interviews of KIPP staff who were administrators or staff members at the three KIPP pre-K programs serving as the focus of our pre-K impact estimates. Interviewees served in various roles during the time when our sample was attending pre-K, including those of principal, instructional leader, or teacher. Interviews focused on the characteristics of the programs; the instructional approaches, curriculums, and assessments the programs employed; the supports the programs provided to families; and the experiences and training provided to program staff during the 2012–2013 school year, when the students in our impact sample were typically 4 years old, attending their second year of pre-K (PK4). We conducted the interviews in February 2017—five

years after the time period of interest. Although staff were unable to answer some questions, they could often consult other staff or historical documents to provide accurate reporting on the period of interest.

C. A note of caution

Although the experimental estimates of the impact of KIPP after five years (Research Question 1) are based on a rigorous methodological approach, the sample size for that analysis is relatively small. Thus, we may not have sufficient power to detect impacts that are not large in magnitude. In addition to having a small sample, the contrasts we used to address Research Questions 2 and 3 are not experimental, meaning that any impacts we observe may be due to, or influenced by, factors other than KIPP pre-K. For example, the students in our pre-K cohort differed in key, observable ways from those in our kindergarten sample, including gender and race. Although we controlled for these observed differences in our analysis, there may also be unobserved differences across the groups that can affect outcomes. For Research Question 3, changes in the size of the impact of KIPP over time could also result from other changes occurring during the same time period—for example, a new program implemented in the district attended by the comparison students between the two data collection periods. Further, for this research question, we restricted our analysis to a subsample of students we tested at both kindergarten and grade 2. Unobserved characteristics related to students' likelihood of completing tests at both time periods may mean that any patterns we observed for this subsample of students are less representative of those for the full sample. Still, the analyses in this report represent exploratory attempts to learn whether the impacts of KIPP pre-kindergarten might persist for longer than those observed for other programs. If we find suggestive evidence that this persistence is true, it suggests that further, more rigorous study of KIPP and similar pre-K programs is merited.

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

III. IMPACTS OF KIPP PRE-K AND EARLY ELEMENTARY SCHOOL

In this chapter, we address the study's three key research questions defined in Chapter I. We find rigorous evidence that the cumulative academic impacts of KIPP pre-K and early elementary school are large and statistically significant after five years. Although the size of the analytic sample limits our ability to draw conclusive findings, we also find suggestive evidence that KIPP pre-K may provide an added benefit above and beyond that of KIPP elementary school without pre-K. Finally, looking at trends over time, we find suggestive evidence that students who won an offer of admission to KIPP pre-K appear to maintain an academic advantage over their peers who did not win such an offer on one measure of reading achievement (Letter-Word Identification) as they reach grade 2, although the size of their advantage on the other measure (Passage Comprehension) appears to decrease, but not disappear, over time. Taken together, these results provide preliminary evidence that KIPP pre-K positively affects student achievement and the impact persists to some degree once students reach grade 2.

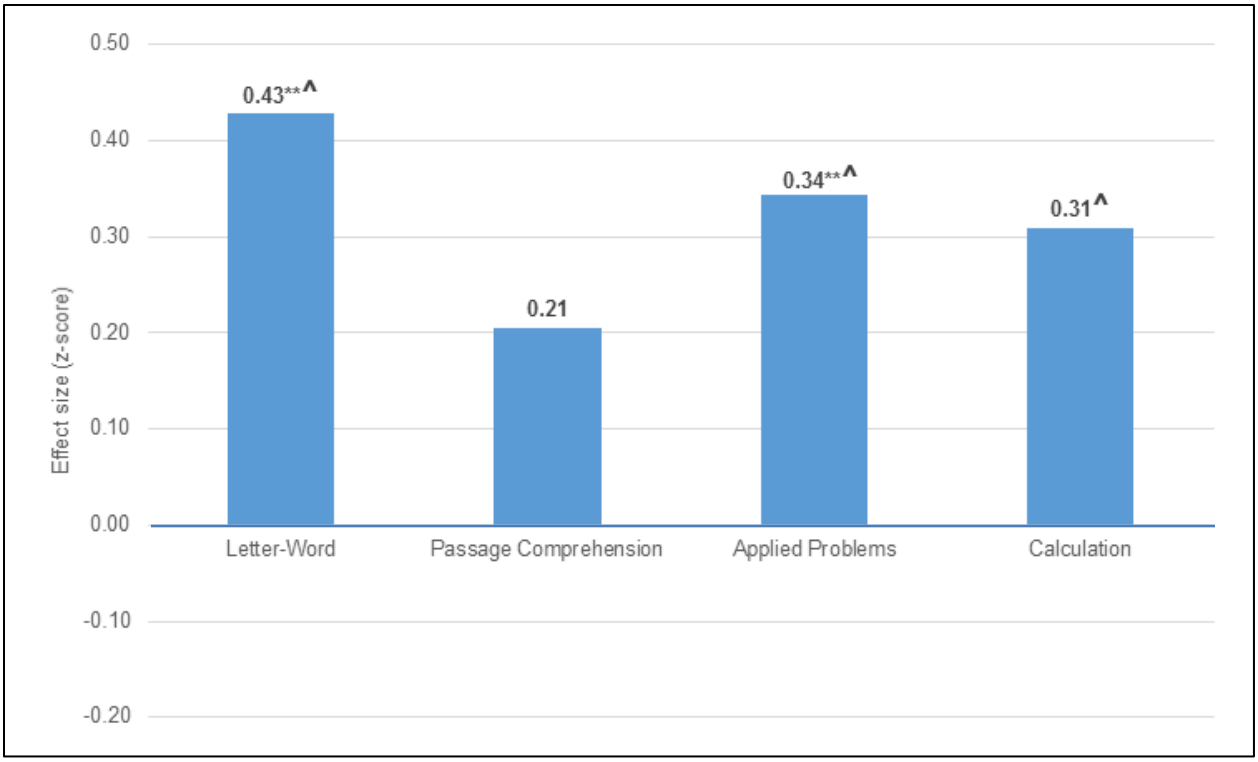
A. After five years, KIPP pre-K and KIPP early elementary has positive and statistically significant impacts on reading and math achievement (Research Question 1)

Being offered admission to KIPP pre-K had a large and statistically significant positive impact on students' Letter-Word Identification score five years after admission, when most students were in grade 2 (Figure III.1). The increase of 0.43 standard deviation units for the treatment group is approximately equivalent to a student moving from the 66th to the 80th percentile.[7] For the Passage Comprehension assessment, an offer of admission had a positive but not statistically significant impact of 0.21 standard deviation units after five years. The precision of our impact estimates is constrained by the size of the pre-K sample we followed from the KIPP i3 Evaluation; we anticipated being able to detect impacts only as small as 0.32 standard deviation units. Nonetheless, this estimate is approximately equivalent to a student moving from the 29th to 36th percentile on the assessment.

In math, an offer of admission to KIPP pre-K had positive impacts of similar magnitude on both outcome measures. On the Applied Problems assessment, this offer had a statistically significant positive impact of 0.34 standard deviation units—approximately equivalent to moving a student from the 47th to the 60th percentile. On the Calculation assessment, the impact of 0.31 standard deviation units was not statistically significant; however, it did meet the What Works Clearinghouse (WWC)'s threshold for being substantively important, defined as an effect size of 0.25 or larger, regardless of statistical significance. This effect size is approximately equivalent to a student moving from the 46th to the 58th percentile on the assessment.

[7] For each outcome measure, we use the percentile corresponding to the control group students' mean score to show average student achievement without the intervention.

Figure III.1. KIPP academic impacts after five years



Source: Study tests administered in spring and summer 2016.

Notes: Outcomes measured on WJ-III Tests of Achievement, administered in the spring of the fifth follow-up year, when most students were in grade 2. All impacts are displayed in z-scores and represent ITT estimates based on regression models that control for baseline covariates. The analytic sample comprises 96 students from the treatment group and 147 from the control group that are part of the pre-K cohort.

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

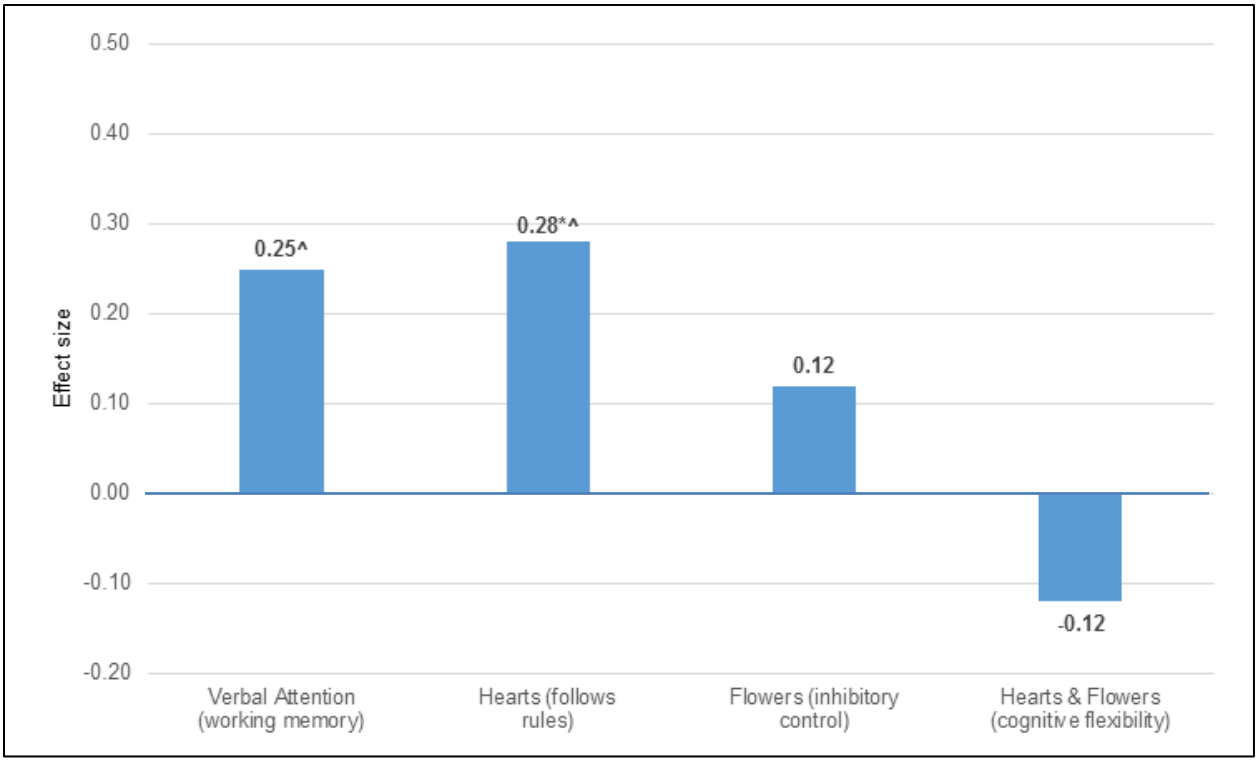
**Impact estimate is statistically significant at the 0.01 level, two-tailed test.

