**Student Characteristics
and Achievement in 22 KIPP
Middle Schools**

Final Report

June 2010

Christina Clark Tuttle
Bing-ru Teh
Ira Nichols-Barrer
Brian P. Gill
Philip Gleason

MATHEMATICA
Policy Research, Inc.

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

**Student Characteristics
and Achievement in 22 KIPP
Middle Schools**

Final Report

June 2010

Christina Clark Tuttle
Bing-ru Teh
Ira Nichols-Barrer
Brian P. Gill
Philip Gleason

MATHEMATICA
Policy Research, Inc.

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

ACKNOWLEDGMENTS

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

CONTENTS

TABLES

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

FIGURES

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

EXECUTIVE SUMMARY

The Knowledge Is Power Program (KIPP) is a bold effort to create a network of charter schools designed to transform and improve the educational opportunities available to low-income families. KIPP schools seek to actively 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. Ultimately, the goal of KIPP is to prepare students to enroll and succeed in college. The KIPP Foundation is guiding this effort by selecting and training school leaders, promoting the program model, and supporting the KIPP network schools. KIPP's "Five Pillars" distinguish its approach:

- ***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*** on learning, both in academics and extra-curricular activities, each day, week, and year

- ***Power to lead*** for school principals, who are accountable for their school's budget and personnel

- ***Focus on results***, by regularly assessing student learning and sharing results to drive continuous improvement and accountability

KIPP has grown from a core of two middle schools established in the mid-1990s to a nationwide network of 82 schools in 19 states and the District of Columbia. In the wake of this growth, the KIPP Foundation, its funders, and other stakeholders are eager to rigorously assess the effectiveness of the program and identify which school practices may be positively related to student outcomes. The Foundation is sponsoring the National Evaluation of KIPP Middle Schools, conducted by Mathematica Policy Research, to examine the impacts of KIPP on the achievement and attainment of its students.

This report presents preliminary findings from a matched, longitudinal analysis designed to estimate KIPP's effect on student achievement. Our preliminary work estimates effects in 22 KIPP middle schools—making this the first report that applies a rigorous (nonexperimental) methodological approach across a nationwide sample of KIPP schools. We selected schools for which we were able to collect longitudinal, student-level data, and that were established by the 2005-06 school year or earlier to ensure that a minimum of two entering cohorts of students per school would be observed for multiple years.

We find that students entering these 22 KIPP schools typically had prior achievement levels that were lower than average achievement in their local school districts. **For the vast majority of KIPP schools studied, impacts on students' state assessment scores in mathematics and reading are positive, statistically significant, and educationally substantial**. Estimated impacts are frequently large enough to substantially reduce race- and income-based achievement gaps within three years of entering KIPP. We describe these findings in more detail in the pages below.

This report is the first in a series of reports that will be produced by the study, which is currently scheduled to continue through 2014. The report concludes by describing our plans for the remainder of the study, which will expand the sample to more KIPP schools, make use of randomized experimental research designs in a subset of schools, and incorporate additional outcome measures beyond state test scores.

Student Characteristics in 22 KIPP Middle Schools

To examine the characteristics of the students who enter KIPP middle schools, we compared the fourth grade characteristics of future KIPP students to those of non-KIPP students in the same districts and in the same elementary feeder schools (in other words, elementary schools attended by students who later enrolled in KIPP). We found no evidence that KIPP middle schools are systematically enrolling more advantaged students from their districts. In Figures ES.1 and ES.2, we show the overwhelming percentage of students at each school in our sample who are identified as racial/ethnic minorities and eligible for the federal free or reduced-price lunch (FRPL) program. Over 90 percent of students at 19 of 22 KIPP schools are black or Hispanic and over 70 percent of students are FRPL-eligible at 14 of 15 KIPP schools for which we had access to that measure. **On average, KIPP middle schools have student bodies characterized by higher concentrations of poverty and racial minorities, but lower concentrations of special education and limited English proficiency (LEP) students, than the public schools from which they draw**.

Figure ES.1. Percentage of Black and Hispanic Students in KIPP Schools
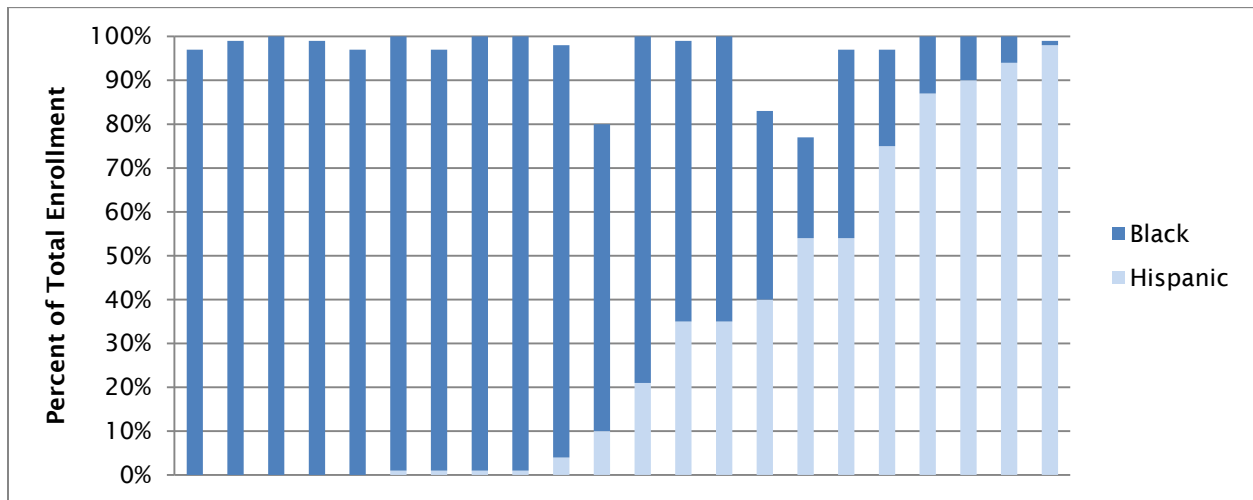
Figure ES.2. Percentage of Students Eligible for FRPL in KIPP Schools
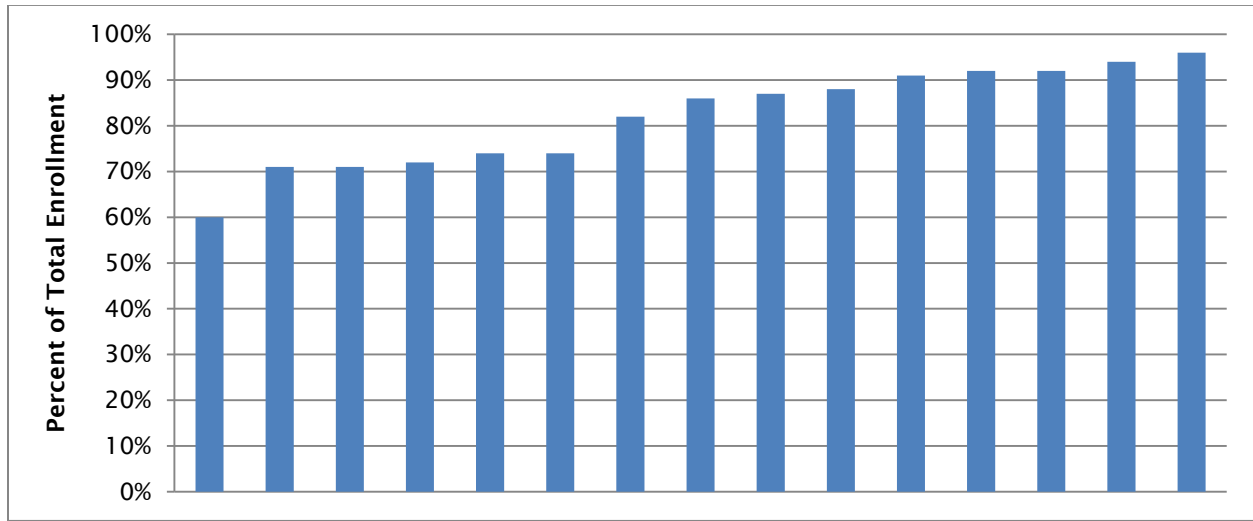
The prior achievement of students entering KIPP schools varies, but KIPP schools most often enroll students whose average fourth-grade achievement is lower than the districtwide average. Figure ES.3 shows whether KIPP students were higher or lower achieving than other students in the district, separately for each KIPP school included in the study. In cases in which the bar—which reflects the difference in baseline test scores between KIPP and district students—is above zero, this implies that KIPP students were higher achieving than district students on average; when the bar is below zero, KIPP students were lower achieving than district students. The size of each bar represents the magnitude of the difference.

Figure ES.3. Distribution of Baseline Math and Reading Differences: KIPP vs. District
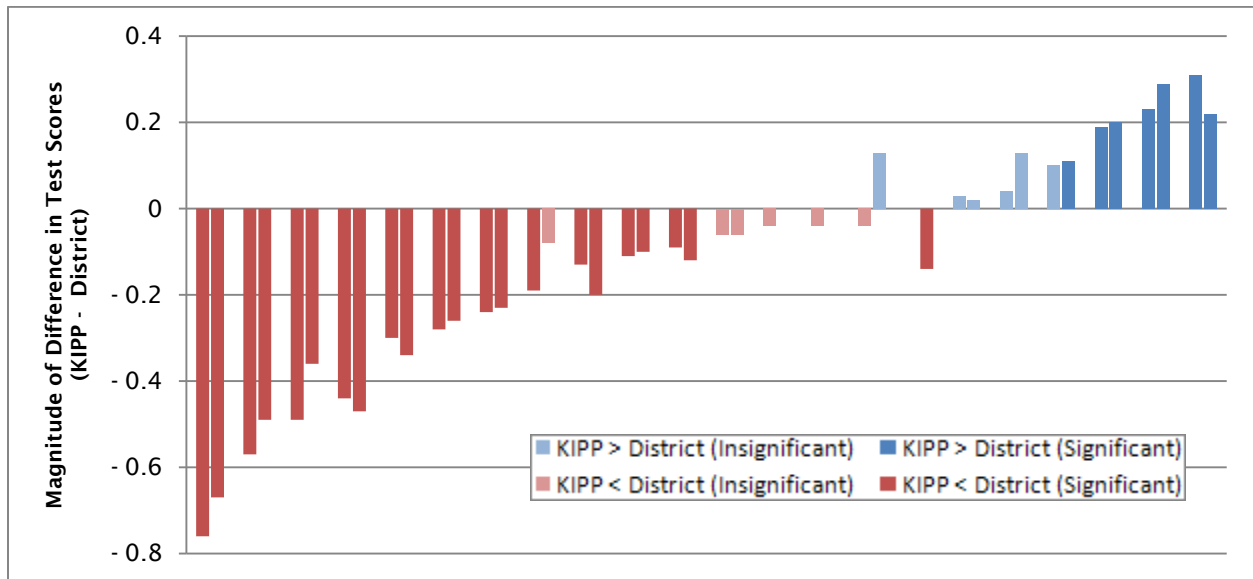
Notes: The bars on the left represent baseline differences in math scores, and the bars on the right represent differences in reading scores. The dark-red and dark-blue bars indicate differences from the district population that are considered statistically significant at the five percent level. Test scores were converted to standardized z-scores to allow comparisons across sites.

We also examined whether students leave KIPP schools prior to completion at unusually high rates. **Cumulative rates of attrition vary widely in different KIPP schools, but we did not find systematically higher (or lower) levels of attrition among these KIPP middle schools as compared with other schools within their districts**. In approximately one-third of cases, attrition was higher at KIPP than at other district schools by a statistically significant margin; in another third, attrition was lower at KIPP than at other district schools by a statistically significant margin.

Grade repetition rates, by contrast, are consistently elevated at KIPP middle schools as compared to district public schools, particularly in fifth and sixth grades. These differences likely reflect differences in policies toward student promotion rather than differences in student performance or achievement. In particular, the differences likely capture KIPP's philosophy that students should be promoted to the next grade level only after they have demonstrated mastery of their current grade's material.

Effects On Test Scores In 22 KIPP Middle Schools

The key outcomes in this matched, longitudinal analysis of the effectiveness of KIPP middle schools are students' scores on the state-administered assessments in reading and math. Our approach to estimating KIPP's effect involves examining the achievement trajectories of KIPP students before and after they enter KIPP schools, and comparing those trajectories to the trajectories of other students in their local school districts. Within this general framework, we employed a variety of different comparison groups and statistical models to explore the robustness of results to different assumptions.

Our preferred, benchmark approach uses propensity score matching techniques to identify other district students who, in elementary school, closely resemble the KIPP students in our sample but do not end up attending KIPP schools. After identifying the matched comparison group, our regression models then control statistically for any remaining differences in the characteristics of the two groups, as related to gender, race/ethnicity, poverty status, special education status, limited English proficiency, and baseline achievement (that is, mathematics and reading test scores for two years prior to KIPP entry). We examine the achievement trajectories of the KIPP students and comparison students in each of the first four years after KIPP entry (as KIPP middle schools typically serve four grades, from fifth through eighth).

Our benchmark estimates reflect the effect of having ever enrolled at KIPP, even if a student subsequently withdraws. This approach—under which all students who spend a year in a KIPP school continue to count as part of the "treatment group," even if they withdraw from KIPP prior to completing eighth grade—is necessary to ensure that impact estimates are not inflated by relying exclusively on students who remain enrolled, given that those who stay are likely to be doing better than those who leave. Although this approach is necessary to avoid an upward bias in impact estimates, it means that the impact estimates implicitly hold KIPP accountable not only for students who stay, but also for students who spend a year at a KIPP school and subsequently leave. If KIPP in fact has positive impacts, then this is a conservative approach that will underestimate the full effect of actually attending KIPP for students who stay.[1]

---

[1] We also estimated the effect of KIPP on students who remain enrolled, by year, but these estimates (reported in the Appendix) are likely to be biased upward.

Our impact estimates suggest two key results across 22 KIPP schools:

Key Finding One: Impacts for large majorities of the 22 KIPP middle schools included in the study are positive in both reading and math in all four years after students enter KIPP schools.

The overall pattern of estimates of the effect of KIPP is consistently positive. This is reflected in Figures ES.4a and ES.4b, which show the proportions of KIPP schools in the study with impacts on math and reading achievement that are positive, negative, and not statistically distinguishable from zero. Within two years after entry, students are experiencing statistically significant, positive impacts in 18 of 22 KIPP schools in math and 15 of 22 KIPP schools in reading. Meanwhile, only two KIPP schools register a significant negative impact in reading in any year of treatment. The single school showing a significant negative impact in mathematics (in year 1) actually reverses that impact by year 3, when it becomes significantly positive.

These findings are consistent across a variety of alternative specifications. Strong positive estimates of KIPP's impact are evident in models using districtwide comparisons (with statistical controls) as well as those using matched comparisons; in models using imputation for missing data as well as those that eschew imputation; and in models that make different assumptions about the test scores of grade repeaters.

Key Finding Two: The magnitude of KIPP impacts is often substantial.

To illustrate the cumulative effects of KIPP schools, we show estimated impacts after three years in Figures ES.5a and ES.5b. The observed effects are especially large in math (which often shows larger impacts in schoolwide interventions; see, e.g., Hoxby, Murarka, and Kang 2009; Dobbie and Fryer 2009; Abdulkadiroglu et al. 2009). By year three, half of the KIPP schools in our sample are producing math impacts of 0.48 standard deviations or more, equivalent to the effect of moving a student from the 30th percentile to the 48th percentile on a typical test distribution. Compared to national norms during this grade span, a 0.48 effect size after three years represents 1.2 years of accumulated extra growth in mathematics over the three year period (Bloom et al. 2008). For comparison, the black-white test score gap in math is typically estimated as approximately one standard deviation at fourth grade and eighth grade (Bloom et al. 2008). Half of these KIPP schools are producing impacts large enough to cut that gap in half within three years.

Three-year impacts in reading are also large in many KIPP schools, if not as large as the effects in math. Half of the KIPP schools in our sample show three-year reading effects of 0.28 standard deviations or more. An effect size of 0.28 standard deviations represents an estimated 0.9 years of additional instruction, or about one-third of the black-white gap in reading achievement in fourth grade and eighth grade (Bloom et al. 2008).

These effect sizes also compare favorably to the effects of other prominent and successful educational interventions. The achievement effects of class-size reduction are often used as a benchmark for other educational interventions. After three years of treatment (grades K-2) in classes one-third smaller than typical, average student gains amounted to 0.20 standard deviations in math and 0.23 standard deviations in reading (U.S. Department of Education, 1998)—less than the effect size achieved in 17 of 22 KIPP schools in math and 13 of 22 KIPP schools in reading.

Figure ES.4a: Percentage of KIPP Schools with Significant Effects in Math, by Year
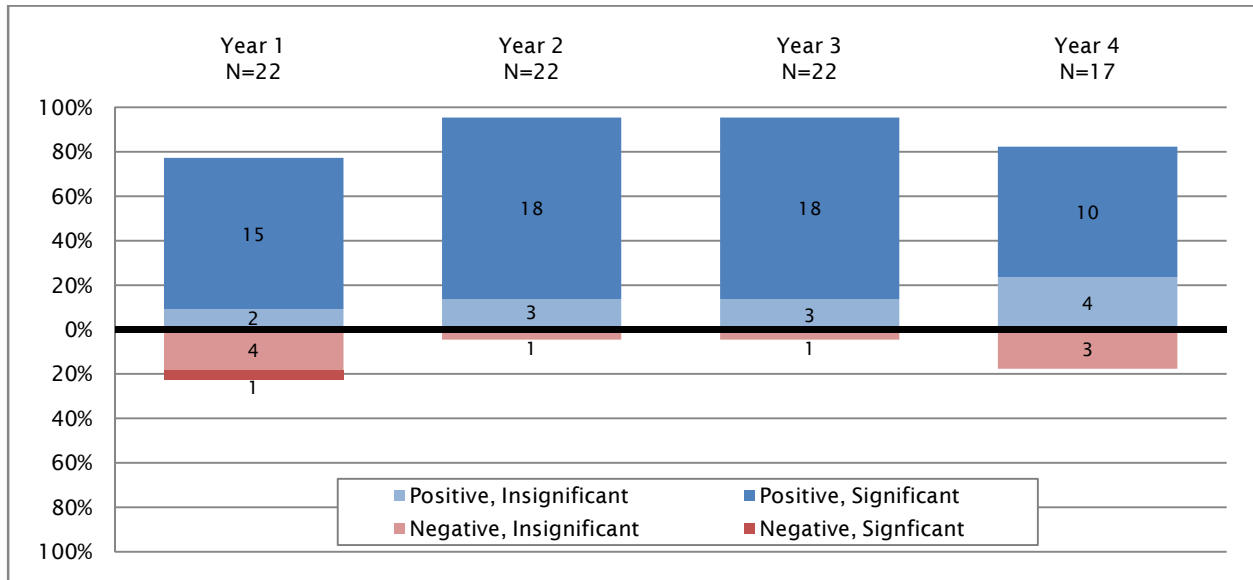
Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact is positive versus negative in a given year. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Figure ES.4b: Percentage of KIPP Schools with Significant Effects in Reading, by Year
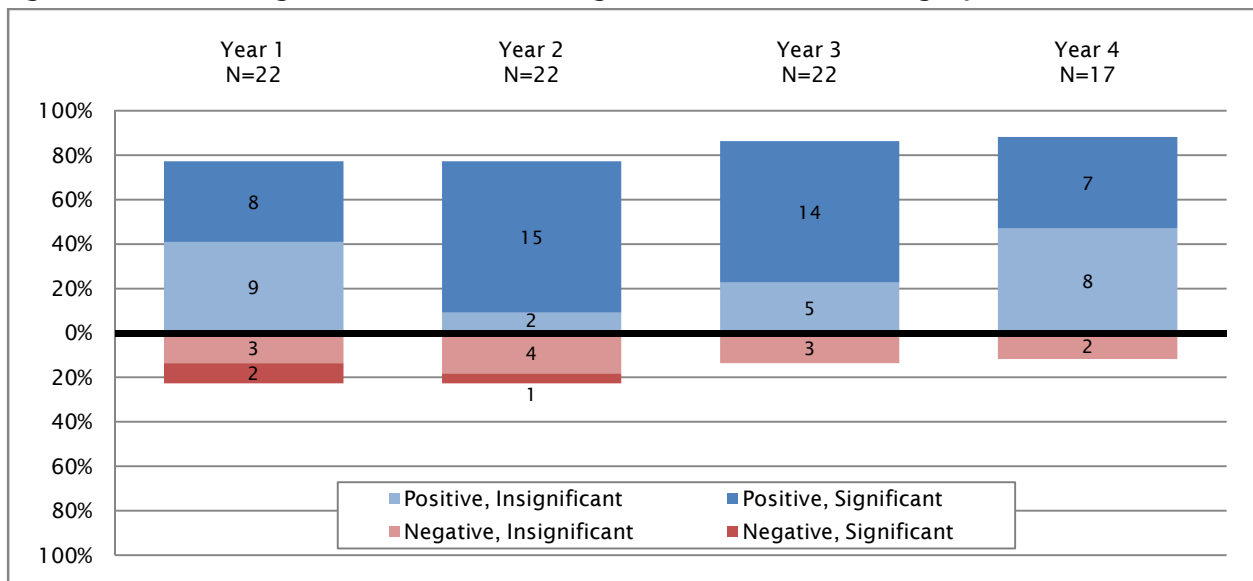
Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact is positive versus negative in a given year. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Figure ES.5a: Distribution of Test Score Effect Sizes After Three Years in Math
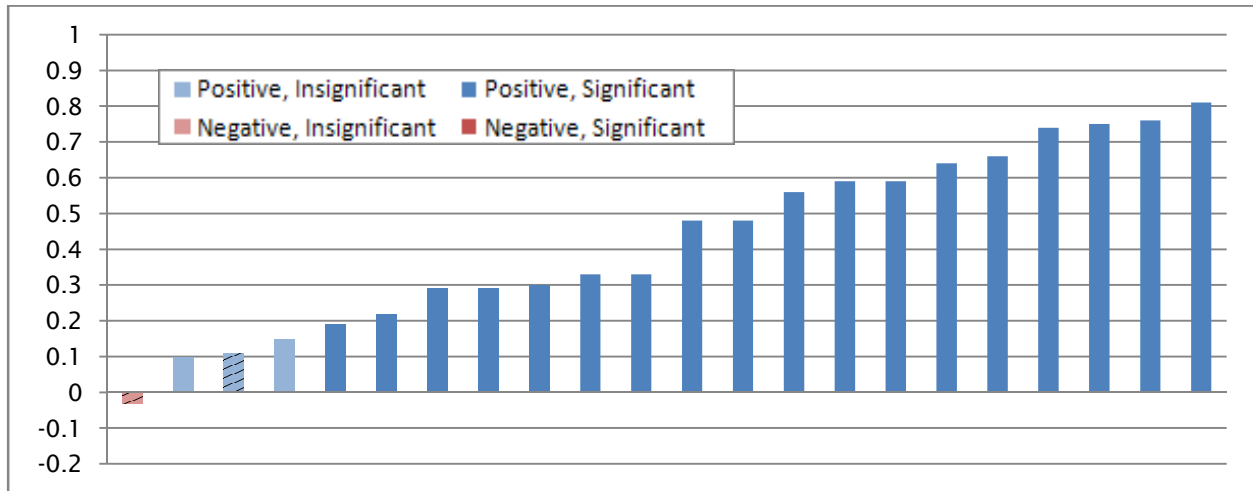
Note: Bars with crosshatching represent schools that had their KIPP affiliation removed. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Figure ES.5b: Distribution of Test Score Effect Sizes After Three Years in Reading
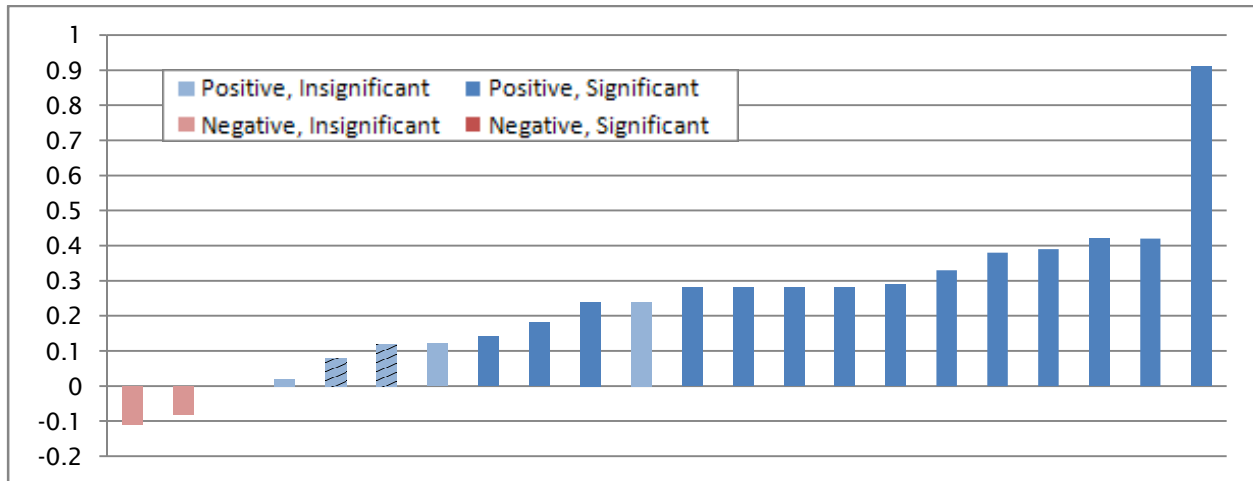
Note: Bars with crosshatching represent schools that had their KIPP affiliation removed. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