^Impact estimate is substantively important (effect size >= 0.25 standard deviation units).

B. KIPP pre-K and KIPP early elementary may also have a positive impact on students' executive function (Research Question 1)

Although most impacts on students' executive function are not statistically significant, we find some suggestive evidence that an offer of admission to a KIPP pre-K may enhance some of those skills, which are widely believed to be related to students' long-term academic success. An offer of admission to KIPP pre-K had a positive and substantively important impact of 0.25 standard deviation units on students' scores on the Verbal Attention assessment after five years, when most students were in grade 2. This impact was not statistically significant. Results on the Hearts and Flowers assessments varied. An offer of admission to KIPP pre-K had a statistically significant positive impact on students' ability to follow simple instructions (0.28 standard deviation units). The impact on students' inhibitory control was also positive but smaller and not significant. On the other hand, an offer of admission to KIPP pre-K had a negative (but not statistically significant) impact on students' cognitive flexibility (0.12 standard deviation units). Figure III.2 shows complete results for the executive function impacts.

Figure III.2. KIPP executive function impacts after five years



Source: Study tests administered in spring and summer 2016.

Notes: Outcomes are measured on the WJ IV Tests of Cognitive Abilities and the Hearts & Flowers assessment from Adele Diamond's lab at the University of British Columbia. We administered both measures in the spring of the fifth follow-up year, when most students were in grade 2. All outcomes are displayed as z-scores and represent ITT estimates based on regression models that control for baseline covariates. The analytic sample varies by outcome measure—it is between 93 and 96 students for the treatment group and between 144 and 147 students for the control group (all from the pre-K cohort).

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

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

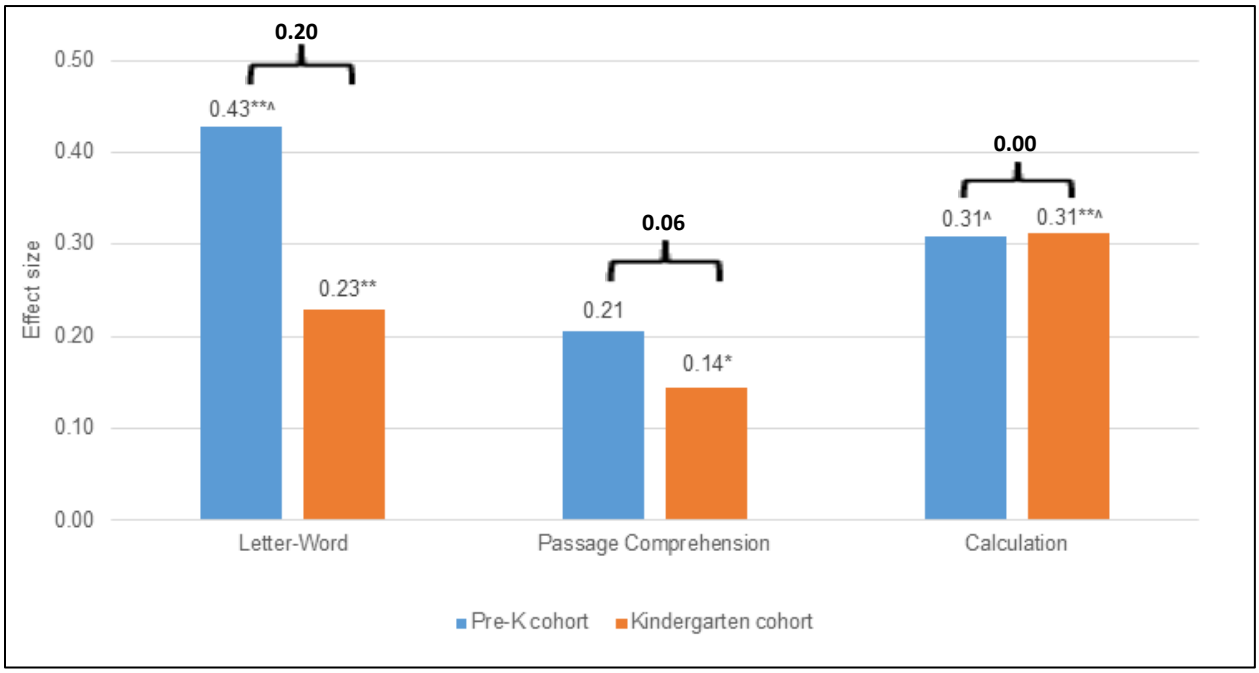
^Impact estimate is substantively important (effect size >= 0.25 standard deviation units).

These latest findings are based on assessments that require students to demonstrate specific skills and consequently minimize reference bias. Executive function measures used in previous studies of KIPP were limited in that they were self-reported by students and parents, and therefore may have suffered from reference bias, wherein KIPP students or parents of students attending KIPP schools may have a consistently different frame of reference when answering survey questions about their attitudes, behavior, and experiences (Tuttle et al. 2013; Tuttle et al. 2015).

C. KIPP pre-K may provide an additional benefit for reading achievement above and beyond KIPP elementary school (Research Question 2)

In reading, the magnitude of the positive impact was larger for the pre-K cohort than the kindergarten cohort on both the Letter-Word and Passage Comprehension tests administered in grade 2 (by 0.20 and 0.06 standard deviation units, respectively, Figure III.3). Neither of these differences is statistically significant; however, the study did not have sufficient power to detect differences of this magnitude. Thus, they may be suggestive of some additional benefit in reading resulting from an offer of admission to KIPP pre-K, above and beyond the impact of an offer to KIPP in kindergarten. In math, however, the impacts for both samples are identical, suggesting no additional benefit of KIPP pre-K beyond the impact of a KIPP elementary school.[8]

Figure III.3. Isolated impact of KIPP pre-K in grade 2



Source: Study tests administered in spring and summer 2016 for the pre-K cohort and in spring and summer 2014 for the kindergarten cohort.

Notes: Outcomes are measured on WJ-III Tests of Achievement, administered in the spring and summer of 2014 for the kindergarten cohort and the spring and summer of 2016 for the pre-K cohort, when most students in both samples were in grade 2. All impacts are displayed in z-scores and are ITT estimates based on regression models that pool all schools and control for baseline covariates. Differences in impact estimates between cohorts are not significant for each test outcome. The analytic sample for the pre-K cohort comprises 96 students from the treatment group and 147 from the control group. For the kindergarten cohort, the analytic sample varies from 176 to 177 students for the treatment group and from 195 to 204 students for the control group depending on the outcome measure.

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

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

^Impact estimate is substantively important (effect size >= 0.25 standard deviation units).

[8] All three kindergarten cohort impacts were positive and statistically significant. Appendix Table 6 provides more detail on results specific to the kindergarten cohort. These cohort-specific results are consistent with previous research on KIPP schools that has found positive and statistically significant impacts of KIPP on academic achievement (Tuttle et al. 2013; Tuttle et al. 2015).

Although these results are exploratory and based on a small sample of KIPP schools that offer pre-K, they suggest that earlier and longer exposure to KIPP improves reading outcomes. It is possible that KIPP pre-K has a stronger focus on reading achievement relative to other programs and/or its pre-K curriculum is better aligned with early elementary grades, resulting in a larger cumulative impact at grade 2. We explore these themes further in Chapter IV.

D. The KIPP impact on Letter-Word Identification scores persists over time, but impacts on Passage Comprehension largely dissipate by grade 2 (Research Question 3)

Students offered admission to KIPP in pre-K scored statistically significantly higher on Letter-Word Identification than students not offered admission. The students who won admission continued to score significantly higher than those who did not when they were tested in grade 2 (Figure III.4).[9] The change in the impact estimates over time (1.58 points) was not statistically significant, meaning that the positive impact of KIPP measured in kindergarten was maintained in grade 2.[10] Though the students admitted to KIPP in pre-K continued to outperform their peers over time, both groups of students experienced similar levels of growth on this measure over time.

The results for the Passage Comprehension assessment tell a different story (Figure III.5). In kindergarten, students who won admission to KIPP pre-K had higher scores on the Passage Comprehension assessment than those who did not. The difference in scores was statistically significant. By grade 2, however, that difference decreased considerably and was no longer statistically significant. Although both groups scored higher in absolute terms in grade 2 than in kindergarten, the students who did not win admission to KIPP largely caught up to the students who did by grade 2. The difference in impacts between kindergarten and grade 2 (-8.46 points) is statistically significant.

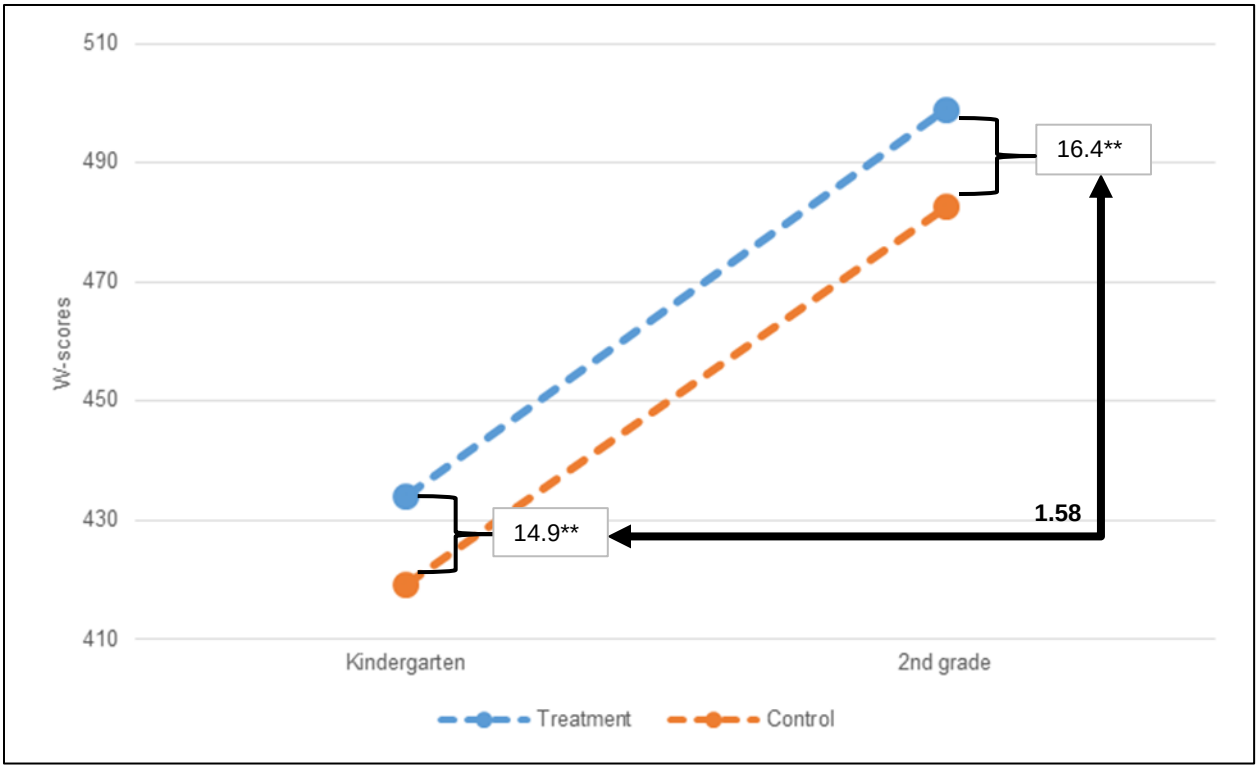
As we mentioned in Chapter II, these analyses are exploratory and should be interpreted with caution. In particular, the sample for this longitudinal analysis is composed of students who completed study-administered testing at two different time periods; this sample differs from the full sample of students who applied to KIPP pre-kindergarten. We discuss these differences in more detail in the appendix.

Taken together, the results in this chapter provide early evidence that KIPP pre-K produces lasting, positive impacts on student achievement. In Chapter IV, we contextualize these findings by describing the experiences of students in the study groups.

[9] Figures III.4 and III.5 display the mean results for the students who won an offer of admission to a KIPP pre-K (blue dots) and students who did not (orange dots) at two time points: three years after the admissions lotteries (when most students in our sample were in kindergarten) and five years after the lotteries (when most students in our sample were in grade 2).

[10] For this analysis, the mean scores for the treatment and control groups are presented as W scores. W scores are equal-interval scores—that is, a change at one point on the scale is equal to a change at another point on the scale. The scores reflect the relative difficulty of the items such that children who correctly respond to more difficult items receive credit for knowing more challenging information. As a result, the W scores allow us to visualize absolute student progress over time.

Figure III.4. Changes in Letter-Word Identification scores over time



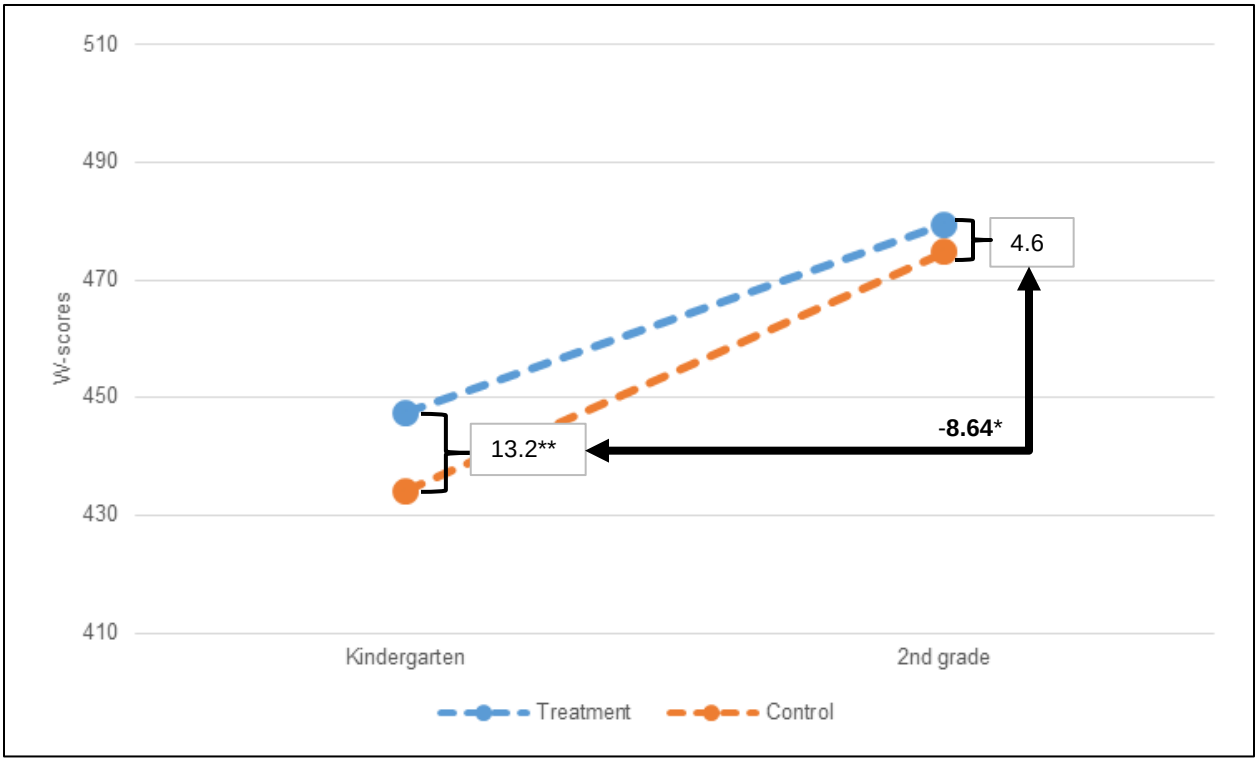
Source: Study tests administered in spring and summer 2014 and 2016.

Notes: Letter-Word Identification is measured on the WJ-III Tests of Achievement, administered in the spring and summer of 2014 for the pre-K cohort in kindergarten and the spring and summer of 2016 for the pre-K cohort in grade 2. Impacts estimates are displayed in the boxes as W-scores, which adjusts for the grade-level difficulty of the test, and are ITT based on regression models that pool all schools and control for baseline covariates. The dotted lines illustrate trends in students' scores on this outcome over time. The bold value at the intersection of the bold arrows displays the change in the size of the impact estimate measured in grade 2 compared with that measured in kindergarten. A positive number indicates the size of the impact of KIPP increases over time; a negative number means that the size of the impact decreased. The sample comprises 78 students in the treatment group and 121 students in the control group from the pre-K cohort.

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

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

Figure III.5. Changes in Passage Comprehension scores over time



Source: Study tests administered in spring and summer 2016.

Notes: Passage Comprehension is measured on the WJ-III Tests of Achievement, administered in the spring and summer of 2014 for the pre-K cohort in kindergarten and the spring and summer of 2016 for the pre-K cohort in grade 2. Impacts estimates are displayed in the boxes as W-scores, which adjusts for the grade-level difficulty of the test, and are ITT based on regression models that pool all schools and control for baseline covariates. The dotted lines illustrate trends in students' scores on this outcome over time. The bold value at the intersection of the bold arrows displays the change in the size of the impact estimate measured in grade 2 compared with that measured in kindergarten. A positive number indicates the size of the impact of KIPP increases over time; a negative number means that the size of the impact decreased. The sample comprises 78 students in the treatment group and 121 students in the control group from the pre-K cohort.

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

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

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

IV. KIPP PRE-KINDERGARTEN EXPERIENCE

Although most students in the study attended a school-based educational program at ages 3 and 4, students in the treatment group were much more likely to be enrolled in KIPP at some point during the five years after admissions lotteries for KIPP pre-K programs. These students also experienced more years of exposure to KIPP on average. Regarding the pre-K experience KIPP schools provided to students in our sample, several program features may have contributed to its large impacts and their greater persistence over time. In particular, KIPP pre-K programs attended by students in our sample focused heavily on academics. Other specific features of the KIPP schools may have contributed to alignment between the educational experience of students in pre-K and early elementary grades. We discuss each of these findings in more detail in this chapter.

A. Schools attended by students in the sample

Among the students in the analytic sample for the pre-K cohort, 81 percent of lottery winners (treatment students) ever enrolled in a KIPP school, whereas 19 percent of those who did not win an admissions lottery to one of our sample schools in spring and summer 2011 (control students) still ended up attending KIPP at some point during the follow-up period (Table IV.1). The 62 percentage-point difference in enrollment rates provides a clear contrast between treatment and control students in exposure to KIPP schools. Contrasting the enrollment for our analytic sample to the enrollment for the baseline sample based on the original lotteries, we find that students who ultimately enrolled in KIPP were more likely to remain in the study than students who did not, regardless of whether the students were initially assigned to the treatment or comparison groups.

The enrollment contrast gradually decreased over the study period but remained large. By the time we measured impacts for the pre-K cohort in grade 2, 60 percent of treatment students in our analytic sample were still enrolled in a KIPP school, compared to 16 percent of control students in our pre-K sample. Although the percentage-point difference in enrollment rates decreased over time, there is still a strong difference between the KIPP exposure of students in the study groups—on average, students in the treatment group attended KIPP for 3.31 years, compared to an average of 0.51 years for students in the comparison group.

Although enrollment rates at KIPP differed, most students in both the treatment and comparison groups attended some form of school-based educational program (at a charter, traditional public, or private school) at ages 3 and 4 (Table IV.2). The large majority of students in both the treatment and control groups (86 percent and 88 percent, respectively) regularly attended a center-based early childhood program when they were ages 3 and 4, according to their parents (not shown). The remaining 14 percent of treatment students and 12 percent of control students did not attend an early childhood program regularly during the pre-K years. At age 3, almost a quarter of students in the control group attended a non-school-based program, such as a day care or nursery school (24 percent), compared with only 12 percent of students in the treatment group; by age 4, fewer than 10 percent of both groups attended a day care or a nursery school.

Table IV.1. Sample enrollment at KIPP

	Analytic sample		Baseline sample	
Enrollment at KIPP	**Treatment**	**Control**	**Treatment**	**Control**
Percentage ever enrolled at KIPP	81	19	71	16
Percentage enrolled at each time period (grade)[a]				
Fall 2011 (PK3)	78	5	68	4
Spring 2012 (PK3)	76	5	67	4
Spring 2013 (PK4)	66	6	59	6
Spring 2014 (Kindergarten)	65	7	55	7
Spring 2015 (grade 1)	65	16[b]	54	13[b]
Spring 2016 (grade 2)	60	16	50	13
Mean years of enrollment at KIPP	**3.31**	**0.51**	**2.84**	**0.43**

Source: We collected enrollment data from 2011–2012 through 2013–2014 school years for the KIPP i3 Evaluation from the individual KIPP schools in our study sample. We collected enrollment data from the 2014–2015 and 2015–2016 school years for the KIPP pre-K study from the KIPP Foundation; this collection included data from all KIPP schools in the cities where the schools in our study sample were located.

Note: To measure mean years of enrollment at KIPP, we treated students as enrolled in a KIPP school for a given year if they were listed on a roster for a KIPP school in the spring of the year of interest. PK3 = pre-K, age 3; PK4 = pre-K, age 4.

[a]"Grade" refers to the enrollment grade for the typical sample student in each year after random assignment.

[b]A large increase in the rate of KIPP enrollment for our comparison group occurred in the spring of 2015, because the enrollment data included all KIPP elementary schools in the cities where the schools in our study sample were located, starting in the 2014–2015 school year. The KIPP enrollment increase in that year is driven by enrollment at non-study schools. Because the roster data from previous years included only schools in the study sample, our enrollment estimates in the first three years of the study likely underestimate enrollment in KIPP schools, particularly for students who did not win a lottery for admission to a school in our sample.