More directly relevant are three recent studies of charter schools that found positive impacts. An experimental 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, Murarka, and Kang 2009). If accumulated over three years, these would amount to 0.27 standard deviations in math and 0.16 standard deviations in reading—substantially less than most of these KIPP schools are producing. The KIPP results more closely resemble the results from studies of Boston charter schools and the Harlem Children's Zone's Promise Academy charter schools—results that have been viewed by some observers as spectacular (see, e.g., Brooks 2009). Charter middle schools in Boston are estimated as producing annual achievement impacts of 0.09 in reading and 0.18 in math (Abdulkadiroglu et al 2009), which, if accumulated over three years, would put them squarely in the middle of the KIPP performance distribution (using a method that is less conservative than ours,

and likely to produce larger impact estimates). The Promise Academy middle school produced three-year impacts estimated as 0.27 standard deviations in reading—near the middle of the KIPP distribution—and 0.77 standard deviations in math—near the high end of the KIPP distribution (Dobbie and Fryer 2009).

Finally, as noted above, our estimates are likely to underestimate KIPP's full impact on students who remain enrolled, because we use a conservative approach in which students who have withdrawn from KIPP schools after a year enrolled remain classified in the KIPP "treatment group."

Additional Findings

To better understand the effects of KIPP on student achievement, we conducted some preliminary analysis of how these effects vary across years after enrolling in KIPP, for KIPP schools that have closed, and across student subgroups. Key findings that have emerged from this preliminary analysis include:

- In most KIPP schools, cumulative positive effects increase for at least the first three years after KIPP entry. In math, 18 of 22 KIPP schools show larger cumulative effects in year 3 than year 1, and in reading, 19 of 22 show larger cumulative effects in year 3 than year 1. But the largest single-year impacts are often in the first year, especially in math.

- Of only three schools that never demonstrate a statistically significant positive impact in either mathematics or reading in any year, two are schools from which the KIPP Foundation withdrew the KIPP affiliation. Both schools subsequently closed.

- We find no evidence that KIPP impacts are higher or lower for specific subgroups of students. We examined impacts for the following subgroups: higher versus lower-performing students on test scores at baseline; LEP students; male students; black students and black male students; and Hispanic students and Hispanic male students. We did not find clear patterns suggesting that KIPP impacts for any of these subgroups differed systematically from average impacts for all KIPP students.

Next Steps

The ultimate goal of this evaluation, when completed, is to produce the best possible estimate of the average impact of KIPP middle schools on their students' academic outcomes. Achieving this goal requires using multiple analytic methods in a coordinated way to produce an estimate that has the greatest possible causal rigor while also representing the largest possible sample of operating KIPP middle schools (that is, to produce an estimate that is both internally and externally valid). The overall study was designed to achieve this aim by complementing the strong causal rigor of randomized experiments—also known as randomized controlled trials, or RCTs—based on the admissions lotteries conducted in a subset of KIPP schools, with the greater comprehensiveness that can be achieved using nonexperimental methods in the full population of KIPP middle schools. In future reports, we will expand our analyses by estimating lottery-based experimental impacts and using them to attempt to validate the nonexperimental methods.

In addition to incorporating experimental impact estimates, future reports will also expand the scope of the evaluation by including a larger population of KIPP middle schools; incorporating additional student outcomes beyond state test scores; and exploring aspects of the operation of KIPP schools that may be related to producing larger impacts on students.

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

I.   INTRODUCTION

The Knowledge Is Power Program (KIPP) is a rapidly expanding network of public charter schools designed to transform and improve the educational opportunities available to low-income families. Ultimately, the goal of KIPP is to prepare students to enroll and succeed in college. The achievement levels of KIPP students are often substantially higher than those of schools serving similar populations of low-income, minority students, earning KIPP the attention of educators, policymakers, philanthropists, and the press (Mathews 2009). Indeed, the promise seen in KIPP schools and other charter networks that use similar approaches is a prominent reason that the Obama administration is making the expansion of high-quality charter schools a central component of its nationwide educational improvement agenda. Over the last year, high-quality charter schools and the policies that govern both their establishment and operation received special emphasis from the Obama administration in its goals for the American Reinvestment and Recovery Act funds as well as in its blueprint for the reauthorization of the federal Elementary and Secondary Education Act.

In this chapter, we provide background on KIPP and discuss prior research on its effectiveness. We then describe our multi-year evaluation of KIPP schools in this context, including the contributions of this initial report to the knowledge base.

A.   The Knowledge is Power Program (KIPP)

KIPP schools seek to actively 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's "Five Pillars" distinguish its approach:

- *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* on learning, both in academics and extra-curricular activities, each day, week, and year

- *Power to lead* for school principals, who are accountable for their school's budget and personnel

- *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 inner-city 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 Doris and Don Fisher 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 supports and services to KIPP schools and regions, including legal support, real estate, technology, finance, corporate governance, operations,

communications, marketing, and development. With the support of the KIPP Foundation, the network has grown dramatically. Currently, there are a total of 82 KIPP schools operating in 20 different states, including the District of Columbia. Most of the early schools were middle schools, but KIPP now offers instruction at all grade levels; the network includes 16 elementary schools, 55 middle schools, and 11 high schools. Collectively, these schools currently enroll over 21,000 students.

With the expansion of KIPP nationwide, the KIPP Foundation, its funders, and other stakeholders are eager to rigorously assess the effectiveness of the program and identify which school practices may be positively related to student outcomes. A handful of studies of KIPP's academic performance have been conducted, but they have thus far examined only a small number of KIPP schools. The Foundation is sponsoring this study, which is the first rigorous, large-scale examination of KIPP's academic and non-academic impacts in its schools across the country. The study focuses on two primary research objectives to examine in KIPP middle schools:

1. ***Achievement.*** Does the KIPP approach work? Do students attending KIPP schools perform better academically and socially than they would have had if they attended another public school? How do KIPP students' academic and nonacademic outcomes compare with similar students at other schools?

2. ***Attainment.*** Does the performance of KIPP students—both within and beyond the core academic areas—suggest that they are on a path toward college attainment and persistence?

In addressing one component of the first research question—Do students attending KIPP schools perform better academically?—this preliminary report is the first in a series that the study will produce over the next several years. In particular, this report presents findings from a nonexperimental analysis designed to estimate the effect of a sample of KIPP middle schools on reading and math scores on state assessments. In addition, we describe the evaluation approach we will use to address both research questions in more detail over the next several years of the evaluation.

B. Findings from Prior Research

Previous studies of KIPP schools have used methods of varying rigor, and have involved only small numbers of schools, but they have consistently pointed to favorable results.[2] Several studies of individual KIPP schools, including two quasi-experimental analyses in Baltimore and Memphis that controlled for baseline differences between students, noted that KIPP students' test scores exceed the average for public schools in the same district (MacIver and Farley-Ripple 2007; McDonald et al. 2008). Two descriptive studies attempted to compare results across several different KIPP schools. The first of these, published in 2002, observed that fifth grade students at four newly founded KIPP schools were achieving test score gains similar to the levels observed at the two flagship KIPP schools—KIPP Academy New York and KIPP Academy Houston (Doran and Drury 2002). Similarly, a cross-district survey of Stanford Achievement Test results for students attending 24 KIPP schools, conducted by the Educational Policy Institute, reported that growth rates in average

---

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

test scores at KIPP schools significantly exceeded growth rates observed in a nationally normed sample (Educational Policy Institute 2005).

One criticism of these studies has been that researchers did not fully control for the possibility that KIPP might be selectively enrolling higher achieving students. 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 a quasi-experimental analysis of student impacts at three KIPP schools located in the San Francisco Bay Area. The authors used propensity score matching to compare KIPP students at these schools to a comparable set of control students at traditional public schools. This study has been certified by the What Works Clearinghouse as meeting its standards with reservations, the highest possible rating for a quasi-experimental study design. Across the three Bay Area schools, the authors found large and statistically significant impacts in both math and reading (with effect sizes ranging from 0.16 to 0.68 standard deviation units in reading, and from 0.19 to 0.88 standard deviation units in math) for students in the first year after enrolling in KIPP middle schools (Woodworth et al. 2008). Another report, published in 2010, included results from an experimental study of impacts conducted at the KIPP Academy in Lynn, MA. Researchers used the school's admission lottery to identify a statistically comparable group of treatment and control students. Comparing the test scores of admission lottery winners to students who were not offered admission, the authors found large, statistically significant impacts in mathematics and smaller, significant impacts in language arts. The reported experimental effect sizes for one year of KIPP instruction were 0.35 in mathematics and 0.12 in reading (Angrist et al. 2010).

C. Overview of Multi-Year KIPP Evaluation Design

The goal of the overall evaluation is to produce the best possible estimate of the average impact of KIPP middle schools on their students' outcomes. Achieving this goal requires a method that has the greatest possible causal rigor while also representing the largest possible sample of KIPP middle schools, thereby producing an estimate that is both internally and externally valid.

The challenge in achieving this aim is the tension between maximizing internal validity and maximizing external validity. Estimating KIPP's achievement effects involves the same methodological dilemma that exists in all studies of charter schools. Because KIPP schools, like other schools, are schools of choice—students are not assigned to attend them but instead make an active decision to do so—it is difficult to know whether their students' observed outcomes are attributable to the schools or to the underlying characteristics of the students and their families. The best way to rule out selection bias in estimates of charter school impacts is to use a randomized experimental design in which a student's opportunity to attend the school is determined by lottery. Experimental designs are possible in some charter schools (including KIPP schools) that have more applicants than spaces, if lotteries are used to determine which eligible applicants are admitted. Several existing studies of charter school impacts have relied on admission lotteries to conduct experimental analyses, comparing the academic achievement of lottery winners to lottery losers (e.g., Hoxby and Rockoff 2005; Hoxby et al. 2009; Dobbie and Fryer 2009; Abdulkadiroglu et al. 2009; Angrist et al. 2010).

At issue for our study and other charter school studies, however, is that not all charter schools are suitable for lottery-based analysis. Not all schools are substantially oversubscribed, and even those that are differ in their admission rules and priorities (sometimes favoring applicants from the local neighborhood, for example). An experimental analysis is possible only in a subset of KIPP schools, which are not necessarily typical of all KIPP schools. Indeed, the schools that are most-

heavily oversubscribed might be those that are the strongest performers—a possibility that received some empirical support in the only study that has thus far considered the issue (Abdulkadiroglu et al. 2009).

In later reports, our study will use "gold standard" randomized experimental analysis in the subset of KIPP schools for which admission lotteries can be used to identify control groups of students who applied to, but did not attend, KIPP schools. We will complement these experimental methods with state-of-the-art nonexperimental methods (those used in the current report) that can be applied more broadly to all (or nearly all) KIPP middle schools.

Our ultimate intent is not to produce two different estimates of KIPP's impact (experimental and nonexperimental), which might or might not be similar. Instead, we eventually plan to use the different methods in concert to produce a single estimate of KIPP's impact that takes advantage of the best features of both methods. In particular, we will seek to replicate experimental impact estimates with our nonexperimental methods in the subset of sites where experimental methods are viable. Two recent studies (Abdulkadiroglu et al. 2009; Bifulco et al. 2009) have used both experimental and careful nonexperimental methods to estimate the impacts of schools of choice, with similar results produced by the two methods—suggesting the promise of this approach. Success in replicating the experimental estimates would provide additional validation for applying the nonexperimental method to all KIPP schools and students, thereby providing a single, KIPP-wide estimate of impacts on student achievement.

D. What This Report Contributes to the KIPP Literature

In this report, we use the best available nonexperimental approaches to provide timely estimates of the effect of a national sample of 22 KIPP middle schools. The successful replication of experimental school impact estimates by careful nonexperimental methods in two existing studies of school choice provides sufficient encouragement to make nonexperimental analyses worthwhile even before there is an opportunity to validate them against our own experimental estimates (which are scheduled to be available in 2012). The nonexperimental methods we employ in this report make use of finely-grained, student-level, longitudinal data to identify appropriate comparison groups; estimate models representing the state of the art in nonexperimental impact analysis; and test the robustness of results to a series of sensitivity analyses.

Our nonexperimental methods account for the pre-KIPP (baseline) characteristics of students who subsequently enter KIPP schools: not only demographic characteristics such as race/ethnicity, gender, poverty status, and special education status, but also their achievement test results for two years prior to entering KIPP. We examine the achievement levels of the students for up to four years after entering KIPP schools. In our preferred models, we compare these trajectories to the achievement trajectories of a matched set of local public school students who have similar achievement test results and demographic characteristics in the baseline period (typically third and fourth grades) but who do not enroll in KIPP. We explore a variety of alternative specifications of the statistical models used to analyze the results. We address the possibility of selection bias in the attrition of students from KIPP schools, as well as bias that could result from differential rates of grade repetition between KIPP students and comparison students.

In sum, this report provides the first analysis of KIPP achievement impacts that applies a consistent, rigorous (nonexperimental) methodological approach across a nationwide sample of schools. Analyzing impacts using a common approach at 22 schools across several states provides a

far greater degree of external validity than previous studies when drawing conclusions about the entire KIPP network (especially as compared to studies that only examine a single school district). By applying rigorous nonexperimental methods, and comparing alternate methodological approaches to ensure the robustness of estimated impacts, this report also provides the highest level of internal validity possible until lottery-based analyses become available.

In future reports, we will expand our analyses by including a larger population of KIPP middle schools; incorporating additional student outcomes beyond state test scores; estimating lottery-based experimental impacts and using them to attempt to validate the nonexperimental methods; and exploring factors in the operation of KIPP schools that may be related to producing larger impacts on students.

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

II.   SAMPLE SELECTION AND DESCRIPTION

In this chapter we describe the type of students who attend KIPP, including their demographic characteristics, poverty level, special education status, and test scores prior to enrolling in middle school. Some critics have argued that KIPP schools "cream" or attract the best students; we explore this issue by conducting two sets of descriptive comparisons. First, we compare the demographic characteristics and (pre-KIPP) elementary school test scores of KIPP middle school students to their district counterparts. Second, we examine how the enrollment patterns of KIPP students vary from the patterns of other public students in terms of grade repetition and attrition.

These potential differences also inform the study's impact analyses (presented in the next chapter), as differences between students at KIPP middle schools and those at other non-KIPP public schools (both traditional and charter) may serve as sources of bias when comparing relative school performance. These differences may arise before KIPP enrollment (through the selection of higher- or lower-performing students) or after KIPP enrollment (through selective student attrition). Prior to enrollment, KIPP may attract students who differ from their peers in districts or schools in terms of achievement, race/ethnicity, poverty, special education status, or limited English proficiency status.

After students enter KIPP, there are also several potential pathways through which KIPP enrollment patterns could complicate comparisons between KIPP schools and non-KIPP public schools. One possibility is student attrition. If KIPP students transfer out of middle school prematurely at higher (or lower) rates, as compared to other district public schools, this could indicate a type of selection bias in the composition of students remaining at KIPP through the end of eighth grade. Differing degrees of grade repetition could also create substantial complications in the comparison of KIPP students to non-KIPP students.

In this chapter, we examine each of these factors in turn. The discussion is divided into three sections: first we describe the study's sample selection process for this round of the nonexperimental analysis; second, we describe the baseline characteristics of KIPP students compared to other public school students in their districts and in their own elementary schools; third, we describe KIPP students' enrollment patterns after entering middle school, again compared to those of other public school students in their districts.

A.   Sample Selection Process

The 22 KIPP middle schools included in the study represent a sample that meets two criteria. First, all included schools had to be established in the 2005-06 school year or earlier to ensure that a minimum of two cohorts[3] of students per school would be observed for multiple years. Second, the KIPP 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 traditional public and

---

[3] Throughout this report, 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.

charter schools.[4] Obtaining student-level longitudinal data was necessary to track individual KIPP and non-KIPP students in the baseline years prior to middle school enrollment. All of the data obtained from jurisdictions was de-identified; each student received a unique identifier code to permit longitudinal analyses. The variables obtained from the jurisdictions' administrative data systems included test scores in reading and mathematics (middle school scores represent a primary outcome and elementary school scores represent a key baseline covariate), demographic characteristics (used as baseline covariates), and schools attended (identifying students' exposure to KIPP).

Within each jurisdiction, we requested data for all available school years beginning two years before KIPP first opened a middle school. This information was used to obtain baseline elementary school data for all cohorts of KIPP students. We were able to obtain data that met the two criteria for inclusion for 22 of the 35 KIPP middle schools that were open by 2005. Table II.1 describes the full set of 22 schools in the current sample, the year each school opened, and the years of administrative data included in each obtained data set. (The subset of cohorts included in our impact analysis sample is detailed in Appendix A.)

As the table makes clear, it was not possible to obtain administrative data for all of the student cohorts that have attended the 22 KIPP middle schools in the study. Some states administered reading and mathematics tests only in a subset of target cohorts' year and grade combinations; for cohorts enrolling in KIPP prior to 2003, in particular, annual statewide testing data was often not available (a summary of each jurisdiction's data limitations is presented in Appendix A). Nevertheless, across all schools in the sample, the obtained data includes a minimum of three student cohorts with at least one year of data after KIPP enrollment and baseline elementary school data.

B. Sample Baseline Characteristics

To investigate how KIPP middle school students may differ from other public school students based on observable traits, we examined the fourth grade characteristics of future KIPP students and non-KIPP students. We compared KIPP students first to the districtwide student population,[5] and then to students observed at the subset of district elementary schools (feeder schools) attended by future KIPP students. This approach addresses two related issues: 1) identifying baseline differences between KIPP students and non-KIPP students in the same district; and 2) investigating whether KIPP attracts a different type of student within the elementary schools from which the network draws its student population.

---

[4] Some states were unable or unwilling to provide de-identified student-level data at the time of data collection. In these cases, we collected records from the school districts in which KIPP schools were located; however, not all of these districts maintain data for their resident charter schools (including KIPP schools). Although we are collecting records from KIPP schools independently to merge with data provided by districts, these schools were omitted from this first round of analyses.

[5] Where statewide data was available, the comparison group for the KIPP sample includes all students in the district where the KIPP school is geographically located, as well as students in contiguous districts that encompass elementary schools attended by KIPP students.

Table II.1. Nonexperimental Sample: Descriptive Analyses

State	Jurisdiction Supplying Data	KIPP School (Year Opened)	Number of KIPP Cohorts in Data (School Years)
AR	Arkansas Department of Education	Delta College Preparatory (2002)	3 (2005-06 to 2007-08)
CA	San Francisco Public Schools	Bayview Academy (2003)	6 (2003-04 to 2008-09)
		SF Bay Academy (2003)	6 (2003-04 to 2008-09)
DC	Washington, DC Public Schools	DC KEY Academy (2001)	8 (2001-02 to 2008-09)
		DC AIM Academy (2005)	4 (2005-06 to 2008-09)
GA	Atlanta Public Schools	WAYS Academy (2003)	4 (2004-05 to 2007-08)
		Achieve Academy (2004) – *closed*	3 (2004-05 to 2006-07)
GA	Fulton County School District	South Fulton Academy (2003)	5 (2003-04 to 2007-08)
IL	Chicago Public Schools	Ascend Charter (2003)	5 (2003-04 to 2007-08)
		Youth Village Academy (2003) – *closed*	3 (2003-04 to 2005-06)
MA	Massachusetts Department of Education	Academy Lynn (2004)	4 (2004-05 to 2007-08)
NY	New York City Public Schools	Academy NY (1995)	5 (2003-04 to 2007-08)
		STAR College Preparatory (2003)	5 (2003-04 to 2007-08)
		AMP Academy (2005)	5 (2003-04 to 2007-08)
		Infinity Charter (2005)	5 (2003-04 to 2007-08)
OK	Oklahoma City Public Schools	Reach College Preparatory (2002)	6 (2002-03 to 2007-08)
PA	Philadelphia School District	Philadelphia Charter (2003)	5 (2003-04 to 2007-08)
TX	Texas Education Agency	3D Academy (2001)	5 (2003-04 to 2007-08)
		Aspire Academy (2003)	5 (2003-04 to 2007-08)
		Austin College Preparatory (2002)	5 (2003-04 to 2007-08)
		Academy Middle School (1995)	5 (2003-04 to 2007-08)
		TRUTH Academy (2003)	5 (2003-04 to 2007-08)

　　　　To begin, we examined the demographic composition of the 22 KIPP middle schools in our sample. These characteristics include gender, race, free and reduced price lunch eligibility (FRPL), special education status, and limited English proficiency (LEP)—all of which are measured in fourth grade, prior to KIPP entry (at "baseline"). Detailed information on each of these variables is included in Appendix Tables B.1 and B.2, which compare KIPP students both to the districtwide middle school population and also to the students who attended the same elementary schools as KIPP students.

　　　　KIPP middle schools enroll a high concentration of racial minorities. As Figure II.1 demonstrates, black and/or Hispanic students come close to constituting the entire student body at many of the 22 KIPP middle schools. Across all of the KIPP schools in the sample, only 3 had a combined proportion of black and Hispanic students below 90 percent and none was below 75

percent. Within these schools, it was also common for one ethnic group to be predominant: black students represented more than 80 percent of the student body at 10 of the sample schools, and Hispanics represented more than 80 percent of the student body at 4 of the sample schools.

Figure II.1. Black and Hispanic Enrollment at KIPP Schools
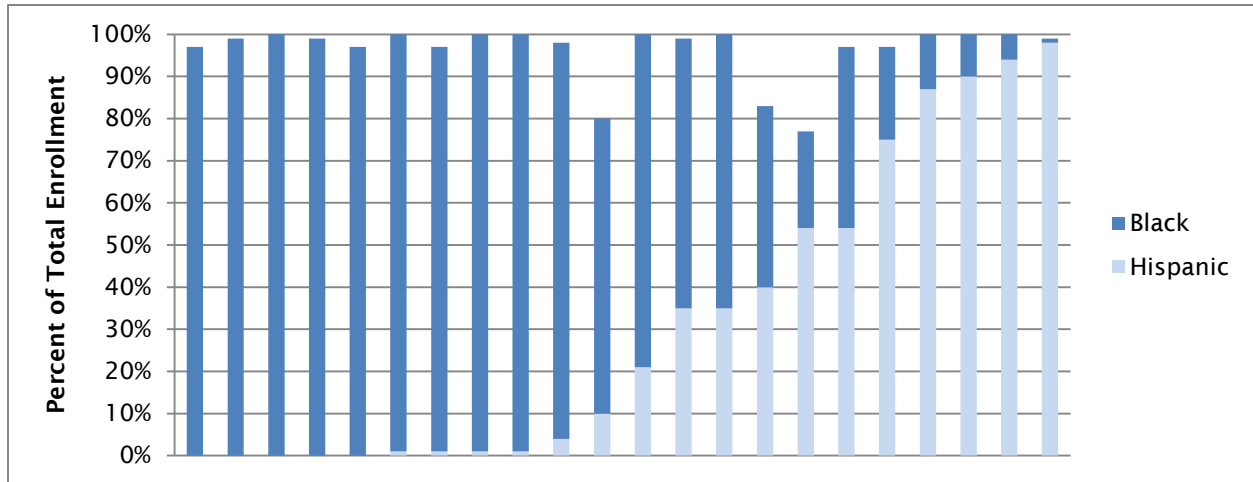
The concentration of racial minorities at every KIPP middle school in the sample is also significantly higher than that found in the KIPP host districts. Figure II.2 depicts the difference between the proportion of blacks and Hispanics at KIPP middle schools and the proportion found in other district middle schools.

Figure II.2. Differences in Minority Enrollment: KIPP vs. District[6]
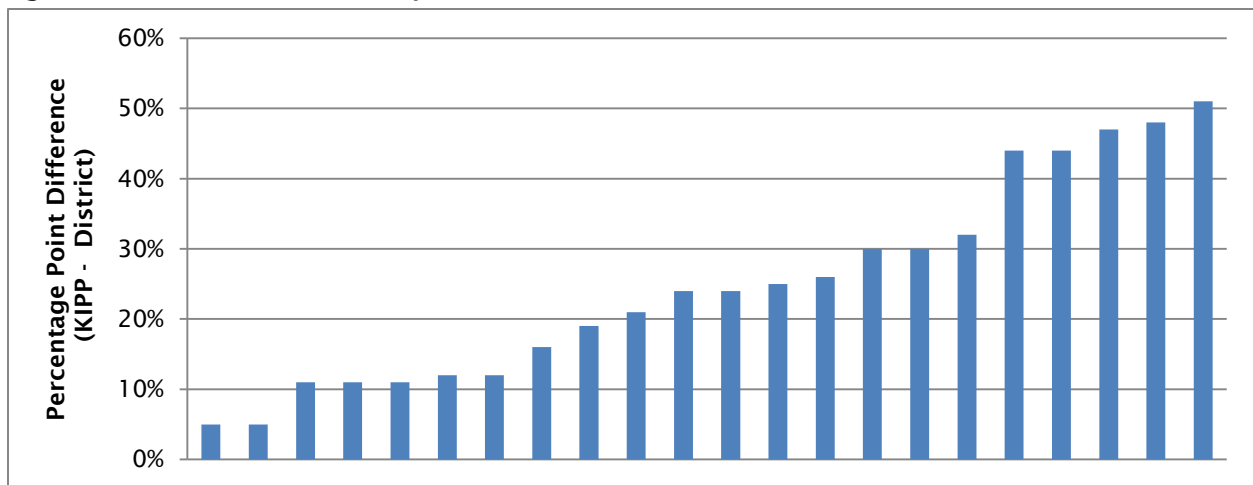
Note: All are statistically significant differences at the five percent level.

---

[6] Throughout this chapter, all differences described as "significant" represent disparities (calculated separately for each KIPP school) with t-test p-values below 0.05. All t-test significance calculations are two-tailed. Each chart used in this chapter presents the individual KIPP schools in a different order, typically ranked from lowest to highest on the given dimension or difference.

In all cases, the proportion of minority students is significantly elevated at KIPP schools compared to host districts. In almost all cases, the proportion of minority students at KIPP is at least 10 percentage points higher than the host districts. KIPP students are also more likely to be black or Hispanic than students at their elementary schools who did not go on to attend KIPP. The proportion of minorities at these feeder schools typically fell between that of the KIPP students and that of the district as a whole. In other words, KIPP's feeder elementary schools have elevated levels of black and Hispanic enrollment, and within these elementary schools, KIPP attracts an even higher proportion of minorities to enroll in its middle school network (see Appendix Table B.1).

Unlike the distribution of minority groups, gender balance at KIPP schools is almost always similar to the ratios found at traditional public schools. Out of the 22 KIPP middle schools in our sample, 19 had gender ratios that were not significantly different from host districts. In the three KIPP schools that did have significant differences, the schools enrolled more girls than boys; the highest observed female enrollment at a KIPP school was 62 percent.

Eligibility for FRPL is a useful proxy indicator for the level of poverty experienced by KIPP students and their families. Where FRPL data was available in obtained administrative records, we observed that a high proportion of KIPP middle school students were eligible for FRPL in fourth grade, and that these FRPL rates were often higher than those found in KIPP host districts. Specifically, over 70 percent of students are FRPL-eligible at 14 of the 15 KIPP schools for which we had access to that measure (Figure II.3). At a majority of these schools, the proportion of students with FRPL-eligibility was over 80 percent.

Figure II.3. Percentage of Students Eligible for FRPL in KIPP Schools
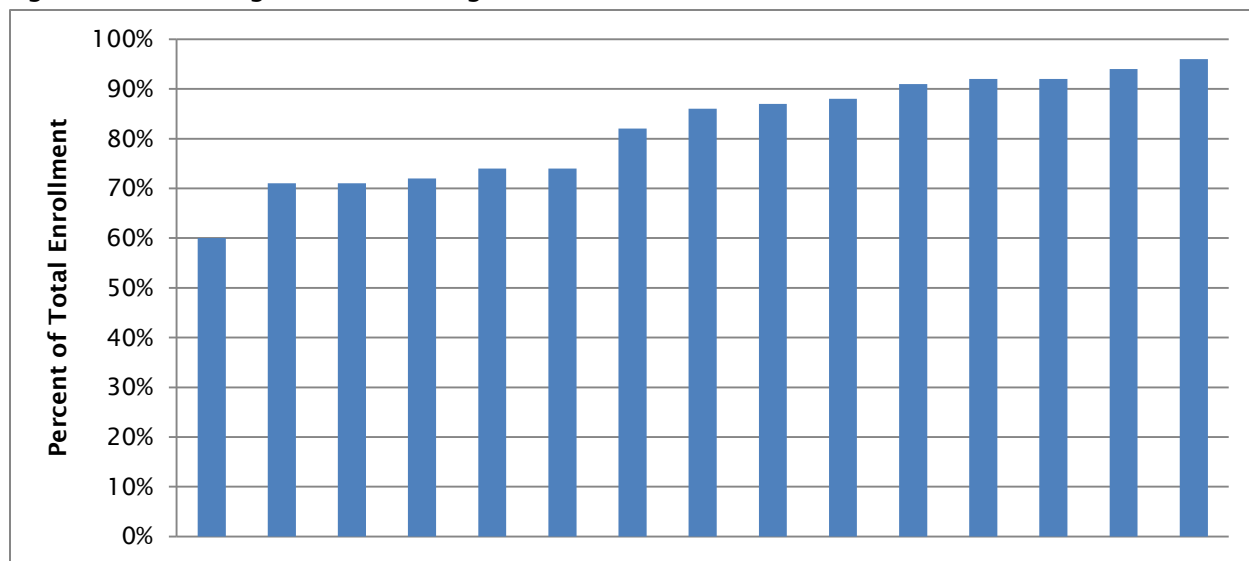
All but 4 of the sampled schools had FRPL-eligibility rates that were significantly higher than the levels found in the districts where KIPP is located, and none was significantly lower (Figure II.4). FRPL rates at KIPP's feeder elementary schools are also higher than districtwide rates in 13 of 15 instances, but this does not fully explain the high FRPL levels found at KIPP schools. In 11 of these cases KIPP students had significantly higher FRPL rates than students from the same elementary schools who did not attend KIPP; there were only two schools where the FRPL rates of KIPP students were significantly lower than the FRPL rates of students from feeder elementary schools (see Appendix Table B.2).

Figure II.4. Differences in FRPL Eligibility: KIPP vs. District
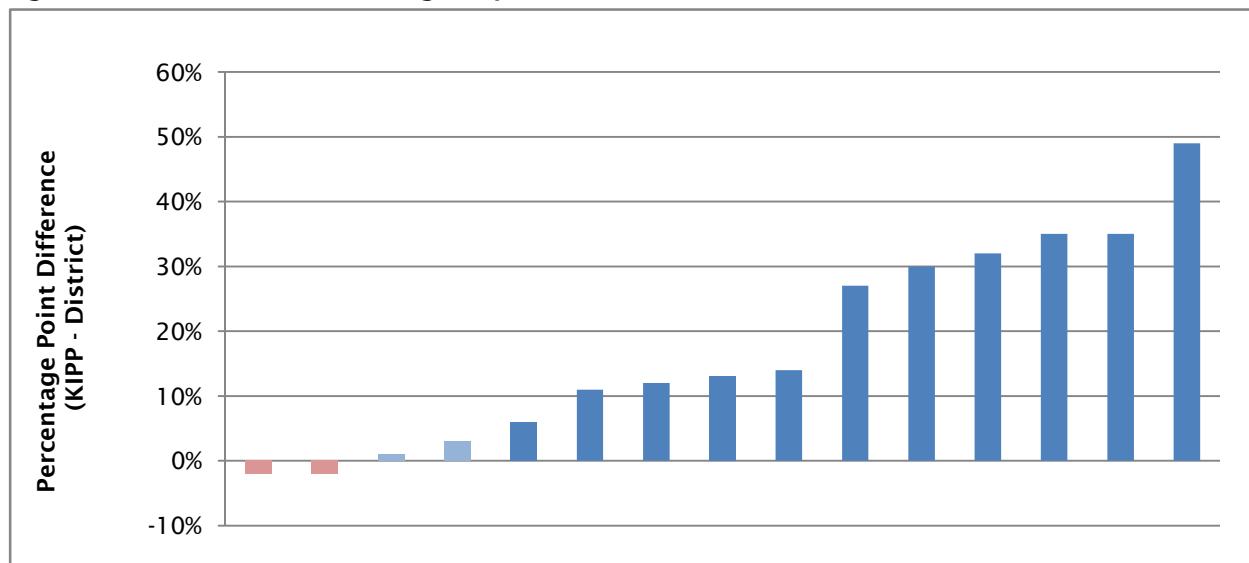
The baseline proportion of students in special education or with limited English proficiency tends to be lower at KIPP, as compared to both districtwide levels and also levels observed in KIPP's feeder elementary schools (a detailed table of average FRPL, special education, and limited English proficiency levels at KIPP schools, feeder elementary schools, and host districts can be found in Appendix Table B.2). Figures II.5 and II.6 compare the special education and limited English proficiency levels, respectively, of KIPP students to the districtwide population of middle school students. The proportion of students enrolled in special education was significantly lower in 12 out of the 22 schools in the sample. KIPP enrolled a significantly lower proportion of LEP students in 13 of 17 schools located in districts with data on this indicator.

Finally, we examined the baseline test scores of KIPP students as compared to their host districts and feeder elementary schools. Across the 22 middle schools in the sample, the pattern of baseline test scores varied greatly by school (the full results are included in Appendix Table B.3). A comparison of KIPP students' test scores to district test scores follows in Figure II.7.

Figure II.5. Differences in Special Education Status: KIPP vs. District
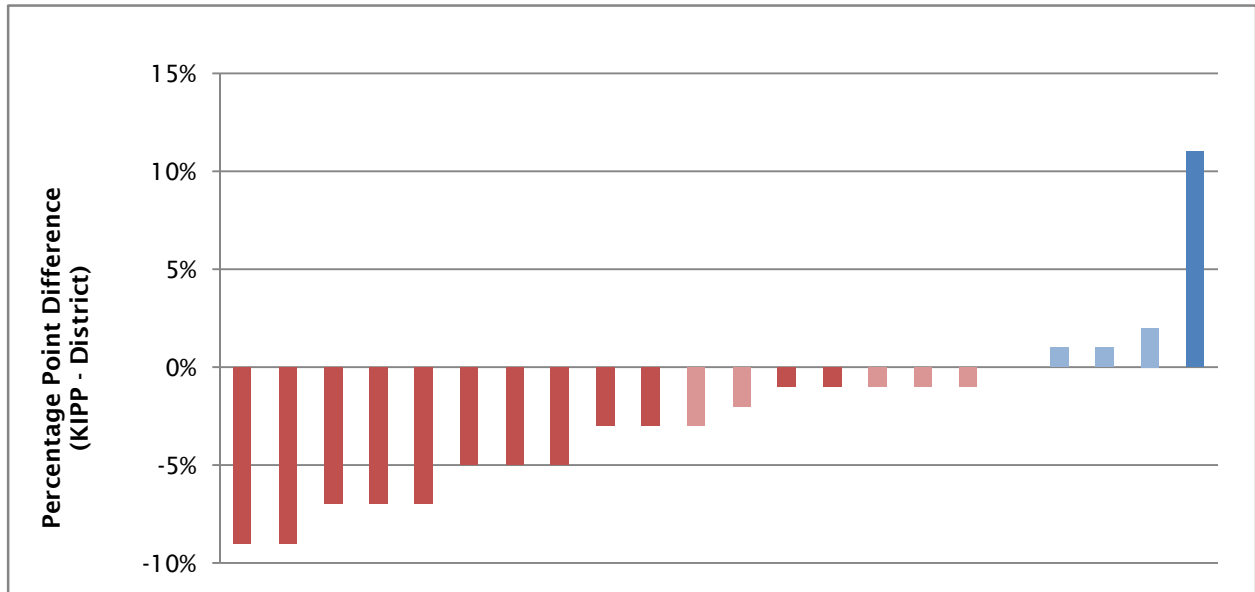
Note: The dark-red and dark-blue bars indicate differences from the district population that are considered statistically significant at the five percent level.

Figure II.6. Differences in LEP: KIPP vs. District
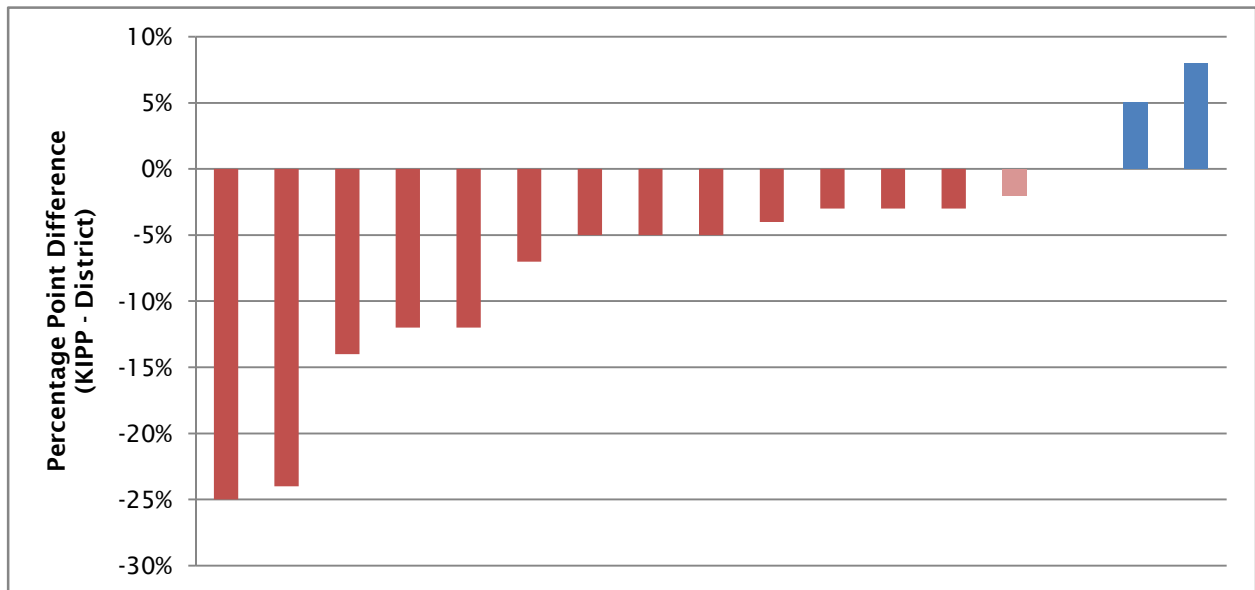
Note: The dark-red and dark-blue bars indicate differences from the district population that are considered statistically significant at the five percent level.

Figure II.7. Distribution of Baseline Math and Reading Differences: KIPP vs. District
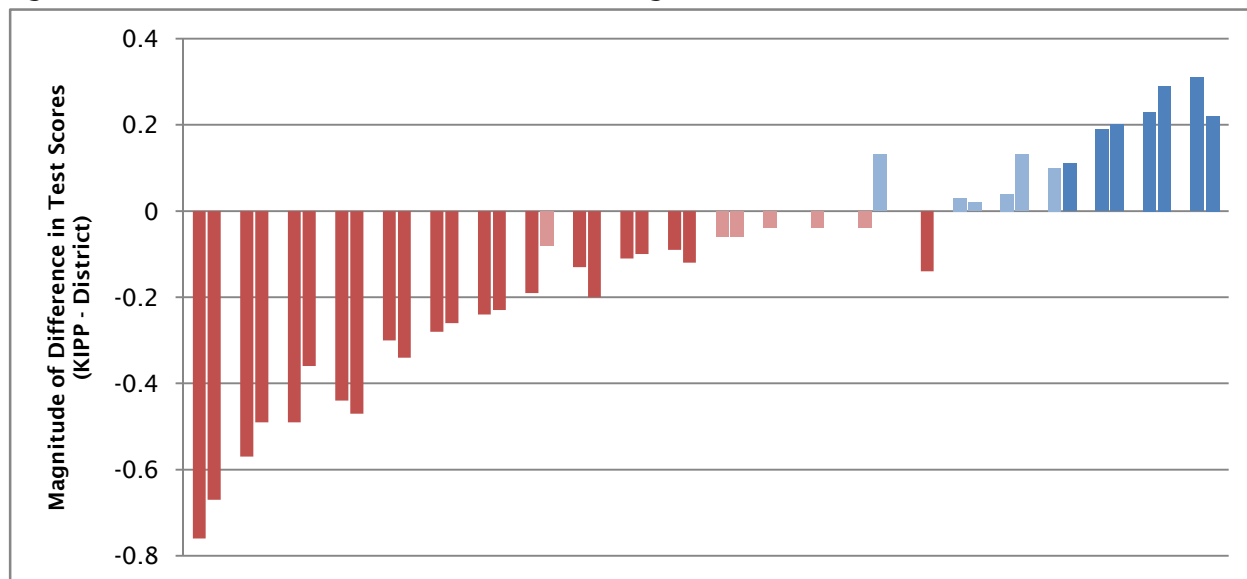
Notes: The left bar in each pair represents baseline differences in math scores for a given school, and the right bar represents differences in reading scores for that school. The dark-red and dark-blue bars indicate differences from the district population that are considered statistically significant at the five percent level.

Although results vary for different schools, KIPP schools in our sample were more likely to enroll students who were lower-achieving than district averages than to enroll higher-achieving students. Approximately half of KIPP schools attract students with significantly lower baseline test scores than the districtwide average (out of 22 sample schools, 11 schools enrolled students with significantly lower scores in reading, and 11 schools enrolled students with significantly lower scores in math). Only 4 schools attracted students with scores that were significantly higher than the district average in at least one subject. In several cases, KIPP students' reading and math test scores were over 0.4 standard deviations lower than students in comparison districts at baseline. Among the 4 KIPP schools enrolling students with higher average baseline test scores, scores were an average of roughly 0.2 standard deviations higher than non-KIPP students in the same districts. Restricting the comparison group to students attending the same elementary schools as KIPP students reveals a more evenly mixed pattern of schools enrolling both higher and lower scoring students. For baseline mathematics scores, for example, approximately the same number of KIPP schools enroll students with significantly higher scores (n=8) as the number that enroll students with significantly lower scores (n=7) within feeder elementary schools (see Appendix Table B.3).

In sum, the evidence suggests that prior to enrollment, KIPP students are more likely to score below the district average. Compared to their counterparts in feeder elementary schools, however, there is no consistent pattern of differences between the elementary school achievement levels of KIPP and non-KIPP students across schools in the sample.

C. Student Enrollment Patterns

Our final set of descriptive analyses examines student enrollment patterns at KIPP schools in terms of attrition and grade repetition. Attrition is an important potential pathway for student selection. If a KIPP middle 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 would bias the estimated KIPP effects if they were based only on the sample of students who remain.

To compare attrition rates at KIPP middle schools to other public schools, we defined attrition as school transfers (either in-district or out-of-district) occurring during or immediately after each grade served by KIPP. To measure the cumulative attrition rate between grades five and eight, we undertook a "life table" analysis, presented in Appendix Table B.4. Under this approach, we calculated a rate of attrition at each grade level for students and used these grade-specific attrition rates to derive the cumulative probability that a given student will change schools before completing eighth grade.[7] 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 fifth grade to a middle school in sixth grade.