Students in the treatment group were much more likely to have attended a KIPP school in pre-K than students in the control group, according to both roster data and parent report. As for non-KIPP pre-K, control group students were much more likely to attend either a traditional public school or a non-KIPP charter school than students in the treatment group (roughly one-third of control group students were enrolled in non-KIPP charters or traditional public schools at both time periods, whereas only 5 to 6 percent of treatment students were enrolled at other charters, and 3 to 12 percent at traditional public schools). No students in the treatment group and very few control group students (5 to 7 percent) were enrolled in a private school. Thus, the impacts of an offer of admission to a KIPP pre-K presented in Chapter III are measured relative to the impacts of the landscape of other early childhood education options available to students in the study cities—primarily relative to the impacts of other school-based educational programs as opposed to home-based care or day care or nursery school.

At grade 2, five years after random assignment, the majority of students in the treatment group still attended a KIPP school (62 percent), whereas the majority of students in the control group were attending a traditional public school (55 percent, Table IV.3). However, almost one-third of treatment group students were attending a traditional public school for grade 2 (30 percent), and almost half of the students in the control group were attending a charter school (19 percent at a KIPP school and 23 percent at a non-KIPP charter).

Table IV.2. Type of early childhood program attended by KIPP pre-K applicants

Percentage (among students with non-missing data on school type)	Age 3		Age 4	
	Treatment	Control	Treatment	Control
KIPP[a]	80	8	74	16
Non-KIPP charter school	5	33	6	34
Traditional public school	3	27	12	37
Private school	0	7	0	5
Other center (such as daycare or nursery school)	12	24	9	8
Sample size (n)	**65**	**84**	**68**	**92**

Source: Parent surveys administered in conjunction with consent forms for spring/summer 2016 study-administered testing.

Note: Includes data for the sample for the pre-K cohort. Proportions reflect the schools that students attended during the 2011–2012 and 2012–2013 school years—the first and second years following admissions lotteries. Among students included in the analysis who attended a program regularly at ages 3 and 4, 82 percent of students in the treatment group had non-missing data on the school they attended at age 3, and 88 percent had non-missing data at age 4. Among students in the control group, the percentage with non-missing data was smaller; 68 percent had non-missing data on the school they attended at age 3, and 78 percent had non-missing data at age 4. We determined the type of school using the National Center for Education Statistics Common Core of Data. We treated programs not listed in the public or private school data sets as "other" center-based programs unless they were public or private schools known to have closed.

[a]The percentage of students whose parents reported they attended KIPP at ages 3 and 4 is similar to the estimates based on roster data reported in Table IV.1 for the treatment group, but the rates of KIPP attendance reported by parents are higher for students in the control group, particularly at age 4. This finding suggests that parents of some students in the comparison group chose to enroll their child in other KIPP schools when they did not win a lottery for a KIPP school in the study sample.

Table IV.3. Type of elementary school attended by KIPP pre-K applicants

School type (grade 2)	Percentage	
	Treatment	Control
KIPP	62	19
Non-KIPP charter school	8	23
Traditional public school	30	55
Private school	0	3

Source: Parent surveys administered in conjunction with consent forms for spring/summer 2016 study-administered testing.

Note: Includes data for the analytic sample for the pre-K cohort. Proportions reflect the schools that students attended during the 2015–2016 school year—the fifth year following admissions lotteries. Three percent of treatment students and 4 percent of control students had missing data on school type. We determined the type of school using the National Center for Education Statistics Common Core of Data.

B. Characteristics of KIPP pre-K during the study period

Based on interviews with staff at the KIPP schools in our sample, here we describe several characteristics of the study pre-K programs during the 2012–2013 school year, when the students in our impact sample typically were attending PK4.[11] Although all of the schools offering KIPP pre-K were part of the same network of charter schools, the characteristics of the programs sometimes differed by site (Table IV.4). We describe six key features of these programs below, highlighting those that might have differed from other, more traditional pre-K programs at the time.

Table IV.4. Features of KIPP pre-K programs (2012–2013 school year)

	Program A	Program B	Program C
Grades served	PK3-K	PK3-grade 4	PK3-grade 4
Number of students	200	239	269
Number of lead teachers	9	7	6
Percentage of students with IEP or IFSP	7	0	0
Percentage of students who spoke another language at home	0	51	82
Number of PK3 classes (schedule options)	5 (full day)	6 (half day; 3 morning and 3 afternoon)	6 (half day; 3 morning and 3 afternoon)
Number of PK4 classrooms (schedule options)	4 (extended day)	5 (full day)	6 (half day; 3 morning and 3 afternoon)
Top three school priorities[a]	1. Mathematics knowledge & skills 2. Literacy knowledge & skills 3. Language development	1. Literacy knowledge & skills 2. Mathematics knowledge & skills 3. Social & emotional development	1. Social & emotional development 2. Language development 3. Literacy knowledge & skills
Accreditation	None	State accreditation agency	State accreditation agency
Year founded	2007	2004	2008
Proportion of teachers with B.A. or higher	100	100	100
Proportion of teachers with M.A.	11	0	0
Principal's highest level of education	M.A.	M.A.	M.A.

Source: Semi-structured interviews with local KIPP staff, February 2017.

Notes: IEP = Individualized education plan; IFSP = Individual family service plan; B.A. = bachelor's degree; M.A. = master's degree; PK3 = pre-kindergarten, age 3.

[a]Interviewees ranked their school's top three priorities during the 2012–2013 school year from among the following options: (a) physical development & health; (b) social & emotional development; (c) language development; (d) literacy knowledge & skills; (e) mathematics knowledge & skills; (f) science knowledge & skills; and (g) creative arts expression.

[11] Two staff from Mathematica coded these interviews to identify similarities and differences across each of the pre-K programs, and then met to resolve disagreements.

1. The structure of the schools supported alignment across school levels

The grade-level organization of the schools varied across the programs in our sample. Two pre-K programs were a part of a larger KIPP elementary school. Both schools offered pre-K for 3-year-old students (PK3) through grade 4. In the third program, there was a separate school for the early childhood and elementary grades. However, the pre-K school was still located on the same site as the local KIPP elementary school.

For all three programs, staff reported that the co-location of the pre-K and elementary grades contributed to students and parents feeling comfortable and familiar with the students' school. Whereas the pre-K and elementary grades were part of separate schools at one site, staff reported that parents did not seem to distinguish between the two schools. Staff reported that the established level of comfort seemed to reduce parents' and students' stress levels, as they were familiar with the building and regularly encountered the same staff and students on campus over time. The co-location also meant that teachers could observe their students as they advanced into the elementary grades and monitor their progress over time.

The schools in our sample had the autonomy to determine their own school leadership structure, and schools structured leadership differently across the three programs. Across all sites, separate individuals often provided oversight for the pre-K versus elementary grades. Two of the programs had a principal who oversaw both the pre-K and elementary grades; however, the junior leadership at these schools was generally organized by grade level, with one administrator responsible for earlier grades and another for later grades. For example, at one school, the principal was responsible for the school overall, one assistant principal was responsible for PK3 through grade 1, and the second was responsible for grades 2 through 4. At the third site, there were separate principals for the pre-K and elementary grades.

School leadership structure also contributed to the degree of continuity in instruction between the pre-K and elementary grades. For example, administrators at two schools provided instructional leadership for both the pre-K and early elementary school grades, which meant that teachers in both grades were receiving similar support and likely implementing similar instructional strategies and approaches. At the two pre-K programs that were part of the larger elementary school, more opportunities existed for collaboration between the pre-K and elementary grades. At these sites, teachers met regularly with their counterparts in the grades above and below them to facilitate alignment between their curricula and ensure that students learned the skills needed for the subsequent grade. In the third pre-K program, where the pre-K grades were in a separate school from the later elementary grades, staff reported having less involvement in their students' elementary school experience.

Responsibility for instructional leadership and ongoing support for teachers also varied by program. At different sites, assistant principals, instructional coaches, or grade-level chairs had responsibility for providing instructional leadership. These leadership responsibilities included activities such as setting broad goals for the school, providing coaching and professional development, and ensuring alignment of instruction to standards and across grade levels. In two programs, the principal or school leader worked with other administrative staff to provide instructional leadership; at the third program staff reported that the school leader did not provide instructional leadership at that time; however, they emphasized that the school leader since has taken on responsibility for instructional leadership at the school.

2. KIPP pre-K programs were heavily focused on academics—particularly emphasizing foundational reading and math skills—during the study period

Staff at all three sites emphasized the academic focus of their pre-K programs, particularly in PK4. When asked to rank their top three priorities at the school, staff at all three programs ranked literacy knowledge and skills in their top three priorities for their students (Table IV.4). Two programs also prioritized mathematics knowledge and skills, and two prioritized language development. One staff member described their school as more heavily academic than most pre-K programs, including other charter programs. All sites indicated that PK4 was more heavily academic than PK3; sites reported using more play-based activities in PK3 to familiarize students with program values and expectations for behavior, as well as to develop early language and math skills, before transitioning to the more structured and academic programming in PK4. All sites used a mix of instructional approaches, including small- and large-group instruction, centers, and indoor and outdoor play. One respondent explained that their program used play for developmentally appropriate learning and embedded academics within all play activities. In the two programs for which we have data on time for free and outdoor play time, students also spent an average of 45 minutes in free play and at least 30 minutes in outdoor play. Staff from two sites mentioned an emphasis on the ultimate KIPP goal of preparing students to enroll in and succeed in college, even in the pre-K grades.

All three programs were in the early stages of the development of their pre-K programs; staff emphasized that they were still honing their academic approach when our student sample was enrolled in pre-K. They stressed that their schools were midway through a transition toward using more play-based, developmentally appropriate instruction.

3. Curriculum and assessments were mostly teacher developed and contributed to alignment in instruction across grades

At the time, most of the academic curricula at the three pre-K sites were teacher developed or borrowed from other curricula. All three sites used thematic units and targeted specific academic and behavioral skills. Staff in all three programs emphasized that curricula were developed to ensure students had the necessary skills for the subsequent grade. The curricula at these programs were developed in collaboration with the teachers in the subsequent grade, sometimes with support from principals or instructional coaches, to vertically align instruction across grade levels.

Teachers developed assessments to measure progress toward the skills students would need in the subsequent grade. Two programs created their own assessments to measure students' progress toward developing these skills and administered them four times a year. The third program used a compilation of established assessments to measure vocabulary development, math skills, social-emotional development, and literacy skills and administered these tests only once a year. Programs also used assessments to inform small-group assignment for students and, in some cases, initiate referrals to in-house support services for students.