Given KIPP's unique grade span, comparing the cumulative attrition rates in this manner may overstate the levels of attrition at KIPP relative to the rest of the district schools. While some proportion of non-KIPP students also attend schools serving grades five through eight inclusive (e.g., K-8, K-12, or other 5-8 schools), the majority attend an elementary school through fifth 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 fifth grade and moved on to sixth 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.

Over the entire course of middle school, cumulative attrition rates for students who enter in fifth grade at KIPP schools range from a low of 10 percent to a high of 76 percent. This represents a wide degree of variation in terms of the proportion of entering fifth graders who finish eighth grade at the same KIPP school. Attrition at traditional public middle schools, meanwhile, ranged from a low of 20 percent in one district to a high of 57 percent in a different district. As Figure II.8 demonstrates, there is also a great deal of variation when comparing attrition levels at KIPP middle schools to their respective district public schools.

There is no systematic pattern of higher or lower attrition rates at KIPP schools as compared to non-KIPP schools. Approximately one third of the sample KIPP schools (n=8) recorded attrition rates that were significantly lower than district rates in a majority of grades. However other schools showed the reverse pattern, with traditional public schools recording significantly lower attrition rates than KIPP schools (n=6) in a majority of grades. A considerable number of KIPP schools (n=7) also showed attrition rates that were not significantly different from traditional public schools. In total, these attrition analyses did not reveal a consistent pattern of differences between cumulative attrition rates at KIPP middle schools and district middle schools. Some KIPP schools have higher attrition rates, some have lower attrition rates, and others do not differ.

---

[7] Because many districts did not provide data on mid-year school transfers, these attrition calculations do not include student transfers that take place during the 8th grade school year.

Figure II.8. Distribution of Differences in Cumulative Middle School Attrition: KIPP vs. District
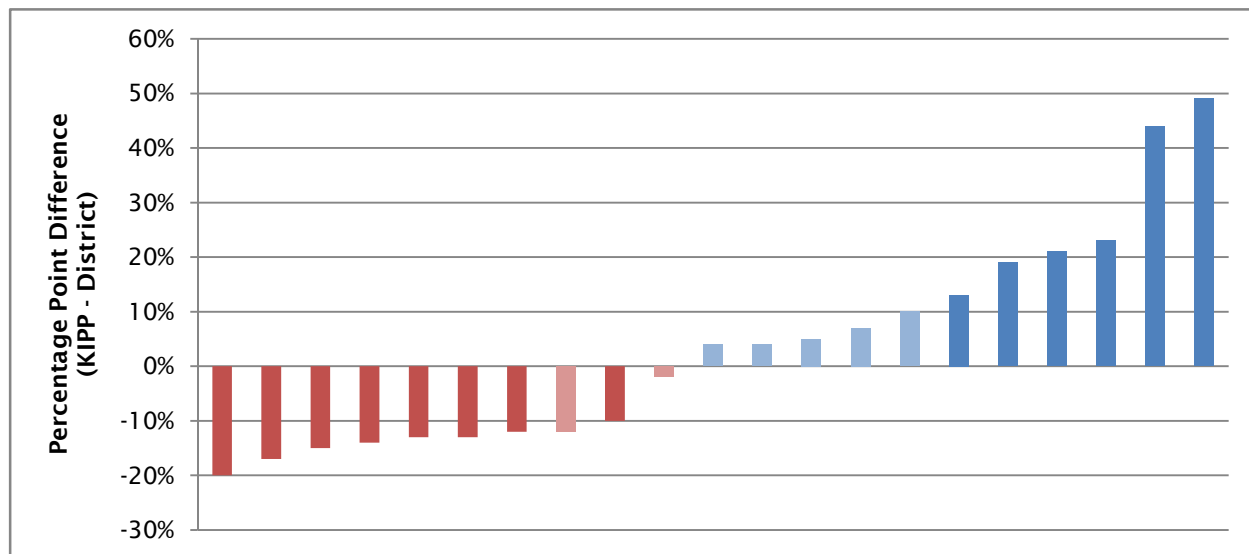
Note: All columns represent cumulative attrition rate differences derived from grade-specific attrition rates during middle school. Dark-colored bars indicate a majority of the grade-specific attrition rate gaps share the same sign and are significant at the five percent level.

To further investigate whether attrition represents a form of student selection, we compared the baseline test scores of students who transfer to those of students who stay at the same middle school through eighth grade. Almost all of the KIPP cumulative attrition rates were approximately evenly divided between within-district transfers and out-of-district transfers (see Appendix Table B.5 for a detailed comparison of these attrition types). Students who transfer within-district tend to have lower baseline test scores than students who do not transfer at all. For KIPP, the baseline scores of students transferring in-district were significantly lower at 12 schools (in at least one subject); none of the KIPP schools recorded higher baseline scores for students transferring in-district. The pattern at non-KIPP schools was even more pronounced: compared to those who do not transfer, students transferring in-district had baseline scores that were significantly lower in at least one subject in *all* 22 sites. For out-of-district transfers, the pattern is mixed. At KIPP, out-of-district transferring students at 17 schools had test scores that were not significantly different from those who stay. Within comparison districts, out-of-district transferring students had significantly lower baseline scores in 14 districts and significantly higher scores in 5 districts. Detailed comparisons of these scores can be found in Appendix Tables B.6 and B.7. Overall, these descriptive findings suggest that lower performing students are more likely to transfer schools before completing eighth grade; at KIPP, this is especially evident for in-district transfers. However, there is little evidence that the transfer pattern of low-performing KIPP students differs from the pattern at other public schools.

In contrast to the high levels of variation in KIPP attrition patterns, we did find a systematic pattern in the rates of grade repetition at KIPP middle schools. The KIPP instructional model generally holds that students should be promoted to the next grade only after they have demonstrated mastery of their current grade's material. Repeating a grade represents a different approach to addressing the needs of underperforming students, involving a dramatic expansion in instructional time and resources. Comparisons between KIPP grade retention rates and districtwide retention rates are reported separately by grade in Table II.2.

As the table demonstrates, KIPP middle schools retain students in grade at significantly higher rates than traditional public schools in the same districts. This pattern is especially evident in fifth and sixth grades, when KIPP middle schools first enroll students who had attended non-KIPP public schools. Retention rates range from 2 to 18 percent in fifth grade, and one to 12 percent in sixth grade at KIPP schools; other public schools in only one district retain students at rates higher than 4 percent in either year. The difference in grade repetition rates compared to district public schools is significantly higher in all but one school. Differences in grade repetition rates make it challenging to compare the achievement of KIPP students to district students using conventional, grade-specific standardized tests. We discuss our approach to addressing this issue in chapter IV.

D. Summary of Descriptive Findings

Across the descriptive analyses presented in this chapter, there is little evidence that KIPP middle schools are systematically enrolling more advantaged or higher achieving students from their districts. Consistent with the mission statement of KIPP, its schools serve a disproportionate share of low-income students relative to the other students in their districts. The KIPP schools in our sample also serve a proportionally high number of African Americans and Hispanics compared to other local schools. In contrast, these schools serve smaller numbers of limited English proficiency and special education students than do other district schools. The elementary-school achievement levels of students who enter KIPP middle schools vary. Half of the KIPP schools in our sample serve students who, on average, perform less well than their peers in fourth grade. A smaller number of schools serve those who perform better than their fourth-grade counterparts, and still other schools serve students whose performance shows no appreciable difference from this same group of peers. The enrollment patterns examined in this chapter also do not provide evidence that suggests KIPP schools benefit from the effects of student selection. We did not find systematically higher (or lower) levels of attrition across this sample of KIPP middle schools relative to host districts, though the pattern varied in different locations. Grade repetition rates, on the other hand, are consistently elevated at KIPP middle schools compared to district public schools.

Table II.2. Grade Repetition Rates by KIPP School and Grade Level

Grade 5 Retention		Grade 6 Retention		Grade 7 Retention		Grade 8 Retention	
KIPP (1)	District (2)	KIPP (3)	District (4)	KIPP (5)	District (6)	KIPP (7)	District (8)
.02 N=158	.01 N=1,162	.06 N=116	.01* N=750	.00 N=46	.00 N=360	N/A	N/A
.05 N=390	.01** N=29,622	.04 N=332	.02* N=24,513	.03 N=207	.01* N=24,521	.00 N=96	.01** N=18,996
.06 N=322	.03** N=18,188	.01 N=243	.03 N=18,570	.01 N=162	.06** N=19,280	.00 N=71	.05 N=19,532
.06 N=471	.01** N=19,977	.02 N=344	.01 N=15,247	.02 N=205	.02 N=12,452	.00 N=135	.02 N=9,440
.07 N=101	.02** N=146,168	.01 N=70	.06 N=158,560	.00 N=25	.02 N=149,729	N/A	N/A
.08 N=169	.01** N=16,880	.04 N=207	.02* N=15,613	.00 N=118	.01** N=11,463	.00 N=50	.02** N=8,032
.08 N=207	.03** N=209,608	.06 N=143	.02** N=207,850	.00 N=66	.03 N=214,501	N/A	N/A
.09 N=490	.02** N=326,176	.03 N=468	.02 N=311,406	.03 N=461	.02** N=315,513	.01 N=415	.01 N=322,409
.09 N=358	.03** N=361,586	.04 N=310	.02* N=358,328	.03 N=227	.03 N=368,155	.00 N=137	.02 N=371,796
.10 N=366	.01** N=16,931	.08 N=377	.02** N=15,588	.05 N=212	.01* N=11,425	.00 N=107	.02** N=7,985
.10 N=473	.02** N=326,164	.06 N=434	.02** N=311,378	.05 N=321	.02** N=315,493	.02 N=204	.01 N=322,381
.10 N=342	.03** N=439,652	.03 N=375	.02 N=436,450	.02 N=342	.03 N=445,107	.01 N=331	.02 N=447,513
.10 N=162	.03** N=209,612	.01 N=123	.02 N=207,854	.00 N=57	.03 N=214,503	N/A	N/A
.11 N=440	.02** N=146,207	.06 N=322	.06 N=158,549	.05 N=199	.02** N=149,686	.01 N=112	.01 N=144,224
.12 N=678	.03** N=40,684	.05 N=581	.02** N=39,480	.04 N=476	.05 N=34,674	.00 N=335	.04** N=29,694
.13 N=391	.02** N=78,195	.08 N=304	.02** N=79,396	.09 N=202	.03** N=80,536	.03 N=103	.02 N=82,115
.14 N=355	.01** N=9,654	.04 N=256	.02** N=9,877	.02 N=132	.02 N=10,181	.00 N=55	.02 N=10,416
.14 N=381	.02** N=81,101	.05 N=278	.01** N=75,445	.03 N=198	.01* N=76,503	.02 N=120	.01 N=81,371
.16 N=277	.03** N=101,820	.04 N=126	.01** N=91,796	.02 N=65	.02 N=97,904	.00 N=40	.02 N=100,867
.16 N=422	.00** N=23,924	.08 N=379	.01** N=23,232	.04 N=252	.01** N=23,940	.00 N=135	.01 N=24,335
.17 N=392	.01** N=69,725	.12 N=316	.01** N=66,287	.09 N=197	.01** N=67,699	.04 N=112	.01** N=69,894
.18 N=405	.00** N=23,919	.11 N=349	.01** N=23,242	.05 N=195	.01** N=23,955	.00 N=125	.01 N=24,341

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

* Statistically significant at the five percent level.
** Statistically significant at the one percent level.

III. EFFECTS OF 22 KIPP MIDDLE SCHOOLS ON STUDENT ACHIEVEMENT

The key to the causal (internal) validity of our research design is the extent to which it accounts for important differences between students who enter KIPP and students in the comparison group. Our analyses of the effects of KIPP make use of rich baseline data and employ the most rigorous nonexperimental methods possible short of having comparison groups of unsuccessful KIPP applicants. We use longitudinally linked student-level data and control for demographic characteristics and multiple baseline test scores. Moreover, during the baseline period, both the KIPP middle school students and the non-KIPP comparison group attend non-KIPP elementary schools. This allows us to produce estimates with a high degree of internal validity (if not as much as in a randomized experiment) for all KIPP middle schools where data are available—not just for the minority of KIPP schools where experimental analysis is possible. Indeed, these analyses include even a small number of KIPP middle schools from which the KIPP foundation had withdrew their KIPP affiliation and that subsequently closed within a year,[8] allowing us to observe performance in schools that might be viewed as KIPP's failures.

Implementation of this component of the study involved requesting and obtaining de-identified student-level records from the jurisdictions in which KIPP schools are located. Our key outcomes are performance on the state assessments in math and reading, typically administered annually in the spring of the school year. To facilitate making comparisons across schools in states with different tests, we standardize these test scores by subject, grade, and year using information from the entire sample of students from each KIPP district (or districts). For more information on this standardization, see Appendix A.

A. Estimation Strategy Overview

Our analysis of achievement impacts consists of two approaches, a districtwide ordinary least squares (OLS) approach and a matched comparison group approach. Both of these approaches take advantage of longitudinal, student-level achievement data to compare the achievement trajectories of KIPP students and comparison students from year to year. These "difference-in-differences" designs involve collecting achievement data on students in grades prior to KIPP entry and observing differences in subsequent achievement gains (controlling for other student characteristics) for KIPP students and non-KIPP students. Longitudinal data at the student level, with "pre-treatment" baseline scores at non-KIPP elementary schools for students who have not reached KIPP middle school grade levels, permits a much stronger nonexperimental design than would be possible with school-level, cross-sectional data. A student's prior test scores are the strongest predictors of later test scores, and can therefore serve as powerful controls against selection bias.

With each nonexperimental approach, we compare KIPP students who attended public elementary schools prior to entering KIPP to a counterfactual of students who also attended public elementary schools and remained in non-KIPP public schools in the same districts when they entered middle school. Both approaches involve estimating a set of ordinary least squares (OLS) models regressing test score outcomes on a series of treatment indicators indicating the number of

---

[8] A small proportion of KIPP schools are closed and 90 percent of KIPP schools remain in operation at present.

years since a student began attending a KIPP school. This allows us to examine the trajectory of estimated impacts for students over the entire course of KIPP "treatment"—from fifth through eighth grades. In addition, both analyses incorporate demographic controls, including indicators for gender, race/ethnicity, FRPL status, special education status, limited English proficiency, cohort,[9] year of test, outcome test grade level[10] and controls for baseline mathematics and reading test scores one and two years prior to KIPP entry.

The decision to control for two years of baseline of test scores was motivated by a falsification test we conducted. We estimated a model that predicts each student's test score in the year immediately *before* entering KIPP (based on demographics as well as test scores in the second year before KIPP entry). It also indicates if the student is a future KIPP entrant. A KIPP school should not affect students' achievement trajectories before they enroll, so we should not find systematic KIPP effects the year prior to enrollment unless there is selection bias. For most of the KIPP schools, we see no evidence of a spurious KIPP "effect" prior to entry (Appendix Table D.1). But for 2 of 22 schools in math and 2 in reading, we find that KIPP students appear to do better on tests prior to KIPP entry, after controlling for one year of prior scores. This suggests that—in a small number of KIPP schools—even among students who were similar in third grade, those who ultimately entered KIPP experienced a more positive trend in test scores in fourth grade than those who did not enter KIPP. The implication of this result is that accounting for just a single year of baseline test scores may not be sufficient in some sites to ensure that our treatment group of KIPP students is sufficiently similar to the comparison group of non-KIPP students in all ways other than their school attendance; controlling for a second baseline year of test scores to capture trends in students' performance prior to middle school may be important to truly make the KIPP and non-KIPP groups comparable.

The factor that distinguishes between our two general approaches is the group of comparison students used in each set of analyses. A detailed description of each approach follows.

1.   Districtwide OLS Approach

We refer to the first approach as a districtwide OLS approach, because it uses data from all other students in the local district (or districts) for comparative purposes. This is essentially the same approach used by Abdulkadiroglu et al. (2009)—a straightforward set of OLS regressions that control for observed demographic characteristics and baseline (prior to KIPP entry) test scores, using data from the entire population of tested students in each KIPP locality. The main benefit of the districtwide OLS approach (relative to the matched sample approach, discussed below) is that the large comparison group of students has the potential to increase the statistical power of the analysis and the corresponding precision of our impact estimates.

---

[9] Because date of birth is not available, we use an indicator for "year of initial fifth-grade entry" in the districtwide OLS approach and an indicator for "entry grade and year cohort" in the matched comparison approach to proxy as a cohort indicator.

[10] Demographic variables, test score variables, and years and grades of coverage vary from district to district. For details, refer to Appendix A.

2.   Matched Comparison Group Approach

One drawback of an analysis using all non-KIPP students in the district as a comparison group is that these students may be very different at baseline from students who enroll in KIPP. OLS models adjust for these differences statistically, but the success of the adjustments depends on assumptions about the underlying relationships between the different student characteristics and the achievement results. Our second (and ultimately preferred) nonexperimental approach aims to avoid relying on these assumptions by identifying a matched comparison group of students who are similar to the KIPP students in terms of observed demographic characteristics and prior achievement. This is a variant of the difference-in-differences matching strategy described in Smith and Todd (2005).

Our treatment group consists of any student who attended one of our sample KIPP schools at any point from grades five through eight, as identified in the administrative data.[11] The comparison group is carefully selected by considering all students across the district (in the appropriate grade and year) as potential comparison students, but retaining in the actual comparison group only those students whose characteristics and achievement during the baseline period (the period before the treatment group enters the KIPP school) match those of treatment group students. This approach can yield unbiased impact estimates if this comparison group closely matches the treatment group on these characteristics, and if these characteristics fully capture the relevant differences between the treatment and comparison groups (that is, there are no unmeasured differences between the two groups that are directly related to test scores during the follow-up period). Cook, Shadish, and Wong (2008) suggest that a closely matched comparison group can increase the likelihood that non-experimental estimators will successfully replicate experimental results.

More specifically, among cohorts of fourth and fifth grade students in each district/state, we estimated a logistic regression model that predicts whether a given student enters KIPP in fifth and sixth grade respectively. Models were estimated separately in each district/state and for different KIPP schools within the same district/state to account for differences in the available predictor variables and differences in the characteristics of students drawn to KIPP schools across sites.

We performed an iterative propensity score estimation procedure by identifying a list of baseline demographic and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. We then calculated propensity scores for KIPP entry using the variables identified by the abovementioned procedure and performed nearest neighbor matching (without replacement) of comparison group students to treatment group students separately by their KIPP entry grade and year. Table III.1[12] summarizes how well the means of baseline test scores and other covariates that entered the propensity score matching procedure are balanced across the treatment and control groups.

We required baseline test scores in both subjects to be balanced across treatment and control groups; in addition, at least 90 percent of the other selected model covariates, higher order and

---

[11] Students enrolled at district KIPP schools not included in this round of analysis (those opened after 2005) were dropped from the student sample.

[12] To both protect schools' confidentiality and satisfy the requirements of data use agreements, we do not identify individual school-level impacts or characteristics by name.

interaction terms must also be balanced across treatment and control groups.[13] In Table III.1 we show that through the use of propensity score matching, the comparison group in the matched analysis looks considerably more like the treatment group than the districtwide averages discussed in chapter II. Columns 1-4 of Table III.1 show that the mean baseline math and reading scores of KIPP students in each of the 22 schools are not significantly different from their matched comparisons at the five percent level. In addition, while the figures in chapter II show that KIPP students' demographic characteristics are different from all non-KIPP district students in the majority of districts/states, there are no significant differences between KIPP students' characteristics (as measured by the matching variables) and those of non-KIPP students in our matched comparison sample in 20 out of 22 KIPP sites.

After identifying the matched comparison group, we applied the same OLS model used in the districtwide analysis to control for the small remaining differences between the restricted sample of KIPP students and matched comparison students. This two-step method produced our preferred impact estimates.

The matched comparison approach has less statistical power than the districtwide OLS approach, because it relies on a smaller comparison group. The matched comparison approach is less likely to be subject to biases resulting from specification errors, however, because it does not draw information from comparison students who are very different from KIPP students. In practice, we found remarkably consistent results using the matched comparison (with OLS) and districtwide OLS approaches. In the main body of this report we present only the matched comparison results. The results of the districtwide OLS analyses are included in Appendix D for interested readers.

B. Addressing Threats to Validity

The matched comparison approach should provide estimates of KIPP effects that are methodologically strong, but there are several threats to the validity of these estimates that we must address. Below we describe how we handle three such threats: students moving from KIPP middle schools to other district schools, students who are retained in grade, and attrition from the sample.

1.   Attrition from KIPP Middle Schools

The combination of propensity-score matching and OLS is designed to address initial selection into KIPP; that is, differences between KIPP students and comparison students at the time of KIPP entry. But selection bias is also a potential problem in the departures of students who leave KIPP before completing eighth grade. As we saw in Chapter II, KIPP schools experience attrition of some of their students, and the levels of that attrition vary. If the students who leave KIPP schools early tend to be those who are not doing as well academically—a possibility that is confirmed by the data in Chapter II—then an analysis that includes only the persisting students would overstate KIPP's impacts.

---

[13] To ensure the covariates used in the propensity score estimation are balanced across the treatment and control groups, we compared the means of each covariate between the two groups to confirm that they were not significantly different from each other at the five percent level.

Table III.1. Balance Between KIPP Students and Matched Comparison Students

School	Baseline Reading Score		Baseline Math Score		Percentage of Matching Variables Unbalanced
	KIPP	Comparison	KIPP	Comparison	
	(1)	(2)	(3)	(4)	(5)
School A	-.62 (.04)	-.69 (.04)	-.70 (.04)	-.81 (.04)	0%
School B	-.42 (.04)	-.41 (.05)	-.39 (.04)	-.35 (.04)	0%
School C	-.41 (.04)	-.45 (.04)	-.45 (.04)	-.46 (.04)	0%
School D[a]	-.40 (.09)	-.35 (.10)	-.52 (.08)	-.49 (.08)	0%
School E	-.32 (.08)	-.41 (.08)	-.51 (.07)	-.58 (.08)	0%
School F	-.25 (.06)	-.18 (.05)	-.19 (.05)	-.14 (.05)	0%
School G[a]	-.17 (.08)	-.13 (.09)	-.24 (.08)	-.22 (.08)	0%
School H	-.17 (.05)	-.15 (.05)	-.09 (.05)	-.11 (.05)	0%
School I	-.09 (.04)	-.11 (.04)	-.14 (.04)	-.12 (.04)	0%
School J	-.09 (.04)	-.10 (.04)	-.05 (.04)	-.07 (.04)	0%
School K	-.08 (.04)	-.11 (.04)	-.08 (.04)	-.13 (.05)	0%
School L	-.01 (.05)	.02 (.05)	.07 (.04)	.16 (.04)	0%
School M	.00 (.07)	.06 (.07)	.02 (.07)	.07 (.06)	0%
School N	.01 (.05)	.00 (.05)	-.01 (.05)	.00 (.04)	0%
School O	.03 (.06)	-.01 (.06)	-.11 (.06)	-.15 (.06)	0%
School P	.06 (.06)	.04 (.06)	.10 (.05)	.10 (.06)	0%
School Q	.07 (.05)	-.02 (.05)	-.05 (.06)	-.09 (.06)	7%
School R	.14 (.04)	.20 (.04)	.30 (.04)	.31 (.05)	0%
School S	.19 (.05)	.21 (.05)	.19 (.05)	.25 (.05)	0%
School T	.21 (.07)	.22 (.07)	.02 (.07)	.00 (.07)	0%
School U	.21 (.05)	.17 (.05)	.19 (.05)	.14 (.05)	7%
School V	.29 (.05)	.25 (.05)	.24 (.05)	.31 (.05)	0%

Note: Column 5 reports the percentage of all variables, excluding baseline scores, used to generate propensity scores where the difference in means between KIPP students and control students is significant at the five percent level. Standard errors in parentheses.