4. KIPP pre-K was designed to establish values and build a behavioral foundation for later success at KIPP

Staff reported that the KIPP pre-K programs in the study heavily emphasized establishing common values and behavioral expectations that would serve as the foundation for students'

success at KIPP and beyond. Particularly in PK3, the programs explicitly taught values and behavioral expectations, and wove them into all aspects of instruction.[12] Upon kindergarten entry, staff reported that familiarity with values and expected behaviors allowed students to focus on academic content. One interviewee reported that teachers knew their students would stay at KIPP and so were invested in laying a foundation for their success as they advanced to later grades. Interviewees noted that students who entered KIPP in kindergarten sometimes needed extra support in these areas to master the behavioral expectations. The programs reported recognizing good behavior in addition to imposing consequences for bad behavior. Staff from one program emphasized that teachers called parents to report good behavior; staff from another recognized students for good behavior in weekly assemblies.

Staff reported that strong relationships developed at KIPP pre-K helped to reinforce behavioral expectations. Specifically, established relationships between students and teachers and among students created consistent expectations across grades and facilitated a smooth transition to elementary school. In many cases, students also had siblings attending the same KIPP school, meaning that parents were already familiar with the school's values and behavioral expectations. These values and expectations were also explicitly communicated explicitly to families through home visits and regular communications; staff at one program reported that parents often reinforced its values and expectations by using the language of KIPP values in their interactions with their children. These relationships reinforced the idea of the schools as an extended family for students.

5. Supports for children and families varied across schools, but all schools heavily emphasized building relationships with students and their families

Two programs provided robust child and family services, whereas the third program later developed additional support services. These services included access to a nurse, a speech therapist, an occupational therapist, a physical therapist, a social worker, health screenings, assistance with basic needs, parent education and supports, and referrals to outside services. Although at the time of the interviews the programs did not have a comprehensive screening process or provide training to teachers on connecting students to services, teachers could initiate a referral to additional support services based on student observations.

Staff frequently mentioned a heavy emphasis on building strong relationships with students and their families. In at least one school, this relationship began with a home visit, in which the student's teacher met with the student and his or her parents at their home to discuss expectations and get buy-in from the student's family. Two sites reported holding regular parent education nights to teach parents how they could support their students' learning at home. All three sites described enlisting parents in their students' education through regular communication. Finally, staff reported that because students were on the same campus for pre-K and elementary grades and many had siblings at the schools, parents developed a familiarity with the campus and teachers, reducing stress for both parents and students.

[12] Two of the programs used Children Learning Appropriate Social Skills (Project CLASS), a social and interpersonal skills curriculum developed by the Houston Achievement Place, to teach expected classroom behaviors and appropriate interactions with peers. The third program used a color chart that was sent home daily with students to report on their behavior.

The KIPP pre-K programs also provided supports to families in meeting expectations regarding attendance. For example, staff from one program reported providing parents with transportation support or offering to move the student to a different pre-K schedule option, such as from the morning to the afternoon class. All three programs noted that they rarely needed to enforce attendance policies because issues with attendance were rare.

6. The training provided to staff varied considerably by school, but most teachers were relatively new to teaching and the pre-K grades

The teachers in these programs were relatively new to teaching, with an average of less than three years of previous experience. All teachers had a minimum of a bachelor's degree before becoming teachers at KIPP. The KIPP teachers in the study schools often entered teaching through Teach For America or KIPP's Teacher Residency Program; others came via more traditional teacher education programs or after teaching at traditional public schools. Before working at a KIPP pre-K, many teachers did not have previous pre-K teaching experience, but many had taught at other KIPP schools.

The ongoing training and support provided to teachers varied by pre-K program. In two programs, administrators or instructional coaches regularly observed teachers and provided coaching or feedback. The third school did not provide this type of instructional support at the time our sample was in pre-K but subsequently began providing it. KIPP school-based and regional support staff in the cities where the schools were located led ongoing professional development sessions for KIPP pre-K teachers in group settings. In general, trainings were planned and provided based on teachers' individual or group needs. Staff in one program indicated that teachers could also request permission and funding to attend specific trainings in which they were interested. The KIPP Foundation also provides a two-week summer training, known as KIPP Summit, focused on connecting teachers across schools and shared learning; many of the teachers in the sample schools had attended this training.

V. IMPLICATIONS FOR POLICY AND PRACTICE

Policymakers continue to look to pre-K as a potentially important and cost-effective way to increase student achievement, reduce achievement gaps, and produce lasting positive impacts for participating students. A recent review of the trends in state pre-K funding found that pre-K is popular on both sides of the political aisle and total state funding for pre-K has increased 47 percent over the past five years (Diffey et al. 2017). Local educational agencies are also investing considerable resources in pre-K programs. For example, New York City recently announced a plan to expand the universal pre-K it already offers for 4-year-olds to include 3-year-olds within the next four years.

Recent studies have reported conflicting findings on the long-term academic impacts of pre-K. A recent study of North Carolina's state pre-K found sustained favorable impacts through grade 5 (Dodge et al. 2016), whereas a study of Tennessee's state pre-K found short-term positive impacts that faded out or became negative by the early elementary grades (Lipsey et al. 2015). Our findings contribute to a body of evidence suggesting that pre-K can have lasting positive impacts. Specifically, we find the following:

1. Rigorous evidence that the cumulative impacts of KIPP pre-K and early elementary grades are positive and substantively important
2. Suggestive evidence that KIPP pre-K provides an additional benefit above and beyond the impact of KIPP kindergarten through grade 2
3. Suggestive evidence that the positive impacts of KIPP pre-K may be somewhat sustained, at least until grade 2

In this chapter, we hypothesize why the impacts of KIPP pre-K may be more lasting than the impacts of other, more traditional pre-K programs. We conclude by suggesting topics for future research.

A. Factors that might contribute to the persistence of impacts of KIPP pre-K

Program characteristics, motivations, and contexts vary across different pre-K programs, and these features may have implications for how pre-K programs like KIPP sustain impacts. Based on our interviews with KIPP staff, several potential mechanisms stand out as to how KIPP pre-K brings about lasting impacts:

- The academic focus of the programs, as described by KIPP staff, is consistent with the large positive impacts in reading and math at each follow-up; specifically, the prioritization of literacy knowledge and skills and language development may explain the lasting impacts in the domain of early literacy.
- Leadership at two of the KIPP schools observed teachers regularly and provided coaching and feedback for teaching staff. Many researchers posit that this type of ongoing support and professional development is a key ingredient for maximizing the potential of pre-K programs (Yoshikawa et al. 2013; Phillips et al. 2017).

- Researchers and practitioners frequently cite the importance of elementary school experiences in maintaining the benefits of pre-K participation (see, for example, Phillips et al. 2017). Shared leadership over and/or co-location of the pre-K and elementary grades may have contributed to greater alignment and continuity of experiences between pre-K programs and elementary grades in KIPP schools.
  - The co-location of pre-K and elementary programs seems to have created opportunities for pre-K and early elementary teachers to interact with each other and with families. The additional exposure to KIPP and familiarity with KIPP in general as well as its facilities and staff all likely contributed to increased engagement of families and, in turn, supported more positive transitions from pre-K to early elementary grades (LoCasale-Crouch et al. 2008).
  - The purposeful collaboration between teachers across grades with support from educational leadership has been hypothesized as an important ingredient of pre-K to elementary alignment (U.S. Department of Education 2016a).

Our findings provide support to the growing consensus about effective pre-K programs and factors that may help sustain their benefits. They shed light on key features that researchers can continue to explore in future studies and policymakers and practitioners can consider incorporating in the design of new and existing pre-K programs.

B. Topics for future research

In this report, we find promising suggestive evidence that KIPP pre-K produces lasting impacts on students' academic achievement. Additional research could build on these findings by measuring longer-term impacts of KIPP pre-K, better defining the experience of students who did not have the opportunity to attend KIPP pre-K, and attempting to replicate the findings using a more rigorous analysis. We detail some potential research questions below.

- **Do the impacts of KIPP pre-K persist as students advance past grade 2?** We found evidence that the cumulative impact of KIPP pre-K and early elementary grades was positive and the early impacts of KIPP pre-K persisted in part until grade 2. To see whether the impacts we observed in grade 2 persist to later grades, a future study could continue to measure impacts on student achievement for both cohorts using student test scores from district administrative records.
- **What were the early childhood experiences of students in our comparison group?** For the current study, we collected data about the types of schools attended by students in our comparison group and more detailed information on the features of the KIPP pre-K programs in our sample. However, we could not collect similar information on the features of the pre-K programs attended by students in the control group. A future study could collect data on the features of the non-KIPP pre-K programs to identify similarities and differences in the characteristics of the pre-K programs that might explain the observed differences in impacts.
- **How did the early childhood experience of students in our pre-K cohort differ from the experience of students in our kindergarten cohort?** We found suggestive evidence that there was an additional benefit of attending pre-K, above and beyond the impact of the KIPP

early elementary grades. Key to understanding the implications of this finding is developing a better understanding of the early childhood experiences of students in our kindergarten cohort—those who did not have the opportunity to attend KIPP pre-K. A future study could collect and report data on the early childhood experiences of students in our kindergarten cohort and describe how those experiences may have differed from those of students in the pre-K cohort.

- **Are these findings replicated in a more rigorous analysis?** As detailed in Chapter II, the findings in this report are preliminary and constrained by the limitations of our study sample. Specifically, we isolated the effects of KIPP pre-K by comparing outcomes for students across KIPP sites with and without pre-K. However, this comparison is not experimental, meaning that the impacts we observed may be due to, or influenced by, factors other than KIPP pre-K—specifically, characteristics of the students or features of the schools in our sample. A more rigorous research design could test whether these impacts hold in an experimental analysis (for example, by funding KIPP pre-K slots in cities where it is not currently available and randomly assigning students to KIPP pre-K or delayed enrollment in KIPP at kindergarten).

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

REFERENCES

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. "Inputs and Impacts in Charter Schools: KIPP Lynn." *American Economic Review: Papers & Proceedings*, vol. 100, no. 2, May 2010, pp. 1–5.

Angrist, Joshua D., Parag A. Pathak, and Christopher B. Walters. "Explaining Charter School Effectiveness*." American Economic Journal: Applied Economics*, vol. 5, no. 4, October 2013, pp. 1-27.

Center on the Developing Child at Harvard University. "Building the Brain's 'Air Traffic Control' System: How Early Experiences Shape the Development of Executive Function: Working Paper No. 11." Cambridge, MA: Center on the Developing Child, 2011. Available at http://developingchild.harvard.edu/resources/building-the-brains-air-traffic-control-system-how-early-experiences-shape-the-development-of-executive-function/. Accessed April 28, 2017.

Diffey, Louisa, Emily Parker, and Bruce Atchison. "State Pre-K Funding 2016-17 Fiscal Year: Trends and Opportunities." Denver, CO: Education Commission of the States, January 2017.

Dobbie, William, and Roland G. Fryer, Jr. "The Medium-Term Impacts of High-Achieving Charter Schools." *Journal of Political Economy*, vol. 123, no. 5, October 2015, pp. 985-1037.