[a] KIPP affiliation removed

We address this problem by ignoring early exits from KIPP schools: Any student who ever enrolled at KIPP remains permanently in the KIPP treatment group, regardless of whether the student remains in a KIPP school or transfers elsewhere.[14] In other words, a student who enrolled at KIPP at fifth grade in the 2002-03 school year but left KIPP after completing sixth grade at the end of the 2003-04 school year is included in the treatment group all four years he or she appears in the data (from 2002-03 through 2005-06 inclusive). This approach is analogous to an "intent to treat" analysis conducted in an experimental context.[15] By including all students observed attending a KIPP school, regardless of whether they leave or stay through eighth grade, we avoid the problem of overstating the effect of KIPP. Instead, this method is likely to produce a conservative estimate of KIPP's full impact on the students who continue attending KIPP schools. From the perspective of parents, students, or policymakers, this method appropriately accounts for the fact that not all students entering KIPP schools will in fact finish at KIPP schools.

We also conducted an alternate analysis that adjusts our benchmark estimates in an attempt to estimate the full effect of KIPP treatment on the students who remain enrolled in KIPP schools, described later in this chapter, with results reported in Appendix D.

2. Grade Repetition

As discussed in Chapter II, KIPP's commitment to high expectations of students does not encourage social promotion. KIPP expects students to meet their standards for being academically prepared for the next grade before they will be promoted. Consequently, KIPP middle schools retain students at significantly higher rates than other public schools in the same districts. Grade retention creates a complication for our analysis because the retained students are no longer taking the same tests as the other students in their cohorts. This is in essence a missing data problem. The scores on grade-specific state assessments of students who are retained cannot be directly compared with the scores of others in their cohort who have progressed to the next grade. With differential retention rates of KIPP students and comparison students, our impact estimates could be biased if we simply excluded retained students from the analysis. Instead, our benchmark analysis uses information from the past performance of grade repeaters.

For each grade repeater, in the year of repetition and subsequent years, we impute a score on the cohort-appropriate (rather than grade-appropriate) assessment that is equal to the student's standardized score in the last year prior to grade retention. In other words, we assume that each retained student does neither better nor worse than before retention. If KIPP in fact has a positive impact on the achievement of grade repeaters, this would cause us to underestimate KIPP's impact. If KIPP in fact has a negative impact on the achievement of grade repeaters, this would cause us to overestimate KIPP's impact. We believe the assumption of an unchanged score trajectory is a reasonable compromise, given the need to make some prediction about scores of grade repeaters.

---

[14] The only exceptions are students who disappear from our dataset entirely; this happens to some students in the comparison group as well as the treatment group. We address the issue in more depth below.

[15] Because it is not exactly the same (students who leave KIPP after a year are not the same as those who decline initial admission offers), to avoid confusion we do not use the "intent to treat" language in the context of the nonexperimental analysis.

Later in the chapter, we describe an alternate approach that makes more conservative assumptions about effects on grade repeaters.

3. Analytic Sample Attrition

Analytic sample attrition—as opposed to the observed attrition described in chapter II and above in Section B.1—is another issue we addressed in our estimation strategy. Students who did not have valid data for the year in which the outcome test score was measured were not included in the analysis. The study subjects may disappear from our analytic sample for several reasons including, but not limited to, dropping out of school; moving to a school outside of our data catchment areas; moving into our data catchment areas after the baseline period; moving to a private school; and having missing variable values. We categorized the above cases when students disappear from our analytic sample as out-of-district transfers. When we examined differences between KIPP and non-KIPP students in the cumulative rates of these out-of-district transfers, we found no obvious trend in either direction. The median cumulative rates of out-of-district transfers are 18 and 17 percent for KIPP and non-KIPP students respectively. For details, refer to Appendix Table B.5.

Because attending a KIPP school does not appear to be systematically related to the likelihood of leaving the district, this type of attrition is not likely to lead to bias in our estimates of the KIPP effect. It does imply, however, that the estimated KIPP effect will apply only to those students who remain in a district school (though not necessarily in a KIPP school).

In this study, a large proportion of sample members were missing baseline test scores, especially baseline test scores from two years prior. Like back-end attrition, missing data on the front end may pose a problem when missingness is correlated to treatment status or to student characteristics. To address this problem, we conducted stochastic regression imputation for missing baseline test scores separately by treatment status and used data sets with imputed baseline test scores.[16]

C. Results from Matched Comparison Models

Tables III.2 and III.3 summarize our preferred impact estimates in reading and math for students with one to four years of treatment in 22 KIPP middle schools. These estimates use the matched comparison approach (controlling for two years of prior achievement in both subjects and remaining differences in demographics), the benchmark approach for handling grade repetition, and data sets with imputed baseline test scores.

The included KIPP schools generally produce positive and statistically significant effects on students' test scores in math and reading, and substantially larger effects on math scores than reading scores. This pattern of producing larger effects on math than reading scores is consistent with that found in many randomized experimental studies of educational interventions, including the Boston charter schools study (Abdulkadiroglu et al. 2009), the Harlem Children's Zone Promise Academy study (Dobbie and Fryer 2009), the KIPP Lynn study (Angrist et al. 2010), the national evaluation of Teach For America (Decker et al. 2004) and the New York City charter schools

---

[16] Refer to Appendix C for more information on our imputation strategy, and Appendix D for results limited to the sample of students with non-missing baseline scores.

Table III.2. Test Score Effects in Mathematics, by Number of Years after KIPP Enrollment

School	Year 1	Year 2	Year 3	Year 4	N
School P	-.12** (.05)	.07 (.06)	.29** (.08)	.13 (.09)	448
School D[a]	-.05 (.07)	.03 (.07)	.11 (.08)	.13 (.09)	220
School G[a]	-.02 (.09)	.001 (.10)	-.03 (.10)	-.17 (.18)	220
School B	-.01 (.03)	.13** (.04)	.19** (.06)	.28** (.09)	735
School M	-.003 (.05)	-.02 (.09)	.10 (.12)	-.23 (.13)	329
School U	.02 (.05)	.17** (.06)	.15 (.09)	-.01 (.16)	532
School S	.03 (.04)	.28** (.06)	.33** (.08)	.23* (.10)	501
School L	.10** (.04)	.31** (.05)	.22** (.06)	.14 (.10)	565
School K	.15** (.03)	.33** (.05)	.48** (.06)	.45** (.08)	729
School I	.16** (.04)	.27** (.07)	.29** (.09)	.004 (.16)	540
School N	.19** (.04)	.17** (.05)	.30** (.07)	.22* (.09)	669
School O	.26** (.06)	.33** (.10)	.59** (.13)	N/A	409
School J	.29** (.04)	.52** (.04)	.56** (.06)	.62** (.08)	676
School A	.34** (.03)	.39** (.04)	.33** (.05)	.24** (.08)	688
School F	.35** (.05)	.57** (.06)	.74** (.09)	.52** (.13)	460
School E	.42** (.07)	.39** (.10)	.76** (.22)	N/A	294
School V	.49** (.05)	.60** (.05)	.48** (.06)	.47** (.08)	550
School T	.54** (.06)	.46** (.07)	.59** (.13)	N/A	307
School Q	.58** (.05)	.66** (.07)	.64** (.09)	N/A	442
School C	.59** (.04)	.59** (.04)	.66** (.06)	.66** (.08)	733
School H	.69** (.05)	.75** (.06)	.81** (.08)	N/A	395
School R	.76** (.04)	.73** (.05)	.75** (.06)	.60** (.09)	783

Note: This table reports the coefficients on a regression of standardized middle school math test scores on indicator variables for the number of years after a student's initial enrollment in a KIPP middle school. After grade repetition, students were assigned the same standardized test score received in the last year prior to retention. The comparison group consists of matched students who never enroll in KIPP; matching was conducted by cohort for students who enroll in KIPP in grade 5 or grade 6. Propensity scores for all students were generated using two years of baseline test scores and all available demographic characteristics. Demographic variables, test score variables, and years of data coverage differ slightly by district. For a detailed discussion of data set differences, see Appendix A. Regression controls include two years of baseline test scores in math and reading (baseline scores were imputed if one year of scores was missing) as well as dummy variables for demographic characteristics, the year-of-test, grade of test, and the year each student first entered fifth grade. Regressions use robust standard errors and are clustered on student identifiers.

* Statistically significant at the five percent level
** Statistically significant at the one percent level
[a] KIPP affiliation removed

Table III.3. Test Score Effects in Reading, by Number of Years after KIPP Enrollment

School	Year 1	Year 2	Year 3	Year 4	N
School U	-.14* (.06)	-.12* (.06)	-.08 (.08)	-.12 (.12)	526
School P	-.14** (.05)	-.03 (.06)	.12 (.07)	.18 (.13)	456
School B	-.05 (.04)	.003 (.05)	-.004 (.06)	.01 (.10)	735
School D[a]	-.04 (.07)	-.002 (.08)	.12 (.09)	.10 (.09)	221
School M	-.04 (.05)	-.04 (.08)	-.11 (.10)	-.15 (.12)	325
School G[a]	.03 (.10)	.02 (.09)	.08 (.11)	.18 (.21)	203
School L	.03 (.04)	.17** (.05)	.14* (.06)	.05 (.11)	570
School S	.04 (.03)	.20** (.05)	.28** (.05)	.17* (.08)	504
School N	.04 (.04)	.13** (.05)	.28** (.06)	.15 (.10)	669
School Q	.06 (.05)	.21** (.07)	.24** (.09)	N/A	442
School K	.06 (.03)	-.02 (.05)	.02 (.05)	.13 (.07)	727
School F	.08 (.05)	.24** (.07)	.39** (.08)	.36** (.12)	456
School T	.10 (.06)	.16* (.07)	.24 (.13)	N/A	312
School V	.11 (.06)	.34** (.06)	.33** (.09)	.25** (.10)	544
School J	.14** (.04)	.24** (.04)	.29** (.05)	.19** (.07)	667
School A	.16** (.03)	.29** (.04)	.28** (.05)	.33** (.07)	691
School I	.176** (.04)	.21** (.06)	.178* (.08)	.17 (.16)	546
School R	.18** (.04)	.22** (.05)	.28** (.06)	.36** (.09)	780
School O	.24** (.05)	.22** (.08)	.38** (.12)	N/A	408
School H	.30** (.05)	.43** (.06)	.91** (.12)	N/A	392
School C	.30** (.03)	.33** (.04)	.42** (.05)	.38** (.08)	733
School E	.425** (.06)	.31** (.08)	.421** (.15)	N/A	294

Note: This table reports the coefficients on a regression of standardized middle school reading test scores on indicator variables for the number of years after a student's initial enrollment in a KIPP middle school. After grade repetition, students were assigned the same standardized test score received in the last year prior to retention. The comparison group consists of matched students who never enroll in KIPP; matching was conducted by cohort for students who enroll in KIPP in grade 5 or grade 6. Propensity scores for all students were generated using two years of baseline test scores and all available demographic characteristics. Demographic variables, test score variables, and years of data coverage differ slightly by district. For a detailed discussion of data set differences, see Appendix A. Regression controls include two years of baseline test scores in math and reading (baseline scores were imputed if one year of scores was missing) as well as dummy variables for demographic characteristics, the year-of-test, grade of test, and the year each student first entered fifth grade. Regressions use robust standard errors and are clustered on student identifiers.

* Statistically significant at the five percent level
** Statistically significant at the one percent level
[a] KIPP affiliation removed

evaluation project (Hoxby at al. 2009). Reading scores have often been found more difficult to move than math scores, perhaps because reading skill is more heavily influenced by activities outside of school.

The impact estimates in Tables III.2 and III.3 suggest several key results across the 22 KIPP schools, which we summarize below.

Key Finding One: Impacts for large majorities of the 22 included KIPP middle schools are positive in both reading and math in all four years after students enter KIPP schools.

The overall pattern of estimates of the effect of KIPP is consistently positive (see Figures III.1 and III.2, which show the percentage of KIPP schools with positive, negative and not statistically distinguishable from zero math and reading impacts by number of years in KIPP). Within two years after entry, students are experiencing statistically significant, positive impacts in 18 of 22 KIPP schools in math and 15 of 22 KIPP schools in reading. Meanwhile, only two KIPP schools register a significant negative impact in reading in any year of treatment. The single school showing a significant negative impact in mathematics (in year 1) actually reverses that impact by year 3, when it becomes significantly positive.

These findings are consistent across a variety of alternative specifications. Strong positive estimates of KIPP's impact are evident in models using districtwide comparisons (with statistical controls) as well as those using matched comparisons; in models that define treatment status differently; in models that make different assumptions about the test scores of grade repeaters; and in models using imputation for missing data as well as those that eschew imputation. Results from the alternative models can be found in Appendix D and are discussed in Section D below.

Key Finding Two: The magnitude of KIPP impacts is often substantial.

First-year effect sizes in each KIPP school are charted in Figures III.3 and III.4. After a single year of KIPP treatment, 15 out of 22 schools show positive impacts in math ranging from 0.10 to 0.76 standard deviations. For reading, 8 out of 22 schools demonstrate positive impacts from 0.14 to 0.42 standard deviations.

To illustrate the cumulative effects of KIPP schools, we show estimated impacts after three years in Figures III.5 and III.6. The observed effects are especially large in math (which often shows larger impacts in schoolwide interventions; see, e.g., Hoxby, Murarka, and Kang 2009; Dobbie and Fryer 2009; Abdulkadiroglu et al. 2009). By year three, half of the KIPP schools in our sample are producing math impacts of 0.48 standard deviations or more, equivalent to the effect of moving a student from the 30th percentile to the 48th percentile on a typical test distribution. Compared to national norms during this grade span, a 0.48 effect size after three years represents 1.2 years of accumulated extra growth in mathematics over the three year period (Bloom et al. 2008). For comparison, the black-white test score gap in math is typically estimated as approximately one standard deviation at fourth grade and eighth grade (Bloom et al. 2008). Half of these KIPP schools are producing impacts large enough to cut that gap in half within three years.

Figure III.1. Percentage of KIPP Schools with Significant Effects in Math, by Year
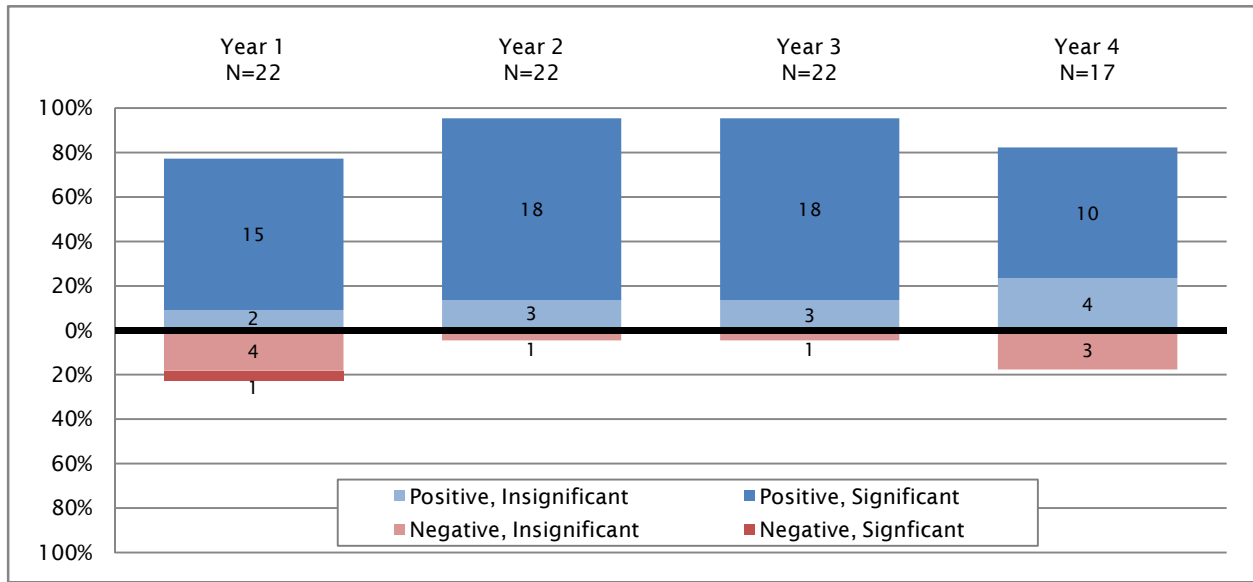
Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact is positive versus negative in a given year. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Figure III.2. Percentage of KIPP Schools with Significant Effects in Reading, by Year
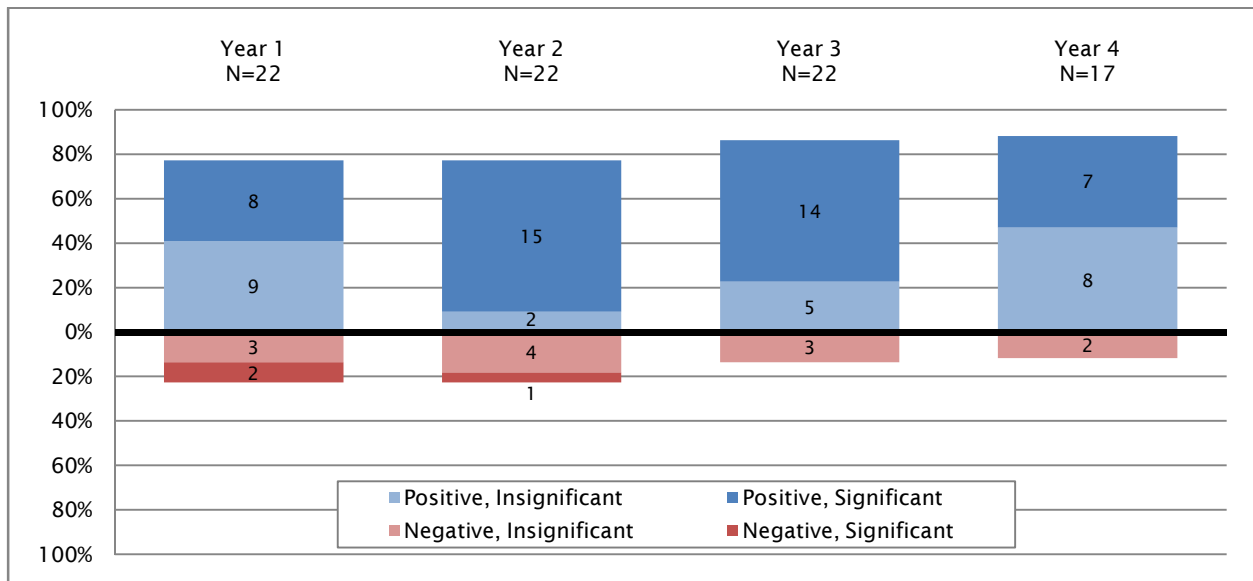
Note: Each bar represents the percentage of schools in the sample where the magnitude of the impact is positive versus negative in a given year. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Figure III.3. Distribution of Test Score Effect Sizes After One Year in Mathematics
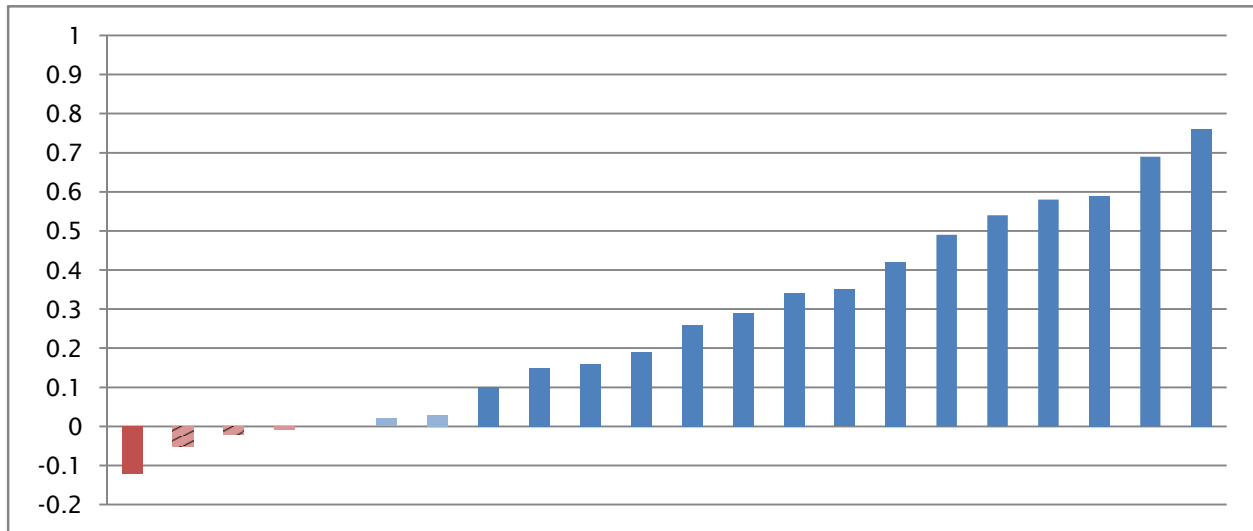
Notes: Bars with crosshatching represent schools that had their KIPP affiliation removed. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Figure III.4. Distribution of Test Score Effect Sizes After One Year in Reading
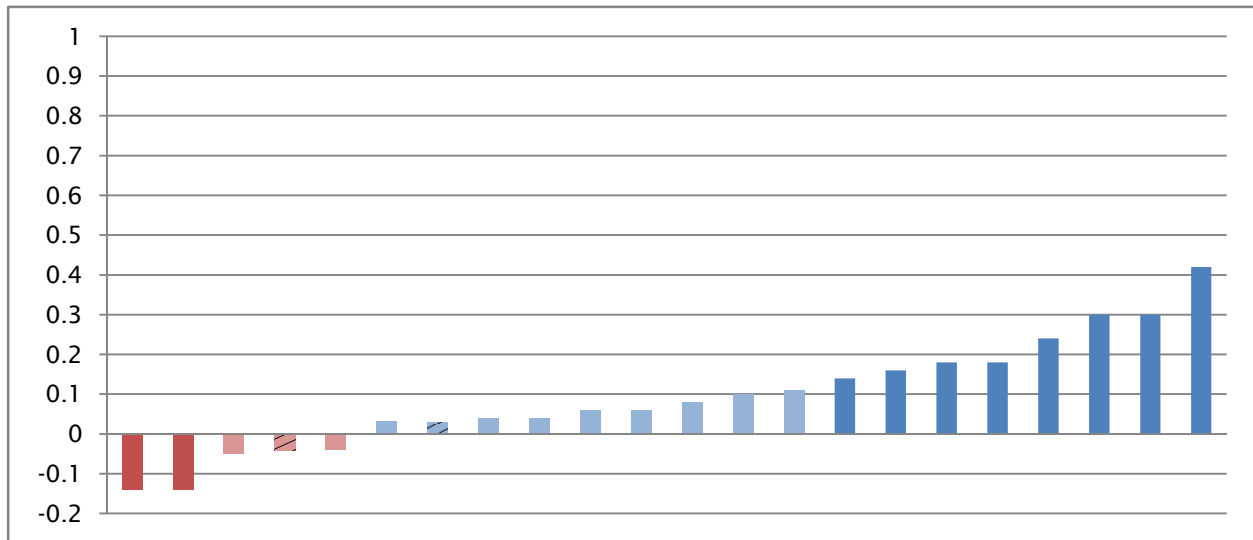
Note: Bars with crosshatching represent schools that had their KIPP affiliation removed. Dark-blue and dark-red colors indicate results that are considered statistically significant at the five percent level (where blue is positive and red is negative).