Dodge, Kenneth A., Yu Bai, Helen F. Ladd, and Clara G. Muschkin, "Impact of North Carolina's Early Childhood Programs and Policies on Educational Outcomes in Elementary School." *Child Development*, 2016. Available at http://onlinelibrary.wiley.com/doi/10.1111/cdev.12645/full. Accessed April 26, 2017.

Elango, Sneha, Andrés Hojman, Jorge Luis García, and James J. Heckman. "Early Childhood Education." In *Means-Tested Transfer Programs in the United States, Volume II*, edited by Robert Moffitt. Chicago: University of Chicago Press, November 2016.

Flowers, A. "Is Pre-K All it's Cracked Up To Be?" FiveThirtyEight, January 5, 2016. Available at http://fivethirtyeight.com/features/is-pre-k-all-its-cracked-up-to-be/?ex_cid=538twitter. Accessed January 21, 2016.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts." Cambridge, MA: Mathematica Policy Research, January 2012.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. "The Evaluation of Charter School Impacts." Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, June 2010.

Gleason, Philip M., Christina Clark Tuttle, Brian Gill, Ira Nichols-Barrer, and Bing-ru Teh. "Do KIPP Schools Boost Student Achievement?" *Education Finance and Policy*, vol. 9, no. 1, 2014, pp. 36–58.

Gormley, William Jr., Deborah Phillips, and Ted Gayer, "Preschool Programs Can Boost School Readiness." *Science*, vol. 320, no. 5884, June 27, 2008, pp. 1723–1724.

Lake, Robin, Melissa Bowen, Allison Demeritt, Moira McCullough, Joshua Haimson, and Brian Gill. "Learning from Charter School Management Organizations: Strategies for Student Behavior and Teacher Coaching." Seattle, WA: Center on Reinventing Public Education, and Washington DC: Mathematica Policy Research, March 2012.

Lipsey, M.W., D.C. Farran, and K.G. Hofer. "A Randomized Control Trial of the Effects of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors Through Third Grade (Research Report)." Nashville, TN: Vanderbilt University, Peabody Research Institute, 2015.

LoCasale-Crouch, J., A.J. Mashburn, J.T. Downer, and R.C. Pianta. "Pre-Kindergarten Teachers' Use of Transition Practices and Children's Adjustment to Kindergarten." *Early Childhood Research Quarterly*, vol. 23, no. 1, 2008, pp. 124–139.

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

Phillips, Deborah A., Mark W. Lipsey, Kenneth A. Dodge, Ron Haskins, Daphna Bassok, Margaret R. Burchinal, Greg J. Duncan, Mark Dynarski, Katherine A. Magnuson, and Christina Weiland. "Puzzling it Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects, A Consensus Statement." Brookings. Available at https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf. Accessed April 26, 2017.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer. "Third Grade Follow-Up to the Head Start Impact Study: Final Report." OPRE Report 2012-45. Washington, DC: U.S. Department of Health and Human Services, 2012.

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

Stipek, D. Clements, C. Coburn, M. Franke, and D. Farran. "PK-3: What Does it Mean for Instruction?" *Social Policy Report*, vol. 30, no. 2, 2017, pp. 1–22. Available at http://srcd.org/sites/default/files/documents/spr_30_2_final.pdf. Accessed April 28, 2017.

Schanzenbach, Diane Whitmore, and Elizabeth U. Cascio. "The Impacts of Expanding Access to High-Quality Preschool Education." *Brookings Papers on Economic Activity*, Fall 2013, pp. 127–178.

Tuttle, C., K. Booker, G. Chojnacki, T. Coen, P. Gleason, L. Goble, V. Knechtel, and I. Nichols-Barrer. "Understanding the Effects of KIPP as it Scales, Volume I: Impacts on Achievement and Other Outcomes." Final report of KIPP's Investing in Innovation Grant Evaluation. Washington, DC: Mathematica Policy Research, September 2015.

Tuttle, C.C., B. Gill, P. Gleason, V. Knechtel, I. Nichols-Barrer, and A. Resch. "KIPP Middle Schools: Impacts on Achievement and Other Outcomes." Washington, DC: Mathematica Policy Research, February 2013.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. "Case Studies of Schools Implementing Early Elementary Strategies: Preschool Through Third Grade Alignment and Differentiated Instruction." Washington, DC: 2016a. Available at https://www2.ed.gov/rschstat/eval/implementing-early-strategies/report.pdf. Accessed April 28, 2017

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. "Preschool Through Third Grade Alignment and Differentiated Instruction: A Literature Review." Washington, DC: 2016b. Available at https://www2.ed.gov/rschstat/eval/disadv/p-3-alignment-differentiated-instruction/report.pdf. Accessed April 28, 2017.

Weiland, Christina, and Hirokazu Yoshikawa. "Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills." *Child Development*, vol. 84, no. 6, 2013, pp. 2112–2130.

Wong, V.C., T.D. Cook, W.S. Barnett, and K. Jung. (2008). "An Effectiveness-Based Evaluation of Five State Prekindergarten Programs." *Journal of Policy Analysis and Management*, vol. 27, no. 1, pp. 122–154, doi:10.1002/pam.20310.

Woodworth, Katrina R., Jane L. David, Roneeta Guha, Haiwen Wang, and Alejandra Lopez-Torkos. "San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement." Final Report. Menlo Park, CA: SRI International, 2008.

Yoshikawa, H., C. Weiland, J. Brooks-Gunn, M. Burchinal, L. Espinosa, W. Gormley, and M.J. Zaslow. "Investing in our Future: The Evidence Base on Preschool Education." New York: Foundation for Child Development, and Ann Arbor, MI: Society for Research in Child Development, 2013.

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

APPENDIX

DETAILED ANALYTIC METHODS

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

This appendix presents additional details about the analysis of the impacts of KIPP elementary schools offering pre-K. We first present information on the sample and the baseline equivalence of students who won an admission lottery (the treatment group) and those who did not win (the control group). We next discuss the data and the analysis and then present complete outcome results. We conclude by discussing results of the sensitivity analyses.

A. Detail on sample

The KIPP Pre-K Study follows an existing student sample from the KIPP i3 Evaluation, which was based on lotteries for admission to KIPP elementary schools in spring 2011. Of the 23 KIPP elementary schools open at that time, 8 were sufficiently oversubscribed to be included in the analysis for the KIPP i3 Evaluation and held admissions lotteries in spring 2011.[13] At 3 of these 8 schools, students enrolled at age 3 (pre-K); at the remaining 5 schools, students enrolled in kindergarten. The different enrollment points created two cohorts of students based on the grade for which KIPP held admissions lotteries: the pre-K and kindergarten cohorts.

Among students in the study sample schools and grades, a total of 1,250 were admitted based on a lottery result.[14] The original lottery sample was larger than required to meet the study's targeted level of statistical power. There was also a substantial imbalance in the size of the treatment and control groups at the school level for several schools. In other words, a larger number of students participating in the lotteries at these schools were offered admission and included in the treatment group than those not offered admission and included in the control group, or vice versa. To conserve resources while simultaneously maximizing our ability to detect impacts, we randomly selected a subsample of 1,097 students at these schools to comprise the baseline sample for inclusion in the study's data collection.

Properly executed randomization should ensure that there are no differences (observed or unobserved) between the treatment and control groups. In principle, we can test the truth of this statement by examining the baseline characteristics of the treatment and control groups for each of our three analytic samples:

1. Students who entered lotteries for admission to KIPP at pre-K and had outcome data in grade 2 (pre-K cohort, Appendix Table 1), including 96 students from the treatment group and 147 students from the control group
2. Students who entered lotteries for admission to KIPP at kindergarten and had outcome data in grade 2 (kindergarten cohort, Appendix Table 2), including 180 treatment students and 206 control students

[13] A ninth school was sufficiently oversubscribed to be included in the study, but ultimately we it dropped from the baseline and analytic samples because more than half of the sample at that site lacked follow-up outcome data.

[14] Students who apply to oversubscribed schools may be guaranteed admission and thus not be eligible for the study. For example, applicants may be admitted to the school outside of the typical lottery process if they have a sibling already attending the school, or if a predetermined number of seats are reserved for district residents and these seats are not oversubscribed.

3. Students who entered lotteries for admission to KIPP at pre-K and had outcome data in both kindergarten and grade 2 (longitudinal sample, Appendix Table 3), including 78 treatment students and 121 control students

To collect baseline characteristics, we administered a survey to the parents of students participating in KIPP lotteries close to the time of random assignment in spring 2011. We filled in missing information from the baseline survey using data from a follow-up parent survey we conducted in spring 2013.

For the pre-K sample, there was one statistically significant difference in baseline characteristics across the treatment and control groups out of 25 characteristics we examined; we would expect at least one difference by chance alone (Appendix Table 1). For this sample, mothers of students in the treatment group were less likely to have a high school or GED diploma as their highest level of educational attainment than mothers of students in the control group. There was a similar difference for mothers of students in our kindergarten sample; mothers of students in the control group were more likely to have attained less than a high school diploma than mothers of students in the treatment group (Appendix Table 2). There were no other statistically significant differences between the treatment and control groups for these two samples. To account for differences in baseline characteristics that might influence students' achievement, the impact models in this report statistically control for all of the baseline characteristics listed in Appendix Table 1.

We restricted the longitudinal analytic sample to students in the pre-K cohort we tested at both kindergarten and grade 2; this restriction resulted in higher attrition for the longitudinal sample than the samples of all pre-K students tested at a given time point (kindergarten or grade 2). The increased attrition could introduce bias into our impact estimates if average attrition patterns were different for the treatment and control groups. To investigate this possibility, we examined baseline equivalence among the same set of variables we examined for the pre-K sample. We found three statistically significant differences between the treatment and control groups: students in the control group were more likely to (1) be Hispanic; (2) have a family with a household income of more than $50,000 a year; and (3) have a mother with only a high school degree or GED relative to the treatment group. Appendix Table 3 provides baseline equivalence results for the pre-K cohort longitudinal sample.

Appendix Table 1. Baseline equivalence of grade 2 analytic sample (pre-K cohort)

Baseline characteristic (proportion, unless otherwise indicated)	Treatment	Control	Difference	*p*-value	SD_t	SD_c	N_t	N_c
Female	0.56	0.58	-0.02	0.76	0.50	0.50	82	117
Age in years	6.15	6.20	-0.05	0.32	0.37	0.33	82	115
Race/ethnicity								
White, non-Hispanic	0.02	0.01	0.01	0.42	0.15	0.09	91	132
Hispanic (any race)	0.28	0.36	-0.08	0.17	0.49	0.48	91	132
Black, non-Hispanic	0.66	0.59	0.07	0.27	0.50	0.49	91	132
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.04	0.04	0.00	0.93	0.23	0.19	91	132
Language spoken at home								
English	0.61	0.58	0.03	0.54	0.50	0.50	82	117
Another language	0.20	0.23	-0.03	0.61	0.48	0.42	82	117
English and another language	0.18	0.19	0.00	0.95	0.42	0.39	82	117
One adult in household	0.32	0.22	0.10	0.11	0.45	0.42	91	130
Family income								
Less than 15K	0.18	0.26	-0.08	0.25	0.41	0.44	79	111
Between 15K and less than 25K	0.20	0.26	-0.07	0.32	0.41	0.44	79	111
Between 25K and less than 35K	0.29	0.21	0.09	0.20	0.45	0.41	79	111
Between 35K and less than 50K	0.12	0.13	-0.01	0.83	0.30	0.33	79	111
50K or greater	0.21	0.14	0.07	0.23	0.39	0.35	79	111
Mother's education								
Less than high school	0.09	0.09	0.00	0.93	0.31	0.28	82	114
High school or GED	0.21	0.36	-0.15*	0.03	0.41	0.48	82	114
Some college	0.31	0.30	0.01	0.86	0.47	0.46	82	114
College	0.39	0.25	0.14	0.05	0.48	0.44	82	114
Schools applied to								
Other KIPP school(s)	0.05	0.04	0.01	0.77	0.26	0.21	81	112
Other charter school(s)	0.22	0.24	-0.02	0.71	0.37	0.43	79	109
Private school(s)	0.02	0.04	-0.02	0.53	0.11	0.19	80	111
Any other school(s)	0.28	0.28	0.00	0.98	0.43	0.45	78	108
Student has access to computer with Internet at home	0.68	0.77	-0.09	0.17	0.48	0.42	82	115
Number of children's books at home	25	30	-5	0.28	20	37	78	105

Source: We drew baseline characteristics from a baseline survey we administered to the parents of students who applied to KIPP schools in our sample in spring 2011. We filled in missing values for age, gender, race/ethnicity, and a single-parent household indicator from the baseline survey using information from a follow-up survey we administered to the same sample of parents in spring 2013, where possible.

Note: All values in this table are based on non-imputed data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

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

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

Appendix Table 2. Baseline equivalence of grade 2 analytic sample (kindergarten cohort)

Baseline characteristic (proportion, unless otherwise indicated)	Treatment	Control	Difference	*p*-value	SD_t	SD_c	N_t	N_c
Female	0.41	0.40	0.02	0.72	0.49	0.49	170	182
Age in years	8.00	8.02	-0.01	0.77	0.33	0.39	162	169
Race/ethnicity								
White, non-Hispanic	0.01	0.01	0.00	0.82	0.08	0.07	172	186
Hispanic (any race)	0.48	0.45	0.03	0.46	0.50	0.50	172	186
Black, non-Hispanic	0.49	0.52	-0.03	0.39	0.49	0.50	172	186
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.02	0.00	0.80	0.15	0.15	172	186
Language spoken at home								
English	0.56	0.56	0.00	0.94	0.50	0.50	163	172
Another language	0.23	0.21	0.02	0.56	0.45	0.41	163	172
English and another language	0.21	0.23	-0.02	0.64	0.42	0.42	163	172
One adult in household	0.29	0.30	-0.02	0.74	0.43	0.46	174	186
Family income								
Less than 15K	0.21	0.29	-0.08	0.09	0.41	0.45	156	166
Between 15K and less than 25K	0.23	0.23	0.00	0.98	0.43	0.43	156	166
Between 25K and less than 35K	0.23	0.18	0.05	0.27	0.41	0.39	156	166
Between 35K and less than 50K	0.22	0.19	0.02	0.61	0.41	0.40	156	166
50K or greater	0.11	0.10	0.01	0.74	0.30	0.30	156	166
Mother's education								
Less than high school	0.08	0.22	-0.14**	0.00	0.32	0.41	160	169
High school or GED	0.23	0.24	-0.02	0.72	0.42	0.43	160	169
Some college	0.36	0.27	0.09	0.07	0.48	0.44	160	169
College	0.33	0.27	0.06	0.25	0.46	0.45	160	169
Schools applied to								
Other KIPP school(s)	0.15	0.20	-0.05	0.17	0.33	0.40	160	165
Other charter school(s)	0.47	0.44	0.03	0.51	0.50	0.50	159	161
Private school(s)	0.08	0.09	-0.01	0.78	0.25	0.28	159	164
Any other school(s)	0.58	0.57	0.01	0.86	0.50	0.50	157	163
Student has access to computer with Internet at home	0.82	0.75	0.07	0.11	0.40	0.43	159	170
Number of children's books at home	45	47	-2	0.70	39	44	149	160

Source: We drew baseline characteristics from a baseline survey we administered to the parents of students who applied to KIPP schools in our sample in spring 2011. We filled in missing values for age, gender, race/ethnicity, and a single-parent household indicator from the baseline survey using information from a follow-up survey we administered to the same sample of parents in spring 2013, where possible.

Note: All values in this table are based on non-imputed data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

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

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

Appendix Table 3. Baseline equivalence of longitudinal analytic sample (pre-K cohort)

Baseline characteristic (proportion, unless otherwise indicated)	Treatment	Control	Difference	*p*-value	SD_t	SD_c	N_t	N_c
Female	0.55	0.58	-0.03	0.72	0.50	0.50	66	101
Age in years	6.09	6.20	-0.11	0.10	0.37	0.33	66	100
Race/ethnicity								
White, non-Hispanic	0.03	0.01	0.02	0.35	0.16	0.09	74	114
Hispanic (any race)	0.24	0.39	-0.14*	0.03	0.49	0.49	74	114
Black, non-Hispanic	0.69	0.57	0.12	0.09	0.50	0.50	74	114
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.04	0.04	0.00	0.89	0.23	0.18	74	114
Language spoken at home								
English	0.65	0.60	0.05	0.41	0.50	0.49	66	101
Another language	0.19	0.24	-0.05	0.47	0.49	0.43	66	101
English and another language	0.15	0.16	0.00	0.96	0.40	0.37	66	101
One adult in household	0.31	0.21	0.09	0.14	0.43	0.41	74	113
Family income								
Less than 15K	0.15	0.27	-0.12	0.10	0.41	0.45	63	97
Between 15K and less than 25K	0.24	0.27	-0.03	0.66	0.45	0.45	63	97
Between 25K and less than 35K	0.26	0.20	0.06	0.38	0.44	0.40	63	97
Between 35K and less than 50K	0.08	0.13	-0.06	0.30	0.21	0.34	63	97
50K or greater	0.28	0.13	0.15*	0.03	0.42	0.34	63	97
Mother's education								
Less than high school	0.10	0.09	0.01	0.85	0.33	0.29	66	99
High school or GED	0.18	0.37	-0.19**	0.01	0.40	0.49	66	99
Some college	0.23	0.29	-0.07	0.41	0.43	0.46	66	99
College	0.49	0.24	0.25	0.00	0.50	0.43	66	99
Schools applied to								
Other KIPP school(s)	0.06	0.04	0.02	0.58	0.24	0.20	65	97
Other charter school(s)	0.27	0.28	0.00	0.97	0.38	0.45	63	94
Private school(s)	0.03	0.04	-0.01	0.76	0.13	0.20	64	96
Any other school(s)	0.33	0.31	0.02	0.76	0.43	0.47	62	93
Student has access to computer with Internet at home	0.74	0.77	-0.03	0.69	0.48	0.42	66	100
Number of children's books at home	28	30	-2	0.64	21	39	62	92

Source: We drew baseline characteristics from a baseline survey we administered to the parents of students who applied to KIPP schools in our sample in spring 2011. We filled in missing values for age, gender, race/ethnicity, and a single-parent household indicator from the baseline survey using information from a follow-up survey we administered to the same sample of parents in spring 2013, where possible.

Note: All values in this table are based on non-imputed data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Control" columns.

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

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

The baseline equivalence results suggest that exposure to the treatment (in this case, attending a KIPP elementary school) could be affecting attrition and biasing our longitudinal impact estimates for the pre-K cohort. To investigate this possibility further, we compared the mean kindergarten test scores for the treatment and control groups from the longitudinal sample—those students in our pre-K cohort tested in both kindergarten and grade 2—to the mean kindergarten test scores for the treatment and control groups from the sample of students tested only in kindergarten (Appendix Table 4, rows 1–6). We also compared mean grade 2 test score outcomes for the treatment and control groups from the longitudinal sample to the mean grade 2 test score outcomes for the treatment and control groups from the sample of students tested only at grade 2. On average, treatment group students tested only in kindergarten had *lower* test scores in kindergarten than treatment group students tested in both kindergarten and grade 2 (row 3). In addition, control group students tested only in kindergarten had *higher* test scores in kindergarten, on average, than control group students tested in both grades (row 6). Our remaining longitudinal sample thus contains treatment students who performed better, on average, than treatment students tested only in kindergarten, and control students who scored worse, on average, than control students tested in kindergarten only. However, students tested only in kindergarten account for a small proportion of the sample, and these patterns may be due to chance. Nevertheless, we controlled for baseline characteristics in our longitudinal analysis and focused this analysis on trends over time rather than the magnitude of the impacts at kindergarten and grade 2.

Appendix Table 4. Average test scores among students based on testing group, by experimental condition

Row	Experimental condition	Sample	Letter-Word Identification (mean)	Passage Comprehension (mean)	Sample size
		Kindergarten outcomes			
1		Students tested only in K	1.12	0.47	26
2	Treatment	Students tested twice	1.59	0.84	78
3		Difference between groups	-	-	
4		Students tested only in K	1.03	0.38	43
5	Control	Students tested twice	0.94	0.20	121
6		Difference between groups	+	+	
		Grade 2 outcomes			
7		Students tested only in grade 2	0.38	-0.39	18
8	Treatment	Students tested twice	1.03	0.38	43
9		Difference between groups	-	-	
10		Students tested only in grade 2	0.48	-0.50	26
11	Control	Students tested in twice	0.41	-0.57	121
12		Difference between groups	+	+	

Source: Study tests administered in spring and summer 2014 and 2016.

Notes: We measured reading outcomes on WJ-III Tests of Achievement and show them in z-scores. Results show the average test score for each subgroup in either kindergarten (K) or grade 2.

Although this attrition also affects our estimates of the impact of KIPP pre-K and early elementary grades after five years, we have fewer concerns about sample attrition biasing these impact estimates because the grade 2 impact sample also contains students who were tested for the first time in grade 2. On average, treatment students tested for the first time in grade 2 scored *lower* on achievement tests than treatment students tested at both kindergarten and grade 2

(Appendix Table 4, row 9). In addition, control students tested for the first time in grade 2 scored *higher* on average than control students tested in both grades (row 12). Essentially, the effect of students who were not retested in grade 2 leaving our sample was likely counteracted in part by the effect of students who were tested for the first time in grade 2 joining our sample.