Three-year impacts in reading are also large in many KIPP schools, if not as large as the effects in math. Half of the KIPP schools in our sample show three-year reading effects of 0.28 standard deviations or more. An effect size of 0.28 standard deviations represents an estimated 0.9 years of additional instruction, or about one-third of the black-white gap in reading achievement in fourth grade and eighth grade (Bloom et al. 2008).

Figure III.5. Distribution of Test Score Effect Sizes After Three Years in Mathematics
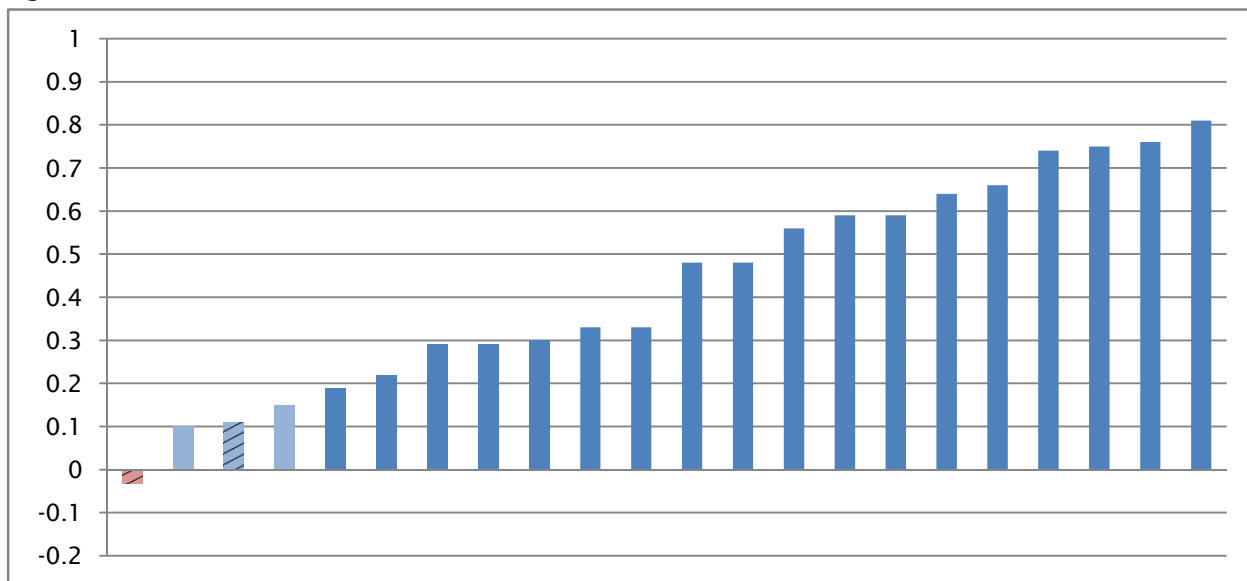
Note: Bars with crosshatching represent schools that had their KIPP affiliation removed. Effect sizes are considered statistically significant at the five percent level.

Figure III.6. Distribution of Test Score Effect Sizes After Three Years in Reading
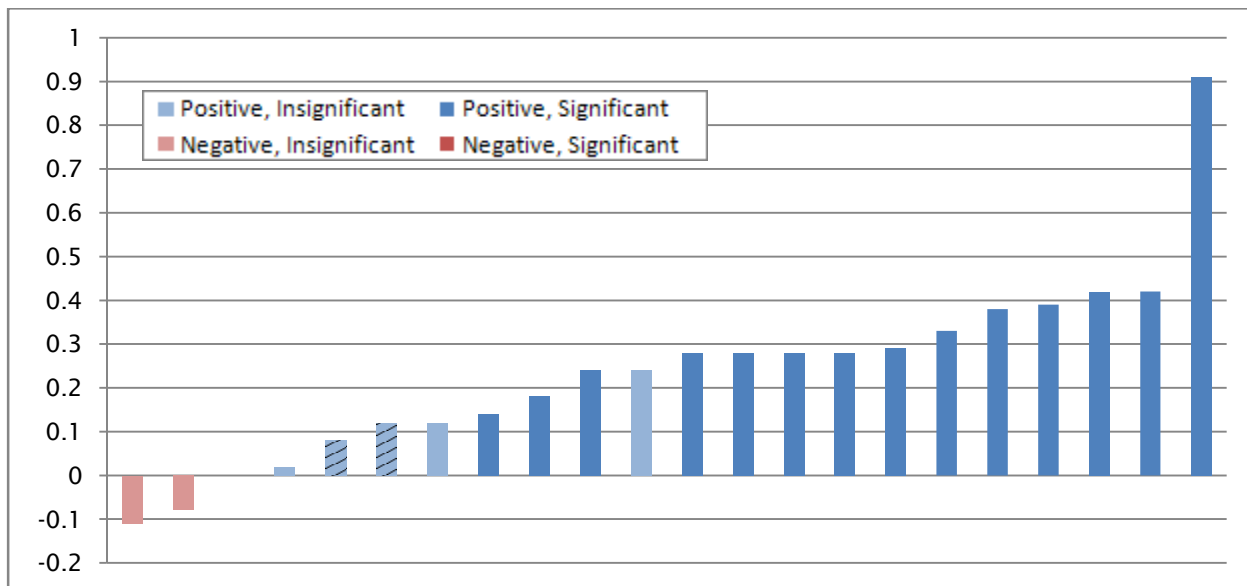
Note: Bars with crosshatching represent schools that had their KIPP affiliation removed. Effect sizes are considered statistically significant at the five percent level.

These effect sizes also compare favorably to the effects of other prominent and successful educational interventions. The achievement effects of class-size reduction are often used as a benchmark for other educational interventions. After three years of treatment (grades K-2) in classes one-third smaller than typical, average student gains amounted to 0.20 standard deviations in math and 0.23 standard deviations in reading (U.S. Department of Education, 1998)—less than the effect size achieved in 17 of 22 KIPP schools in math and 13 of 22 KIPP schools in reading.

More directly relevant are three recent studies of charter schools that found positive impacts. An experimental 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, Murarka, and Kang 2009). If accumulated over three years, these would amount to 0.27 standard deviations in math and 0.16 standard deviations in reading—substantially less than most of these KIPP schools are producing. The KIPP results more closely resemble the results from studies of Boston charter schools and the Harlem Children's Zone's Promise Academy charter schools—results that have been viewed by some observers as spectacular (see, e.g., Brooks 2009). Charter middle schools in Boston are estimated as producing annual achievement impacts of 0.09 in reading and 0.18 in math (Abdulkadiroglu et al 2009), which, if accumulated over three years, would put them squarely in the middle of the KIPP performance distribution (using a method that is less conservative than ours, and likely to produce larger impact estimates). The Promise Academy middle school produced three-year impacts estimated as 0.27 standard deviations in reading—near the middle of the KIPP distribution—and 0.77 standard deviations in math—near the high end of the KIPP distribution (Dobbie and Fryer 2009).

A variety of alternative analyses, the details of which can be found in Appendix D, produce results that likewise suggest large impacts. In addition, the estimates presented above are likely to underestimate KIPP's full impact on students who remain enrolled, because we use a conservative approach in which students who have withdrawn from KIPP schools after a year enrolled remain classified in the KIPP "treatment group." The full effect of KIPP attendance on students who remain enrolled through eighth grade is likely to be even larger than the estimates we report here. An estimate of that attrition-adjusted effect—which is likely to be biased upward, as described earlier—can be found in Appendix Tables D.3a and D.3b.

Additional Findings

In most KIPP schools, cumulative positive effects increase for at least the first three years after KIPP entry. In math, 18 of 22 KIPP schools show larger cumulative effects in year 3 than year 1, and in reading, 19 of 22 show larger cumulative effects in year 3 than year 1[17]. But the largest single-year impacts are often experienced in the first year, especially in math. The trajectory of effects for students who remain continuously enrolled in a KIPP school is often steeper than our benchmark estimates indicate (Appendix Tables D.3a and D.3b).

The two schools in our sample that lost their KIPP affiliation and that subsequently closed, KIPP Achieve Academy in Atlanta and KIPP Chicago Youth Village Academy, showed impacts that were noticeably weaker than those of the majority of KIPP schools that remain open. It is clear that these schools were not performing as well as most of the other KIPP schools. In both mathematics and reading, neither achieved statistically significant results for any of the years for which we could estimate their effects.

---

[17] These counts only consider the magnitudes of the point estimates and do not assess the statistical significance of the differences between year 1 and year 3 estimates.

For various reasons, impacts on select subgroups of KIPP students may be of particular interest. Using our sample from the matched analysis, we examined the following subgroups in turn: higher versus lower-performing students on test scores at baseline; LEP students; male students; racial/ethnic minority students (black or Hispanic); and minority male students (see Appendix E). We did not find clear patterns suggesting that KIPP impacts for any of these subgroups differed systematically from average impacts for all KIPP students.

D. Results from Alternate Models

Below, we briefly describe the results we obtained from estimating alternative models that handle the three threats to validity we discussed in Section B. The results from all alternate models closely resemble those from our benchmark model presented in Section C above. Detailed tables with estimates for all alternate models—including districtwide OLS analyses as well as the three variants on the matched analysis described below—can be found in Appendix D.

1. Effects on Students Who Remain Enrolled in KIPP Schools (Attrition Adjusted Model)

To estimate the effect of KIPP on students who remain enrolled, we applied the same procedure (known as a Bloom adjustment) that is used to convert "intent to treat" effects to "treatment on the treated" effects in experimental studies. This procedure assumes that students who withdraw from KIPP schools experience no continuing effect of their prior enrollment at KIPP. This is a very strong assumption, and if it is not true—as seems likely—then the adjusted estimates will typically overestimate KIPP's full effect on students who remain enrolled. Detailed estimation results can be found in Appendix Tables D.3a and D.3b. In fact, while the estimated impacts from this alternate analysis are generally larger in magnitude than those of our benchmark model, they are very similar overall. We report these results only in an appendix because they are likely to overestimate true impacts. True KIPP effects on students who remain enrolled probably fall somewhere between our preferred estimates and the adjusted estimates in the appendix.

2. Alternate Grade Repetition Model

Appendix D also includes the results from an alternate, more conservative approach to handling grade repeaters, in which we assigned them the test score of a student at the fifth percentile of the analysis sample for that school in the grades they would have attended under a "normal" grade progression for all years after grade repetition was first observed. This method is likely to underestimate KIPP's impacts, but given the uncertainty about true effects on grade repeaters, it is useful to test results under a conservative assumption.

When grade repeaters were assigned the test score of a student at the fifth percentile of the analysis sample for that school in the grades they would have attended under a "normal" grade progression, we find that estimated KIPP impacts, while generally diminished, remained comparable to those of our benchmark model. In math, for example, 13 instead of 15 schools exhibited positive and statistically significant year 1 impacts. The alternate model did not change the number of schools (18) that demonstrated positive and statistically significant math impacts from three years of KIPP treatment. In reading, while the same number of schools (8) exhibited positive and statistically significant year 1 impacts, 15 instead of 14 schools exhibited positive and statistically significant impacts by year 3. Detailed estimation results can be found in Appendix Tables D.4a and D4.b.

3.   Non-imputed Baseline Test Score Model

Results using non-imputed baseline data did not differ substantially from results with imputed data. Detailed estimation results can be found in Appendix Tables D.5a and D5.b.

IV. NEXT STEPS

The findings presented in the previous chapter provide compelling preliminary evidence that KIPP schools have a positive effect on their students' academic performance. Do KIPP schools that were not included in this analysis also have positive effects on their students? Do the positive effects of KIPP hold up when we use other, possibly more rigorous, estimation methods? Does KIPP also have positive impacts on outcomes other than test scores? In this chapter, we present our approach for addressing these questions through the remainder of the evaluation. We also present our schedule for reporting future findings from the evaluation.

A. Expanding the Nonexperimental Analysis

This report presents estimates of KIPP effects on students' scores on mathematics and reading achievement tests based on a nonexperimental analysis of 22 KIPP schools. This represents a broader set of KIPP schools than has been included in previous studies of KIPP impacts, but still represents fewer than half of all KIPP middle schools—those for which data were most readily available. Our study ultimately aims to estimate average achievement impacts across as broad a set of KIPP middle schools and students as possible. In future reports, the number of KIPP middle schools, cohorts of students, and years of follow-up included in the nonexperimental analyses will increase substantially.

We also plan to conduct work designed to assess the validity of the nonexperimental estimates of KIPP effects using experimental methods. This will involve attempting to replicate the nonexperimental estimates of KIPP effects with lottery-based results (described below), in the subset of KIPP sites where both approaches are possible. We will follow the path of other replication studies that start with an experimental study that has been well-implemented and can be credibly believed to have produced unbiased impact estimates. We will then apply our nonexperimental design to estimate the same impact parameters (ideally using the same sample of treatment students). If the nonexperimental impact estimates match the experimental impact estimates, we will have evidence that the nonexperimental design can produce unbiased impact estimates. Successful replication implies that outcomes among the experimental control group and the nonexperimental comparison group should not be significantly different.

B. Experimental Analyses of KIPP Impacts

The centerpiece of our overall KIPP evaluation is an experimental impact analyses at all eligible KIPP middle schools using their admissions lotteries as the random assignment mechanism. The lottery-based method is the most rigorous way to produce causally valid estimates of charter-school impacts, and it will be used as a benchmark to assess the validity of nonexperimental methods that can be applied to the full population of KIPP schools in the study. We can be confident that a well-conducted randomized control trial (RCT) will generate unbiased impact estimates.

The key requirement for using an admissions lottery in an experimental analysis of a school's impact is that there must be a group of applicants to the school for whom the probability of enrollment in the school is determined by a random draw. KIPP schools that have more applicants than open seats (by the time of their application deadline) hold randomized lotteries to determine which applicants will be offered admission to their school. Thus, any differences in later outcomes between those lottery participants who won the lottery and were offered admission and those who

lost the lottery and were not offered admission can be attributed to the impact of the admission offer to the KIPP school.

1. "Intent to Treat" vs. "Treatment on Treated" Impact Estimates

The estimated impact of KIPP based on a comparison of outcomes between lottery winners and lottery losers is referred to as an "intent to treat" (ITT) estimate. It represents the average impact of receiving an offer of admission from the school, regardless of whether the student actually enrolls in and attends the KIPP school. The ITT estimate represents the most rigorous causal impact estimate, and requires relatively few assumptions. The ITT estimate also can provide a window on the likely policy impact of making KIPP schools available. Because KIPP schools are schools of choice, policymakers cannot compel students to attend, so an estimate that accounts for the effect of offering admission can be useful.

Policymakers as well as parents are likely to be interested, however, in the effect of actually enrolling in and attending a KIPP school. This estimate, known generally as the effect of "treatment on the treated," or TOT, is derived from the ITT estimate. In addition to calculating ITT estimates of the KIPP impact, we will conduct a second stage of the analysis that adjusts the ITT estimate to account for non-compliant students; that is, those who win the lottery and are offered admission to a KIPP school but who do not attend that school.[18] Estimating impacts with the TOT approach requires assumptions that are not required by the ITT model, but if these assumptions hold, the TOT approach yields rigorous, internally valid estimates of the impact of attending KIPP.

In our primary experimental impact analysis, we will produce both ITT and TOT estimates of the impact of KIPP. In other words, we will estimate the impact of being offered admission to KIPP, along with the impact of actually enrolling in and attending KIPP.

2. Key Outcomes in the Experimental Analysis

In addition to producing rigorous ITT and TOT estimates of the KIPP effect on students' scores on state mathematics and reading tests in the experimental study, we will be able to examine other outcomes. Because we identify students at the time they apply to KIPP schools and then follow them over time as part of this study, we will have an opportunity to collect a broader range of data than was possible for the nonexperimental analysis presented in this report (or than has been presented in retrospective lottery-based studies). In particular, we are engaging in a variety of data collection activities:

- **Parent and Student Surveys.** At baseline (prior to the lottery), we surveyed parents of applicants to obtain information about student and family characteristics, school of attendance at baseline, and the decision to apply to a KIPP school. In the spring of the second follow-up year after the lottery, we are surveying both parents *and* students to

---

[18] In theory, noncompliant students (sometimes referred to as crossovers) may also include those who lost the lottery but ended up attending the KIPP school anyway. So long as schools follow their lottery outcomes, this type of noncompliance should be quite unusual in our main experimental analysis. In an alternative experimental approach we describe below, however, this type of crossover is more common. Our approach for estimating TOT KIPP impacts accounts for noncompliance of either type.

obtain outcome data on student satisfaction, motivation/engagement, and behavior, as well as parent satisfaction and involvement with their children's education.

- **Administrative Records and State Assessment Scores.** We will obtain data from states and districts (and schools where necessary) on test scores and attendance at baseline and annually for four follow-up years (with an option for a fifth year). For students entering KIPP at fifth grade, this will cover the entire grade span of the typical KIPP middle school (serving grades 5 through 8).

- **Principal Surveys.** Principals of the schools (both KIPP and non-KIPP) attended by students in the sample will be surveyed during the follow-up period to provide contextual information about school characteristics, students, and staff to help us interpret impacts.

- **Study-Administered Test.** We will administer a study-specific test to students in their third follow-up year in order to assess higher-order thinking skills as measured by a common instrument across all students and states.

3. Alternatives to the Experimental Analysis

One limitation of this experimental analysis will be that it will be possible only in those KIPP schools that hold admissions lotteries resulting in waiting lists of lottery losers who are not subsequently offered admission to the school.[19] We would like to produce rigorous impact estimates for a broader set of KIPP schools. Nonexperimental impact estimates such as those presented in this report are one way to examine KIPP schools not included in the experimental analysis, though this approach allows us to examine only student achievement on state assessments. Thus, we have developed two alternatives to the primary experimental design based on samples of applicants to KIPP schools for whom we would be able to collect the full range of outcome data.

a. Alternate Experimental Approach

An alternate experimental (AE) approach was used in the recent study of Boston charter schools conducted by a team of researchers at Harvard and MIT (Abdulkadiroglu et al. 2009) and is being used in Mathematica's National Study of Charter School Management Organizations. The AE approach distinguishes treatment and control groups for ITT purposes based on each applicant's position on the randomized admission list at the time of the lottery. With this method, only students who are *initial* lottery winners—receiving an offer at the time of the lottery—are defined as those intended to be treated. The remainder of the lottery participants—those who are not offered admission at the time of the lottery—are assigned to the control group, even though some of them may receive offers of admission at a later date. With this method, some crossover can occur in both directions. Some initial lottery winners may decline to enroll, while some initial lottery losers may later be admitted from the waiting list and ultimately enroll.

---

[19] Some schools hold admissions lotteries and form initial waiting lists of lottery losers, but then subsequently offer admission to all of these lottery participants on the waiting list as they seek to fill seats when some lottery winners choose not to enroll in KIPP.

The alternate approach can produce an estimate of the school's impact as long as the enrollment rate of applicants initially put on the wait list is sufficiently lower than the enrollment rate of initial lottery winners. If so, the student's randomly determined position on the lottery list affects the probability of enrollment—the key requirement noted above. In practice, we would expect that this condition will often hold. Students who receive late offers are presumably more likely to have made other plans and therefore to decline these offers than are students who receive offers at the time of the lottery.

In essence, the benchmark experimental method and the alternate experimental method are distinguished by the time the intention to treat is determined. For the baseline method, the ITT estimate identifies the effect of receiving an offer of admission at any time. For the alternate method, the ITT estimate identifies the effect of receiving an offer of admission at the time of the lottery (that is, the effect of being an initial lottery winner).

Shifting the time at which intention to treat is determined (and the corresponding interpretation of the ITT analysis) can be useful in the situations described above, depending on the difference in enrollment rates between initial lottery winners and students admitted later. When a school has offered admission to all of the initial lottery losers, the alternate approach permits the possibility of estimating the school's impact experimentally, even though doing so would be impossible with the baseline method. In schools that have offered admission to most but not all of the initial lottery losers, the alternate approach has the potential of increasing the statistical power of the experimental estimate.

b.   Stratified Applicant-Based Design

The stratified design (SD) compares outcomes among a treatment group of students who applied to KIPP and were offered admission with a comparison group of students who also applied to the school but received a lower preference in the school's lottery process. Several KIPP network schools stratify their lotteries, allowing them to give preference to certain groups of students (for example, those residing in-district or those eligible for free or reduced-price lunch) while still employing a fair and impartial admissions mechanism. In this way, schools will randomize all students within one stratum, randomize the students within the next stratum, and so on. In an experimental design, we can include only the students within the single stratum in which the cut-point occurs (that is, where there are students within the stratum who both have been made offers and those who have not), since students in other strata are different along some observable dimension—the strata criteria—that may affect impacts.

However, particularly in the case of in-district versus out-of-district applicants, we can make an argument that, on average, the students are similar, given that both sets of students were sufficiently interested in KIPP to submit an application. Therefore, applicants from other lottery strata represent a potential source of additional data that we can use to supplement the experimental sample. Under this SD, we include in the treatment group strata in which all the students are made offers and in the control group strata in which no students receive offers.

The important caveat concerning this approach is that while the stratum criteria solely determined whether a student was admitted (given that they applied), these criteria (for example, students' residential location) could be correlated with other factors that differentiate the two groups. For example, the socioeconomic status (SES) levels in the two districts may be different and the quality of the elementary schools they attended may also differ. In fact, their level of underlying

motivation may also be different, since a student would presumably have to be more motivated to apply to a school located in a different district than to apply to a school in their own district. In estimating the impacts of KIPP using a stratified design, we will control for observed differences in the characteristics of students in the different strata such as SES and baseline achievement measures. However, we cannot be certain that we will be able to control for all such differences—there may be unobserved differences between students in different strata.