B. Detail on analytic methods

Model specification

To estimate the impact of offering admission to a KIPP elementary school for all students with grade 2 outcomes, we used the following model:

$$(1)\ y_i = \beta_1 T_i + \beta_2 (T_i * COHORT) + \sum_{k=1}^{K} \alpha_k SCHOOL_{i,k} + \gamma X_i + \varepsilon_i$$

where *i* and *k* index students and schools, respectively, and *y* is the student-level outcome of interest. *T* is a binary treatment status variable indicating whether the student was offered admission to the school via the lottery; *T *COHORT* is an interaction term allowing the effect of treatment to vary by cohort for the pre-K and kindergarten samples. *SCHOOL* is a set of binary variables indicating the school to which the student applied and thus the lottery in which the student participated. *X* is a set of student-level demographic and other control variables. The α symbol represents school/lottery fixed effects, which are commonly used in the charter school literature (Angrist et al. 2013; Dobbie and Fryer 2015) and capture differences in outcomes across lottery sites. The parameter β_1 represents the average impact of winning a KIPP elementary school lottery for the pre-K cohort; $\beta_1 + \beta_2$ represents the average impact of winning a KIPP elementary school lottery for the kindergarten cohort. These are intent-to-treat (ITT) estimates, in that not all students offered pre-K KIPP enrollment did enroll; that is, the treatment group includes some students who did not in fact attend KIPP pre-K.

Our analysis includes the following student covariates (represented by X_i in Equation 1):

- Gender
- Student age in years
- Race/ethnicity
- Language spoken at home
- Whether there is only one adult in the household
- Family income
- Mother's education
- Whether the student has access to a computer with Internet access at home
- The number of children's books in the home

To estimate the impact of an offer of admission to a KIPP elementary school for the sample of students who had outcomes in both kindergarten and grade 2 (the "stayers"), we used the following time series model:

$$(2)\ y_{tik} = \beta_1 T_i + \beta_2 YEAR_t + \beta_3 (T_i * YEAR_t) + \sum_{k=1}^{K} \alpha_k SCHOOL_{i,k} + \gamma X_i + \varepsilon_i + \mu_t$$

where *t, i,* and *k* index time period, students, and schools, respectively, and *y* is the student-level outcome of interest at time *t*. *T* is a binary treatment status variable indicating whether the student was offered admission to the school via the lottery; *T *YEAR* is an interaction term allowing the effect of treatment to vary by time period. As in Equation 1, *SCHOOL* is a set of binary variables indicating the school to which the student applied, and thus the lottery in which the student participated, and *X* is a set of demographic and other controls. The parameter β_1 represents the average impact of winning admission to KIPP among the "stayers" in the sample, when most students were in kindergarten in 2014; $\beta_1 + \beta_3$ represents the average impact of winning admission to KIPP among those same stayers, when most students were in grade 2 in 2016. These are ITT estimates. We used the same set of covariates as in the previous model.

Weighting

The impact model incorporates sample weights to account for the fact that not all students in the lottery have the same probability of being offered admission to the KIPP school (that is, being selected into the treatment group). Some students have a higher probability of being offered admission, either based on their inclusion in a particular stratum defined by a student characteristic or because they have a sibling in the lottery. If we do not use sample weights or otherwise account for these student characteristics in the impact model, then the characteristics of students in the treatment and control groups may differ on average, potentially leading to a bias in the impact estimate. For example, because several KIPP schools use sibling preference rules in their lotteries, students with siblings will tend to be overrepresented in the treatment group and those without siblings will be overrepresented in the control group. If having siblings affects student performance directly or is correlated with some other student or family characteristic not accounted for, it could bias the impact estimate.

The creation of the sample weights is based on the procedure used in Gleason et al. (2010). In the simple case, where all students interested in attending a particular KIPP school enter the lottery and no preferences are given for siblings or other characteristics, the sample weight for a given student is based on the probability that he or she ended up in a particular experimental group (that is, either the treatment or control group). This probability is used in calculating each student's *base weight*. In particular, the base weight assigned to treatment group members is set to the inverse of the probability of being selected into the treatment group. The base weight for control group members is set to the inverse of the probability of being selected into the control group. We then normalize this weight to account for the fact that the sample will be representative of the set of all consenting lottery participants at that school. We set this normalization factor such that the weights of each experimental group sum to one-half of the total sample size within the school. Thus, the sum of all students' weights within a school will be equal to the overall sample size in that school (that is, the number of consenting lottery participants), with the sum of weights among treatment group students equal to that among control group students.

In schools with sibling preference rules, the basic approach to calculating sample weights is the same as in the simple case above.[15] The difference, however, is in the calculation of the probability of admission. No longer can we simply use the number of students offered admission divided by the number of lottery participants. The exact probabilities of admission depend on the number of sets of siblings who participate in the lottery at the school and the number of students within each sibling set. With sibling preference rules, each sibling in the lottery has a higher probability of admissions than non-siblings, so the probabilities are adjusted to account for the number of siblings in each affected lottery.

Imputation of baseline characteristics

If there were missing values for the model's covariates, we imputed them based on other baseline information we collected from the student so that he or she could be included in the sample and contribute to our impact estimates. Our imputation procedure, known as multiple imputation by chained equations, uses non-missing values of baseline covariates to estimate plausible values of baseline characteristics for observations with missing baseline data. Specifically, this method first generates multiple data sets with estimated ("imputed") values for missing baseline characteristics. A separate impact estimate is then calculated using each of the imputed data sets. Finally, these impact estimates are combined using procedures described in Rubin (1987) that account for the variability of estimates calculated using the different imputed data sets. The standard error of each combined impact estimate is adjusted to reflect this variability. The imputation procedure and impact estimation using imputed data are conducted using standard commands in Stata; 20 imputations are used. Imputation is conducted separately by treatment and control groups, and all baseline characteristics included as covariates in the impact model are included in the imputation model. Finally, no outcome measures are imputed, only baseline characteristics.

Although we used these imputed baseline covariates in our analysis of KIPP's impacts, we did not include any of the imputed values in the tests of baseline equivalence discussed earlier in the appendix. For the analysis of baseline equivalence, we simply treated students missing data on a given variable as being missing from the sample.

C. Detailed results on outcomes

The following tables present additional details about the results described in Chapter III. Appendix Table 5 provides impacts estimates of KIPP schools five years after admission, when most students were in grade 2, for the pre-K cohort.

[15] An example of sibling preference rules occurs when a school enters two siblings separately in an admissions lottery. If one of the two siblings is drawn as a lottery winner and offered admission to the school, the other sibling is pulled from the lottery pool and also offered admission.

Appendix Table 5. Five-year impacts of offer of admission to KIPP elementary school in pre-K

Outcome	Adjusted mean, treatment	Mean, control	Impact estimate	Standard error	*p*-value
Math achievement					
Calculation	0.21	-0.10	0.31^	0.21	0.14
Applied Problems	0.26	-0.08	0.34**^	0.11	0.00
Reading achievement					
Letter-Word Identification	0.85	0.42	0.43**^	0.12	0.00
Passage Comprehension	-0.36	-0.56	0.21	0.15	0.16
Non-academic outcomes					
Verbal Attention (working memory)	0.15	-0.10	0.25^	0.16	0.12
Hearts (follows rules)	0.21	-0.07	0.28*^	0.14	0.04
Flowers (inhibitory control)	0.09	-0.03	0.12	0.15	0.41
Hearts & Flowers (cognitive flexibility)	-0.10	0.01	-0.12	0.17	0.47

Source: Study tests administered, spring and summer 2016.

Notes: We measured math and reading outcomes on WJ-III Tests of Achievement. We measured executive function outcomes on the WJ IV Tests of Cognitive Abilities and the Hearts & Flowers (H&F) assessment from Adele Diamond's lab at the University of British Columbia. All outcomes are shown in z-scores. All impacts in this table are ITT, based on regression models that control for baseline covariates. Means for the control group are unadjusted; means for the treatment group are equal to the means for the treatment group plus the estimated impact. The analytic sample varies with between 93 and 96 students for the treatment group and between 144 and 147 students for the control group (all from the pre-K cohort).

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

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

^Impact estimate is substantively important (effect size >= 0.25 standard deviation units).

Appendix Table 6 presents detailed results for cohort-specific impacts when most students were in grade 2. The table also shows the difference in impact estimates across the two cohorts, which isolates the effect of KIPP pre-K.

Appendix Table 7 displays detailed results from our analysis of changes in the impacts of an offer of admission to a KIPP pre-K over time. The longitudinal sample includes students tested both in kindergarten and grade 2. Results are shown in W-scores, which adjust for the grade-level difficulty of the exam.

Appendix Table 6. Isolated effect of KIPP pre-K (exploratory analysis)

Outcome (z-scores)	Impact estimates (grade 2)		Isolated effect of KIPP pre-K (Difference)	Standard error of difference	*p*-value
	Pre-K cohort (KIPP with pre-K)	Kindergarten cohort (KIPP without pre-K)			
Math achievement					
Calculation	0.31^	0.31**^	0.00	0.23	0.99
Reading achievement					
Letter-Word Identification	0.43**^	0.23**	0.20	0.15	0.19
Passage Comprehension	0.21	0.14*	0.06	0.15	0.69

Source: Study tests administered in spring and summer 2016 for the pre-K cohort and spring and summer 2014 for the kindergarten cohort.

Notes: We measured outcomes on the WJ-III Tests of Achievement. All impacts in this table are ITT, based on regression models that pool all schools and control for baseline covariates. The pre-K cohort contains 243 students who applied to three KIPP schools. The kindergarten cohort contains 383 students who applied to five KIPP schools.

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

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

^Impact estimate is substantively important (effect size >= 0.25 standard deviation units).

Appendix Table 7. Reading impacts over time for KIPP pre-K cohort

	Mean, treatment	Mean, control	Impact estimate	Standard error	*p*-value
Letter-Word Identification					
Kindergarten	434.14	419.29	14.85**	4.65	0.00
Grade 2	498.93	482.49	16.44**	4.47	0.00
Difference			1.58	4.05	0.70
Passage Comprehension					
Kindergarten	447.43	434.21	13.21**	3.65	0.00
Grade 2	479.41	474.83	4.58	3.71	0.22
Difference			-8.64*	3.60	0.02

Source: Study tests administered in spring and summer 2014 for the pre-K sample and spring and summer 2016 for the pre-K sample.

Notes: We measured outcomes on WJ-III Tests of Achievement. All impacts in this table are ITT, based on regression models that pool all schools and control for baseline covariates. Sample size = 199 students across 3 KIPP schools.

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

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

D. Sensitivity analyses

In addition to estimating impacts using our primary model, we also conducted analyses to see whether impact estimates were sensitive to alternative weighting and imputation approaches. We estimated impacts using two alternative imputation strategies for missing baseline covariates: (1) mean imputation—calculating the covariate mean by school and experimental condition (separately for the treatment and control groups), and (2) no imputation—estimating impacts only on the sample with available baseline data. These results were similar to our main impact estimates, which used multiple imputation. We also estimated school-specific impacts and found no evidence that a single school was driving our overall results.

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

www.mathematica-mpr.com

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

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD

MATHEMATICA
Policy Research

Mathematica® is a registered trademark of Mathematica Policy Research, Inc.