4. Sample for the Experimental Design and Its Alternatives

We recruited KIPP schools and their applicants to fifth and/or sixth grades across two school years: for entry in fall of 2008 (cohort 1) and entry in fall of 2009 (cohort 2). Sixteen KIPP middle schools—representing a cross-section of states, student characteristics, and regional structure—were eligible for the RCT, the AE, and/or the SD. Table IV.1 presents the schools participating in one of these study designs, based on samples of applicants to KIPP schools. The ultimate sample of applicants to participating schools for whom we obtained parental consent comprises 23 school/grade/year combinations in 16 unique schools for a total of 1,661 students.[20] The student-level characteristics of this sample will be described in the interim report.

C. Reporting Schedule

To provide timely but comprehensive reporting of findings, we plan to release a series of three major reports from this study. In addition to this preliminary report, we expect to issue an interim report (in mid-2012) and a final report (expected late 2014).[21] Provided below is a brief description of the anticipated content of these reports.

1. Interim Report

- Estimated impacts from the RCT design, as well as the AE and SD designs, after two years of followup, including (1) state assessments, (2) survey-based outcomes, and (3) performance on the common test of higher-order thinking skills

- Estimated impacts from the nonexperimental analysis for every KIPP middle school operating as of 2007 for which we are able to obtain sufficient data

- Results of the exercise to validate our nonexperimental methods using experimental estimates

2. Final Report

- Estimated impacts from the RCT, AE, and SD designs, after four years of followup

---

[20] One caveat to this projected sample size of schools is as follows: the student-level counts represent the number of applicants with consent. Although we aim for high response rates in all of our follow-up data collection efforts, they can be unpredictable. The sample we use in the analysis will include only students for whom we obtain valid outcome data. In addition, the number of schools that will be included in the analyses for a given outcome is dependent on having respondents from both the treatment and control groups on the instrument (for example, administrative records versus student-level survey).

[21] If the fifth-year option for the evaluation is exercised, the final report will be released in 2015.

- Estimated impacts from the nonexperimental analysis for additional KIPP middle schools and cohorts

Table IV.1. Applicant-Based Sample

School	State	Cohort	Design	Grade	Sample Size	Consent rate
KIPP Aspire Academy	TX	2	RCT	6	21	66%
KIPP Austin College Prep	TX	2	RCT	6	57	89%
KIPP DC: KEY Academy	DC	2	RCT	6	33	49%
KIPP DC: WILL Academy	DC	2	AE	6	17	74%
KIPP Academy New York	NY	2	RCT	5	18	67%
KIPP Academy New York	NY	2	SD	5	164	79%
KIPP Academy Middle School	TX	1	RCT	5	51	69%
KIPP Academy Middle School	TX	2	RCT	5	151	86%
KIPP Academy of Opportunity	CA	2	RCT	5	21	72%
KIPP Academy of Opportunity	CA	2	SD	5	61	84%
KIPP Academy of Opportunity	CA	2	RCT	6	66	69%
KIPP L.A. Prep	CA	2	RCT	6	111	90%
KIPP Academy Lynn	MA	1	RCT	5	71	37%
KIPP Academy Lynn	MA	2	RCT	5	98	45%
KIPP AMP Academy	NY	2	AE	5	33	77%
KIPP Philadelphia	PA	1	RCT	6	56	74%
KIPP Philadelphia	PA	2	RCT	6	68	81%
KIPP South Fulton Academy	GA	1	RCT	6	66	88%
KIPP South Fulton Academy	GA	2	RCT	5	83	75%
KIPP South Fulton Academy	GA	2	RCT	6	94	80%
KIPP Summit Academy	CA	1	RCT	5	55	72%
KIPP Summit Academy	CA	2	RCT	5	69	86%
KIPP Summit Academy	CA	2	SD	5	46	88%
KIPP TRUTH Academy	TX	2	RCT	6	16	76%
KIPP Tulsa College Prep	OK	1	RCT	5	57	78%
KIPP WAYS Academy	GA	2	RCT	6	21	91%
KIPP WAYS Academy	GA	2	SD	6	57	64%
TOTAL					**1661**	**71%**

REFERENCES

Abdulkadiroglu, Atila, Josh Angrist, Sarah Cohodes, Susan Dynarski, Jon Fullerton, Thomas Kane, and Parag Pathak. "Informing the Debate: Comparing Boston's Charter, Pilot and Traditional Schools." Boston, MA: Boston Foundation, January 2009.

Aiken, L.S., S.G. West, D. Schwalm, J. Carroll, and S. Hsuing. "Comparison of a Randomized and Two Quasi-Experimental Designs in a Single Outcome Evaluation: Efficacy of a University-Level Remedial Writing Program." *Evaluation Review*, vol. 22, no. 4, 1998, pp. 207-244.

Anderson, Amy B. and Dale DeCesare. "Opening Closed Doors: Lessons from Colorado's First Independent Charter School." Denver, CO: Augenblick, Palaich and Associates, Inc., May 2006.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. "Who Benefits from KIPP?" *National Bureau of Economic Research Working Paper #15740*, Cambridge, MA: February 2010.

Betts, Julian R. and Y. Emily Tang. "Value-Added and Experimental Studies of the Effect of Charter Schools on Student Achievement." Seattle, WA: Center on Reinventing Public Education, December 2008.

Bifulco, Robert, Casey D. Cobb, and Courtney Bell. "Can Interdistrict Choice Boost Student Achievement? The Case of Connecticut's Interdistrict Magnet School Program." *Educational Evaluation and Policy Analysis*, vol. 31, no. 4, December 2009, pp. 323-345.

Bifulco, Robert and Helen F. Ladd. "The Impact of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance and Policy,* vol. 1, no. 1, 2006, pp. 50-90.

Bloom, Howard, Carolyn Hill, Alison Rebeck Black, and Mark Lipsey. "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions," *MDRC Working Papers on Research Methodology*, October 2008.

Booker, Toby Kevin, Scott Gilpatric, Timothy Gronberg, and Dennis Jansen. "The Impact of Charter School Attendance on Student Performance." *Journal of Public Economics*, vol. 91, nos. 5-6, June 2007, pp. 849-876.

Brooks, David. "The Harlem Miracle." *The New York Times*, May 8, 2009, p.A31.

Center for Research on Education Outcomes (CREDO). "Multiple Choice: Charter School Performance in 16 States." Stanford, CA: Stanford University, June 2009.

Charter School Achievement Consensus Panel. "Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines." NCSRP White Paper Series, No. 2. Seattle, WA: Center on Reinventing Public Education, University of Washington, May 2006.

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.

Dobbie, Will, and Roland G. Fryer, Jr. "Are High-Quality Schools Enough to Close the Achievement Gap? Evidence from a Bold Social Experiment in Harlem," National Bureau of Economic Research Working Paper #15473. Cambridge, MA: NBER, November 2009.

Doran, Harold C. and Darrel W. Drury. *Evaluating Success: KIPP Educational Program Evaluation*, New American Schools Education Performance Network, October 2002.

Educational Policy Institute. "Focus on Results: An Academic Impact Analysis of the Knowledge is Power Program." San Francisco, CA: KIPP Foundation. August 2005.

Gill, Brian, P., Mike Timpane, Karen E. Ross, Dominic J. Brewer, and Kevin Booker. "Rhetoric Versus Reality: What We Know and What We Need to Know about Vouchers and Charter Schools." Santa Monica, CA: RAND Corporation, 2007.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin, and Gregory F. Branch. "Charter School Quality and Parental Decision Making with School Choice." *Journal of Public Economics*, vol. 91, 2007, pp. 823-848.

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.

Hoxby, Caroline M. and Jonah E. Rockoff. "Findings from the City of Big Shoulders." *Education Next*, vol. 5, no. 4, 2005, pp. 52-58.

Imberman, Scott. "Achievement and Behavior in Charter Schools: Drawing a More Complete Picture." *The Review of Economics and Statistics*, forthcoming.

Mac Iver, Martha A. and Elizabeth Farley-Ripple. "The Baltimore KIPP Ujima Village Academy, 2002-2006: A Longitudinal Analysis of Student Outcomes". Center for the Social Organization of Schools, June 2007.

Mathews, Jay. *Work Hard. Be Nice: How Two Inspired teachers Created the Most Promising Schools in America*. Chapel Hill, NC: Algonquin Books, 2009.

McDonald, Aaron J., Steven M. Ross, Jane Abney, and Todd Zoblotsky. "Urban School Reform: Year 4 Outcomes for the Knowledge is Power Program in an Urban Middle School." Paper presented at the Annual Meeting of the American Educational Research Association, March 2008.

Peikes, D., L. Moreno, and S. Orzol. "Propensity Score Matching: A Note of Caution for Evaluators of Social Programs." *The American Statistician*, vol. 62, no. 3, 2008, pp. 222-231.

Sass, Tim R. "Charter Schools and Student Achievement in Florida." *Education Finance and Policy,* vol. 1, no. 1, winter 2006, pp. 91-122.

Smith, Jeffrey and Petra Todd. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, vol. 125, nos. 1-2, 2005, pp. 305-353.

Solmon, Lewis, Kern Paark, and David Garcia. "Does Charter School Attendance Improve Test Scores? The Arizona Results." Phoenix, AZ: Goldwater Institute, 2001.

Trochim, W.M.K. and J.C. Cappelleri. "Cutoff Assignment Strategies for Enhancing Randomized Clinical Trials." *Controlled Clinical Trials*, vol. 13, 1992, 190-212.

US Department of Education, "Research on the Academic Effects of Small Class Size," April 1998. Available at http://www2.ed.gov/pubs/ClassSize/academic.html

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

Witte, John F., David L. Weimer, Arnold Shober, and Paul Schlomer. "The Performance of Charter Schools in Wisconsin." *Journal of Policy Analysis and Management*, vol. 26, no. 3, 2007, pp. 557-573.

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.

Zimmer, Ron, Richard Buddin, Derrick Chau, Glenn Daley, Brian Gill, Cassandra Guarino, Laura Hamilton, Cathy Krop, Dan McCaffrey, Melinda Sandler, and Dominic Brewer. "Charter School Operations and Performance: Evidence from California." Santa Monica, CA: RAND Corporation, 2003.

Zimmer, Ron, Brian Gill, and Kevin Booker. "Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition." Santa Monica, CA: RAND Corporation, March 2009.

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

APPENDIX A

ADMINISTRATIVE DATA

In Appendix A, we describe the data used in this report in greater detail. First, we describe the process by which we collected the administrative data from states and districts ("jurisdictions") represented by the 22 KIPP schools in our sample. Then, we describe the specific years of data and variables collected by school.

Administrative Data Collection Process

We selected 22 KIPP middle schools for inclusion in this report based on two criteria. First, all included schools had to be established in 2005 or earlier, to ensure that a minimum of two cohorts of students per school would be observed for multiple years. Second, included KIPP schools had to be located in jurisdictions providing at least three consecutive years of complete, longitudinally-linked student-level data for traditional public and charter schools.

Obtaining student-level longitudinal data was necessary to track individual KIPP and non-KIPP students in the baseline years prior to middle school enrollment. All of the records obtained from jurisdictions were de-identified; each student received a unique identifier code to permit longitudinal analyses. We requested variables from jurisdictions' administrative data systems including test scores in reading and mathematics, demographic characteristics, schools attended and dates of enrollment, and discipline and attendance, where available. Within each jurisdiction, Mathematica requested data for all school years beginning with the year KIPP first opened a middle school and up to two years prior to capture baseline data for the maximum number of cohorts.

We obtained records for 15 KIPP schools from the school districts hosting each KIPP middle school. Records for the remaining 7 KIPP schools were obtained from statewide data sets in Arkansas, Massachusetts, and Texas. Mathematica restricted statewide data files on a geographic basis to make statewide analyses comparable to the analyses carried out with district-specific data sets. Each statewide data file was restricted to (1) the district where the KIPP school is geographically located, and (2) any contiguous districts that included elementary schools attended by KIPP students (feeder schools). In the case of Texas, which included KIPP schools located in four different metropolitan areas, the statewide data was split into four different analytical files. In Table A.1 we summarize the list of included comparison districts for each sample KIPP school located in Arkansas, Massachusetts, and Texas.

Table A.1: Included Districts for KIPP Schools in Arkansas, Massachusetts, and Texas

School	City	Traditional Public School Districts	Charter School Districts
KIPP Delta	Helena, AR	Barton-Lexa Elaine Helena-West Helena Lee County Marvel	(none)
KIPP Lynn	Lynn, MA	Beverly Lynn Nahant Peabody Revere Salem Swampscott	Marblehead Community Charter Northshore Education Consortium Salem Academy Charter
KIPP Austin College Prep	Austin, TX	Austin Del Valle Leander Manor Pflugerville Round Rock	Harmony Science Academy KIPP Austin College Preparatory School Inc. KIPP Inc. Charter Texas Academy of Excellence
KIPP Truth	Dallas, TX	Cedar Hill Dallas DeSoto Duncanville Garland Lancaster Mesquite	Faith Family Academy of Oak Cliff Harmony Science Academy KIPP Inc. Charter KIPP Truth Academy Life School
KIPP Academy Houston KIPP 3D Academy	Houston, TX	Aldine Alief Channelview Cypress-Fairbanks Fort Bend Galena Park Houston Humble Katy Klein Lamar Consolidated North Forest Pasasdena Pearland Spring Spring Branch Stafford Municipal	Beatrice Mayes Institute Charter School Harmony Science Academy KIPP Inc. Charter SER-Ninos Charter School Varnett Charter School YES College Preparatory School
KIPP Aspire	San Antonio, TX	Edgewood Harlandale North East Northside San Antonio South San Antonio	Brooks Academy of Science and Engineering KIPP Aspire Academy KIPP Inc. Charter La Escuela de las Americas School of Excellence in Education

Note: KIPP Academy Houston and KIPP 3D Academy are both located in Houston and therefore share a comparison set of districts.

District Data Coverage and Data Processing

Once we obtained the administrative data, we implemented a data clarification protocol for each district. This process included confirming data variable definitions with district administrators, assessing data coverage gaps, and merging separate district data sets into a single longitudinally linked analytical data set. The analysis file was structured in "long" form, wherein a given student had a separate record for each year he or she appeared in the data.

The first step in this procedure was standardizing the outcome test scores for each district. During the standardization process, each test score was converted into a z-score—a measure of the relative difference between each student's raw test score and the average score in that student's school district for the year and grade in which the test was administered. To generate this measure, the mean score for all district students taking the test in the same year and grade is subtracted from each raw score, and then divided by the standard deviation of observed scores for that test grade and year.

In addition to standardizing test scores, the data processing stage also addressed gaps in data coverage. Administrative data in several districts did not provide reliable information for certain demographic variables. For example, Chicago Public Schools did not provide free or reduced price lunch eligibility status, and the Texas Education Agency did not provide Limited English Proficiency status. School enrollment information varied in quality and level of detail. Several jurisdictions listed multiple enrollment schools for students in a single school year without specifying the enrollment dates associated with each school.[22] In these cases, we developed a process to select the first and last school attended each year by inference, based on the school enrollment information from the preceding and following years for that student. This approach was not successful in all cases, however; Oklahoma City Public Schools provided multiple-enrollment data that did not appear in a format that supported analyses of within-year transfer and attrition patterns.

Almost all of the variables included in the study's analyses were defined as binary or 'dummy' indicator variables. A student was defined as being enrolled in KIPP if they enrolled in a KIPP middle school at any point during the school year. We defined "LEP" as an indicator variable for limited English proficiency (also referred to as being an English-language learner in some districts). We defined "FRPL" is an indicator variable for free or reduced price lunch eligibility. SPED was defined as a binary indicator variable representing special education status. The list of racial category indicator variables provided varied by jurisdiction.

Table A.2 summarizes the years of data obtained from each district, the outcome tests analyzed for each grade and year combination, the demographic variables used in the analyses, and the number of KIPP cohorts with at least one year of baseline (pre-KIPP) test score data and at least one year of test score data after KIPP enrollment. The table notes provide more detail for individual jurisdictions.

---

[22] These included Atlanta Public Schools, Fulton County School District, and the Massachusetts Department of Education. Additionally, the Arkansas Department of Education did provide quarterly enrollment and attendance information, but not enrollment dates; the process was modified to examine the enrollment information from the preceding and following quarter.

Table A.2. Administrative Data Specifications

Jurisdiction	Years of Data Collected	Outcome Tests Analyzed	Demographic Variables Analyzed	Missing Demographic Variables	KIPP School (Year Opened)	Number of Cohorts with Impact Estimates
Arkansas Department of Education[a]	03-04 to 07-08	Grade-level Benchmark Exams	American Indian Asian Hispanic Black White LEP FRPL SPED		Delta College Preparatory (2002)	3 (05-06 to 07-08)
San Francisco Public Schools	00-01 to 08-09	California Standards Tests	American Indian Asian Pacific Islander Filipino Hispanic Black White LEP FRPL SPED		Bayview Academy (2003) SF Bay Academy (2003)	6 (03-04 to 08-09) 6 (03-04 to 08-09)
Washington, DC Public Schools[b]	00-01 to 08-09	Stanford Achievement Tests, 9th Edition (00-01 to 04-05) District of Columbia Comprehensive Assessment System (05-06 to 08-09)	American Indian Asian Hispanic Black White LEP FRPL SPED		DC KEY Academy (2001) DC AIM Academy (2005)	8 (01-02 to 08-09) 3 (05-06 to 07-08)
Atlanta Public Schools[c]	03-04 to 07-08	Criterion-Referenced Competency Tests, based on Georgia Performance Standards (03-04 to 06-07) Criterion-Referenced Competency Tests, based on Quality Core Curriculum (05-06 to 07-08)	American Indian Asian Hispanic Black White Mixed LEP FRPL SPED		WAYS Academy (2003) Achieve Academy (2004) – closed	4 (04-05 to 07-08) 3 (04-05 to 06-07)
Fulton County School District	01-02 to 07-08	Criterion-Referenced Competency Tests	American Indian Asian Hispanic Black White Mixed LEP FRPL SPED		South Fulton Academy (2003)	5 (03-04 to 07-08)

Table A.2. *(continued)*

Jurisdiction	Years of Data Collected	Outcome Tests Analyzed	Demographic Variables Analyzed	Missing Demographic Variables	KIPP School (Year Opened)	Number of Cohorts with Impact Estimates
Chicago Public Schools[d]	96-97 to 07-08	Iowa Tests of Basic Skills (96-97 to 99-00) Illinois Standards Achievement Tests (00-01 to 07-08)	American Indian Asian Hispanic Black White Mixed LEP SPED	FRPL	Ascend Charter (2003) Youth Village Academy (2003) – closed	5 (03-04 to 07-08) 3 (03-04 to 05-06)
Massachusetts Department of Education[e]	02-03 to 07-08	Massachusetts Comprehensive Assessment System	American Indian Asian Hispanic Black White Mixed LEP FRPL SPED		Academy Lynn (2004)	4 (04-05 to 07-08)
New York City Public Schools[f]	02-03 to 07-08	New York State Testing Program	American Indian Asian Hispanic Black White LEP SPED	FRPL	Academy NY (1995) STAR College Preparatory (2003) AMP Academy (2005) Infinity Charter (2005)	5 (03-04 to 07-08) 5 (03-04 to 07-08) 3 (05-06 to 07-08) 3 (05-06 to 07-08)
Oklahoma City Public Schools[g]	00-01 to 07-08	Oklahoma Core Curriculum Tests	American Indian Asian Hispanic Black White LEP FRPL SPED		Reach College Preparatory (2002)	3 (05-06 to 07-08)
Philadelphia School District[h]	01-02 to 07-08	Pennsylvania System of School Assessment (01-02 to 07-08) Stanford Achievement Tests, 9th Edition (01-02) TerraNova (02-03 to 07-08)	American Indian Asian Hispanic Black White Mixed LEP SPED	FRPL	Philadelphia Charter (2003)	5 (03-04 to 07-08)

Table A.2. *(continued)*

Jurisdiction	Years of Data Collected	Outcome Tests Analyzed	Demographic Variables Analyzed	Missing Demographic Variables	KIPP School (Year Opened)	Number of Cohorts with Impact Estimates
Texas Education Agency[i]	02-03 to 07-08	Texas Assessment of Knowledge and Skills	American Indian Asian Hispanic Black White FRPL SPED	LEP	3D Academy (2001)	4 (04-05 to 07-08)
					Aspire Academy (2003)	4 (04-05 to 07-08)
					Austin College Preparatory (2002)	4 (04-05 to 07-08)
					Academy Middle School (1995)	4 (04-05 to 07-08)
					TRUTH Academy (2003)	4 (04-05 to 07-08)

Notes: A "cohort" is defined as the group of students who first enrolled in a KIPP middle school's minimum offered grade (usually 5th grade) at the beginning of that school year.

[a] In Arkansas, grade 3 and grade 4 test score data did not appear until the 2003-04 school year. As a result, the first KIPP cohort with valid baseline test score data first appeared in the 2004-05 school year.

[b] In Washington, DC, the District of Columbia Comprehensive Assessment System replaced the Stanford Achievement Tests, 9th Edition in the 2005-06 school year. The obtained data did not include test scores of KIPP AIM students in the 2008-09 school year.

[c] In Atlanta, the Quality Core Curriculum standards-based Criterion-Referenced Competency Tests were used for reading in grades 2 through 6 during 2003-04 and 2004-05, and for math in grade 6 during 2003-04 and 2004-05, grades 2 and 7 from 2003-04 to 2005-06, and grades 3, 4, 5, and 8 from 2003-04 to 2006-07. The Georgia Performance Standards-based version replaced the Quality Core Curriculum-based version after the respective school years listed. In Atlanta there was also no single variable recording students' special education status. For this district, the SPED variable was generated using students' 504 plan status as a proxy; all students with active 504 plans were classified as being in special education.

[d] In Chicago, the Illinois Standards Achievement Tests replaced the Iowa Tests of Basic Skills in the 2000-01 school year.

[e] In Massachusetts, the "mixed" race variable was defined to flag students whose observations included two or more distinct race identifiers in the data set.

[f] In New York City, free and reduced price lunch eligibility was not recorded consistently in the data, so this variable was omitted.

[g] In Oklahoma City, baseline grade 3 and grade 4 test score data for KIPP students does not appear until the 2004-05 school year. As a result, the first KIPP cohorts with valid baseline test score data first appeared in the 2005-06 school year.

[h] In Philadelphia, results from all three tests were combined. In 2001-02, the Stanford Achievement Tests, 9th Edition (SAT-9) were used for grades 3, 4, 6, and 7, and the Pennsylvania System of School Assessment (PSSA) was used for all other grades. From 2002-03 to 2004-05, the TerraNova replaced the SAT-9, and was used for grades 3, 4, 6, and 7; the PSSA also was used for the remaining grades during that span. After the 2004-05 school year, the PSSA was the only test used. Free and reduced price lunch eligibility was not recorded consistently in the Philadelphia data, so this variable was omitted.

[i] In Texas, we were not provided a special education variable in the original data for potential disclosure reasons. As a proxy, students were flagged as being 'special education' if they received disability accommodations when taking the Texas Assessment of Knowledge and Skills. Baseline test score data did not appear until the 2003-04 school year in Texas. As a result, the first KIPP cohorts with valid baseline test score data first appeared in the 2004-05 school year.

APPENDIX B

SUPPLEMENTAL TABLES FOR CHAPTER II

In Appendix B, we provide supplementary tables for the descriptive analyses of KIPP middle school characteristics presented in Chapter II. These supplementary tables present the underlying characteristics of each school in terms of baseline demographics, baseline test scores, and attrition. The appendix also includes supplemental analyses of the baseline test scores of students who change middle schools prior to completing eighth grade.

1.   Ethnicity and Gender Characteristics

As discussed in Chapter II, KIPP middle schools enroll a higher proportion of racial minorities than other middle schools in the same districts. Table B.1 reports the percentage of black and Hispanic student enrollment at KIPP middle schools, and compares those figures to the districtwide middle school population and also the pool of students who attended the same elementary schools as KIPP students. The table reports sample means in baseline years for students with at least one observation during middle school. All baseline student characteristics are taken from fourth grade observations. Table B.1 also includes the analogous figures for the percentage of female students at KIPP schools, district schools, and among students who attended KIPP's feeder elementary schools. Unlike the distribution of racial minorities, gender balance at KIPP schools is almost always similar to the ratios at district public schools.

2.   Other Demographic Characteristics

Where data was available, the study also compared KIPP enrollment levels of students who were eligible for free or reduced price lunch, enrolled in special education programs, or of limited English proficiency. The results of the comparison differed for each of these variables. Compared to district public schools and feeder elementary schools, KIPP schools in the study were more likely to enroll students eligible for free and reduced price lunch, but less likely to enroll students enrolled in special education or students of limited English proficiency. Table B.2 compares the enrollment percentages of KIPP schools across each of these characteristics to the percentages found at district middle schools and the percentages among the pool of students who attended KIPP's feeder elementary schools. The table reports sample means in baseline years for students with at least one observation during middle school.

3.   Baseline Test Scores

As discussed in Chapter II, there is no clear pattern in the baseline test scores of KIPP middle school students; the baseline scores of KIPP students tend to be lower than the district average, but there is no systematic pattern comparing the baseline scores of KIPP students to students who attended the same feeder elementary schools. Table B.3 reports the mean baseline test scores at each KIPP school in the study, and compares those scores to the average at district schools and among feeder elementary school students. The results are reported for both mathematics and reading. The table reports sample means in baseline years for students with at least one observation during middle

school. Mean mathematics scores and reading scores represent raw test scores that have been standardized by grade, subject, and year.[23]

4. Middle School Attrition Life Table

To compare transfer rates at KIPP middle schools to other public schools, we defined attrition as follows: school transfers (either in-district or out-of-district) occurring during or immediately after each grade offered. Each middle school was analyzed using its specific grade range, allowing the attrition analysis to disregard school transfers caused by a normal grade progression. For example, the fifth grade attrition rate represents the proportion of students leaving their school unexpectedly during or immediately after fifth grade. To measure the cumulative attrition rate during all of middle school, we undertook a "life table" analysis. Under this approach, we first calculated a rate of attrition at each grade level for students beginning in fifth grade. Then, we used each grade level's observed attrition rate to derive the cumulative probability that a student in fifth grade will leave that school before completing eighth grade.[24] Table B.4 presents the results of these attrition calculations, comparing the transfer rates at KIPP schools to the transfer rates at district middle schools. As discussed in Chapter II, we did not find a systematic pattern of higher or lower attrition rates at KIPP schools as compared to non-KIPP schools.

5. Cumulative Middle School Attrition, by Attrition Type

Almost all of the KIPP and district cumulative attrition rates were approximately evenly divided between within-district transfers and out-of-district transfers. Within-district transfers were defined as attrition transfers to another public school in the same district. All other transfers were classified as out of district transfers; these are comprised of transfers to other districts, transfers to private schools, or students dropping out of the school system. Table B.5 presents the cumulative attrition rates for all of these attrition types at KIPP schools and at district middle schools.

6. Attrition Status and Baseline Test Scores

To further investigate whether attrition represents a form of student selection, we compared the baseline test scores of students who transfer to students who stay at the same middle school through eighth grade. As discussed in Chapter II, students who transfer within-district tend to have lower baseline test scores than students who do not transfer at all. For KIPP, the baseline scores of students transferring in-district were significantly lower at 12 schools (in at least one subject); none of the KIPP schools recorded higher baseline scores for students transferring in-district. The pattern at non-KIPP schools was even more pronounced: compared to those who do not transfer, students transferring in-district had baseline scores that were significantly lower in at least one subject in *all* 22 sites. For out-of-district transfers, the pattern is mixed. Full comparisons of baseline scores by attrition status are shown in Appendix Table B.6 (mathematics) and Appendix Table B.7 (reading).

---

[23] In cases where students repeated their baseline grade the most recent test score was kept; dropping older test scores of grade repeaters caused the districtwide average standardized test score to be slightly above zero in some cases.

[24] Because many districts did not provide data on mid-year school transfers, these attrition calculations do not include student transfers that take place during the 8th grade school year. Also, one KIPP school enrolled students beginning in grade 4. The cumulative attrition comparison for this school and its comparison district schools includes grade 4 attrition hazard rates; to preserve school anonymity, the grade 4 rates in this case are not presented in the table.

Table B.1 Ethnicity and Gender Characteristics

	Hispanic			Black			Female	
KIPP (1)	Feeder (2)	District (3)	KIPP (4)	Feeder (5)	District (6)	KIPP (7)	Feeder (8)	District (9)
.00 N=156	.01** N=1,052	.01** N=1,058	.97 N=156	.83** N=1,052	.84** N=1,058	.62 N=156	.51* N=1,052	.51* N=1,058
.00 N=146	.03** N=8,899	.04** N=14,509	.99 N=146	.94** N=8,899	.84** N=14,509	.56 N=146	.49 N=8,899	.49 N=14,509
.00 N=329	.03** N=11,431	.04** N=14,509	1.00 N=329	.94** N=11,431	.84** N=14,509	.53 N=329	.49 N=11,431	.49 N=14,509
.00 N=630	.05** N=28,581	.09** N=36,999	.99 N=630	.92** N=28,581	.85** N=36,999	.54 N=630	.51 N=28,653	.50 N=37,080
.00 N=413	.24** N=11,594	.34** N=17,256	.97 N=413	.42** N=11,594	.31** N=17,256	.53 N=413	.48 N=11,594	.49 N=17,256
.01 N=144	.05** N=16,231	.35** N=239,477	.99 N=144	.88** N=16,231	.54** N=239,477	.51 N=144	.51 N=16,231	.50 N=239,477
.01 N=284	.03** N=14,756	.09** N=29,307	.99 N=284	.95** N=14,756	.86** N=29,307	.51 N=284	.51 N=14,756	.50 N=29,308
.01 N=423	.09** N=10,082	.10** N=25,031	.96 N=423	.69** N=10,082	.36** N=25,031	.51 N=423	.49 N=10,082	.49 N=25,031
.01 N=403	.21** N=43,333	.36** N=212,800	.99 N=403	.74** N=43,333	.53** N=212,800	.50 N=403	.50 N=43,333	.50 N=212,800
.04 N=159	.13** N=28,783	.40** N=342,706	.94 N=159	.80** N=28,783	.33** N=342,706	.50 N=159	.50 N=28,783	.49 N=342,706
.10 N=387	.29** N=17,949	.23** N=26,352	.70 N=387	.16** N=17,949	.13** N=26,352	.53 N=387	.49 N=17,949	.49 N=26,352
.21 N=376	.19 N=49,574	.16* N=88,072	.79 N=376	.72** N=49,574	.65** N=88,072	.51 N=376	.49 N=49,574	.49 N=88,072
.35 N=223	.55** N=54,190	.54** N=83,826	.65 N=223	.36** N=54,190	.30** N=83,826	.53 N=223	.50 N=54,190	.50 N=83,826
.35 N=346	.54** N=47,478	.40 N=357,095	.64 N=346	.39** N=47,478	.33** N=357,095	.51 N=346	.49 N=47,478	.49 N=357,095
.40 N=409	.25** N=23,601	.23** N=26,352	.43 N=409	.13** N=23,601	.13** N=26,352	.50 N=409	.48 N=23,601	.49 N=26,352
.54 N=274	.59 N=46,779	.40** N=357,106	.43 N=274	.33** N=46,779	.33** N=357,106	.55 N=274	.49* N=46,779	.49* N=357,106
.54 N=338	.30** N=5,642	.24** N=10,997	.23 N=338	.13** N=5,642	.09** N=10,997	.49 N=338	.49 N=5,642	.49 N=10,997
.75 N=202	.68* N=24,039	.40** N=342,706	.22 N=202	.27 N=24,039	.33** N=342,706	.52 N=202	.49 N=24,039	.49 N=342,706
.87 N=371	.50** N=155,479	.48** N=229,458	.13 N=371	.23** N=155,479	.22** N=229,458	.57 N=371	.50** N=155,479	.50** N=229,458
.90 N=384	.57** N=127,745	.48** N=229,458	.10 N=384	.26** N=127,745	.22** N=229,458	.52 N=384	.50 N=127,745	.50 N=229,458
.94 N=381	.45** N=43,116	.42** N=48,296	.06 N=381	.11** N=43,116	.10** N=48,296	.54 N=381	.50 N=43,116	.50 N=48,296
.98 N=365	.87** N=28,650	.73** N=67,475	.01 N=365	.04** N=28,650	.05** N=67,475	.49 N=365	.50 N=28,650	.50 N=67,475

* statistically significant at the five percent level
** statistically significant at the one percent level

Table B.2 Other Demographic Characteristics

FRPL			SPED			LEP		
KIPP (1)	Feeder (2)	District (3)	KIPP (4)	Feeder (5)	District (6)	KIPP (7)	Feeder (8)	District (9)
.60 N=630	.65* N=28,653	.62 N=37,080	.07 N=630	.15** N=28,581	.16** N=37,000	.00 N=630	.03** N=28,581	.05** N=37,000
.71 N=423	.65** N=10,082	.36** N=25,031	.22 N=423	.24 N=10,082	.20 N=25,031	.00 N=423	.04** N=10,082	.05** N=25,031
.71 N=284	.71 N=14,756	.65* N=29,308	.14 N=284	.16 N=14,756	.17 N=29,308	.01 N=284	.02** N=14,756	.06** N=29,308
.72 N=409	.64** N=23,589	.61** N=26,339	.11 N=409	.13 N=23,601	.13 N=26,352	.32 N=409	.35 N=23,601	.34 N=26,352
.74 N=387	.65** N=17,938	.61** N=26,339	.14 N=387	.13 N=17,949	.13 N=26,352	.10 N=387	.34** N=17,949	.34** N=26,352
.74 N=329	.84** N=11,431	.76 N=14,509	.00 N=329	.00 N=11,431	.01 N=14,509	.00 N=329	.02** N=11,431	.03** N=14,509
.82 N=338	.65** N=5,642	.52** N=10,976	.19 N=338	.17 N=5,642	.18 N=10,997	.16 N=338	.16 N=5,642	.11* N=10,997
.86 N=413	.82* N=11,594	.85 N=17,256	.12 N=413	.20** N=11,594	.19** N=17,256	.01 N=413	.18** N=11,594	.26** N=17,256
.87 N=223	.78** N=54,190	.73** N=83,826	.01 N=223	.02 N=54,190	.02 N=83,826	N/A	N/A	N/A
.88 N=146	.85 N=8,899	.76** N=14,509	.00 N=146	.00** N=8,899	.01** N=14,509	.00 N=146	.02** N=8,899	.03** N=14,509
.91 N=371	.63** N=155,479	.59** N=229,458	.01 N=371	.02** N=155,479	.02** N=229,458	N/A	N/A	N/A
.92 N=381	.46** N=43,116	.43** N=48,296	.01 N=381	.04** N=43,116	.04** N=48,296	N/A	N/A	N/A
.92 N=365	.82** N=28,650	.65** N=67,475	.04 N=365	.05 N=28,650	.04 N=67,475	N/A	N/A	N/A
.94 N=384	.75** N=127,745	.59** N=229,458	.01 N=384	.02 N=127,745	.02 N=229,458	N/A	N/A	N/A
.96 N=156	.93 N=1,052	.93 N=1,058	.06 N=156	.11* N=1,052	.11* N=1,058	.00 N=156	.00 N=1,052	.00 N=1,058
N/A	N/A	N/A	.07 N=274	.17** N=46,779	.16** N=357,106	.08 N=274	.16** N=46,779	.12* N=357,106
N/A	N/A	N/A	.09 N=159	.14* N=28,783	.16** N=342,706	.00 N=159	.04** N=28,783	.12** N=342,706
N/A	N/A	N/A	.09 N=202	.15** N=24,039	.16** N=342,706	.20 N=202	.23 N=24,039	.12** N=342,706
N/A	N/A	N/A	.11 N=376	.14 N=49,574	.14* N=88,072	.04 N=376	.07** N=49,574	.07** N=88,072
N/A	N/A	N/A	.11 N=346	.16** N=47,478	.16** N=357,095	.05 N=346	.15** N=47,478	.12** N=357,095
N/A	N/A	N/A	.14 N=403	.19** N=43,333	.19** N=212,800	.00 N=403	.08** N=43,333	.14** N=212,800
N/A	N/A	N/A	.30 N=144	.17** N=16,231	.19** N=239,477	.00 N=144	.02** N=16,220	.12** N=238,712

* statistically significant at the five percent level
** statistically significant at the one percent level

Table B.3 Baseline Test Scores

Reading			Math		
KIPP (1)	Feeder (2)	District (3)	KIPP (4)	Feeder (5)	District (6)
-.63 N=379	-.08** N=17,244	.04** N=25,453	-.71 N=378	-.07** N=17,367	.05** N=25,600
-.45 N=401	.02** N=22,765	.04** N=25,453	-.52 N=401	.02** N=22,901	.05** N=25,600
-.42 N=423	-.32* N=10,072	.05** N=24,965	-.39 N=423	-.29* N=10,076	.05** N=25,021
-.30 N=334	-.13** N=5,461	.04** N=10,780	-.27 N=334	-.11** N=5,467	.03** N=10,788
-.25 N=138	.18** N=16,166	.11** N=238,106	-.38 N=139	.09** N=16,117	.11** N=237,616
-.23 N=142	-.11 N=8,588	.00** N=13,999	-.23 N=143	-.12 N=8,600	.01** N=14,031
-.22 N=273	.01** N=35,834	.04** N=40,559	-.24 N=272	.01** N=34,655	.04** N=39,263
-.18 N=183	-.24 N=20,742	.02** N=313,381	-.11 N=201	-.27** N=23,834	.02* N=337,834
-.11 N=322	-.11 N=25,043	.03* N=59,627	.03 N=312	-.10** N=24,281	.03 N=58,030
-.09 N=335	-.17 N=42,791	.03** N=326,385	-.06 N=343	-.16* N=46,963	.03* N=351,812
-.08 N=403	-.13 N=43,240	.02* N=212,398	-.09 N=403	-.16 N=43,164	.02** N=212,058
-.07 N=208	.05* N=5,119	.01 N=7,669	-.17 N=205	.03** N=5,136	.02** N=7,718
-.02 N=213	-.03 N=50,019	.04 N=75,678	-.02 N=209	-.04 N=49,104	.04 N=74,242
-.01 N=185	-.11 N=10,314	-.01 N=20,880	-.05 N=185	-.12 N=10,374	-.01 N=21,030
.01 N=361	-.08 N=41,131	.01 N=72,273	-.03 N=362	-.10 N=41,672	.01 N=72,977
.06 N=284	-.06* N=100,505	.04 N=190,364	.07 N=276	-.07** N=98,483	.04 N=186,588
.15 N=156	.02 N=1,052	.02 N=1,058	.07 N=156	.03 N=1,052	.03 N=1,058
.15 N=265	.03** N=126,833	.04* N=190,364	.14 N=265	.03* N=124,572	.04 N=186,588
.15 N=158	.00* N=27,713	.02 N=313,381	-.02 N=158	-.09 N=28,438	.02 N=337,834
.20 N=323	-.10** N=11,028	.00** N=13,999	.20 N=322	-.10** N=11,046	.01** N=14,031
.23 N=368	-.04** N=17,493	.01** N=22,810	.32 N=373	-.03** N=17,644	.01** N=22,986
.32 N=264	-.22** N=42,176	.03** N=326,389	.26 N=273	-.22** N=46,300	.03** N=351,816

* statistically significant at the five percent level
** statistically significant at the one percent level

Table B.4. Middle School Attrition Life Table

Grade 5		Grade 6		Grade 7		Cumulative	
KIPP (1)	District (2)	KIPP (3)	District (4)	KIPP (5)	District (6)	KIPP (7)	District (8)
.06 N=209	.14** N=216,370	.02 N=122	.08** N=132,476	.00 N=58	.03** N=63,029	.08	.34
.08 N=178	.14** N=216,383	.02 N=111	.08** N=132,488	.00 N=50	.03** N=63,031	.10	.34
.05 N=378	.14** N=453,383	.05 N=286	.10** N=347,632	.02 N=204	.08** N=264,655	.11	.31
.08 N=314	.14** N=9,617	.05 N=217	.10** N=6,610	.00 N=114	.04** N=4,148	.13	.26
.08 N=368	.15** N=372,524	.04 N=273	.11** N=273,926	.07 N=189	.08 N=195,584	.18	.33
.11 N=447	.13 N=320,539	.08 N=332	.12* N=228,064	.06 N=230	.13** N=176,759	.24	.34
.17 N=309	.25** N=17,904	.06 N=192	.15** N=11,502	.08 N=115	.10 N=6,954	.27	.44
.13 N=342	.16 N=77,300	.10 N=222	.14 N=56,975	.08 N=138	.14* N=41,093	.28	.40
.14 N=361	.11 N=23,852	.13 N=259	.07* N=17,394	.08 N=158	.07 N=13,398	.31	.24
.18 N=340	.13* N=79,804	.11 N=201	.13 N=51,556	.10 N=131	.12 N=37,475	.38	.34
.19 N=155	.12* N=1,150	.25 N=83	.09** N=683	N/A1	N/A	.39	.20
.20 N=337	.19 N=16,737	.17 N=258	.21 N=9,472	.10 N=144	.10 N=5,544	.42	.44
.23 N=393	.16** N=145,119	.17 N=235	.15 N=104,309	.10 N=123	.11 N=74,606	.42	.38
.18 N=428	.13** N=320,523	.16 N=269	.12 N=228,050	.19 N=162	.13 N=176,761	.44	.34
.17 N=611	.22** N=39,814	.18 N=465	.19 N=30,201	.18 N=314	.29** N=23,132	.44	.57
.22 N=340	.11** N=23,849	.21 N=222	.07** N=17,401	.10 N=124	.07 N=13,415	.45	.24
.23 N=372	.11** N=29,459	.23 N=190	.13** N=19,620	.08 N=99	.10 N=13,127	.45	.32
.33 N=241	.26* N=99,412	.15 N=93	.11 N=62,654	.07 N=42	.14 N=46,938	.49	.44
.20 N=329	.13** N=69,255	.32 N=184	.12** N=47,175	.19 N=72	.12 N=36,675	.56	.33
.24 N=440	.28 N=19,722	.32 N=237	.24* N=12,522	.23 N=130	.22 N=9,154	.62	.62
.46 N=157	.13** N=16,692	.26 N=140	.12** N=9,482	.19 N=75	.01** N=5,580	.68	.24
.33 N=39	.10** N=98,873	.36 N=11	.07 N=75,092	N/Aa	N/A	.76	.30

Note: Attrition rates for one KIPP school could not be estimated due to unreliable school enrollment data.

[a] For districts where attrition data was not available in grade seven, columns 7 and 8 compare the cumulative attrition rate at KIPP through the latest observed grade to the cumulative attrition rate at traditional public schools up until that same grade.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table B.5. Cumulative Middle School Attrition, by Attrition Type

KIPP			District		
Overall Transfers (1)	Within District Transfers (2)	Out of District Transfers (3)	Overall Transfers (4)	Within District Transfers (5)	Out of District Transfers (6)
.08	.05	.03	.34	.24	.11
.10	.06	.04	.34	.24	.11
.11	.04	.07	.31	.20	.11
.13	.10	.03	.26	.13	.13
.18	.08	.10	.33	.18	.14
.24	.11	.13	.34	.17	.17
.27	.15	.13	.44	.23	.21
.28	.18	.11	.40	.23	.17
.31	.13	.18	.24	.07	.17
.38	.20	.18	.34	.14	.20
.39	.17	.22	.20	.04	.16
.42	.25	.17	.44	.15	.29
.42	.25	.17	.38	.20	.19
.44	.21	.23	.34	.17	.17
.44	.23	.22	.57	.30	.28
.45	.20	.25	.24	.07	.17
.45	.30	.15	.32	.10	.21
.49	.23	.27	.44	.21	.23
.56	.27	.29	.33	.14	.20
.62	.37	.25	.62	.29	.33
.68	.36	.32	.24	.09	.14
.76	.59	.18	.30	.14	.15

Note: Attrition rates for one KIPP school could not be estimated due to unreliable school enrollment data. For districts where attrition data was not available in grade seven, columns compare the cumulative attrition rate at KIPP through the latest observed grade to the cumulative attrition rate at traditional public schools up until that same grade.

Table B.6. Attrition Status and Baseline Math Test Scores

Non-Transfers		Within-District Transfers		Out-of-District Transfers	
KIPP (1)	District (2)	KIPP (3)	District (4)	KIPP (5)	District (6)
-.64 N=245	.12 N=21,267	-.96** N=77	-.59** N=1,240	-.66 N=56	-.17** N=3,022
-.45 N=300	.12 N=21,267	-.79** N=55	-.59** N=1,240	-.63 N=46	-.17** N=3,022
-.30 N=270	.14 N=19,869	-.53** N=119	-.39** N=1,579	-.59* N=34	-.18** N=3,516
-.22 N=296	.05 N=8,704	-.77** N=28	-.12** N=1,254	-.47 N=10	.02 N=789
-.20 N=200	.10 N=32,111	-.22 N=48	-.16** N=3,071	-.60 N=24	-.29** N=3,944
-.18 N=59	.14 N=194,775	-.57* N=76	-.07** N=31,854	.13 N=5	.07** N=13,511
-.13 N=124	.06 N=5,648	-.22 N=71	-.18** N=1,076	-.30 N=10	.05 N=967
-.07 N=181	.06 N=289,537	-.42 N=13	-.23** N=35,151	-.44 N=7	.01** N=12,798
-.05 N=148	.00 N=14,398	-.19 N=26	-.08** N=5,163	.24 N=11	.19** N=1,404
-.04 N=282	.07 N=267,404	-.21 N=34	-.15** N=47,409	-.01 N=27	.01** N=34,554
-.04 N=277	.07 N=166,733	-.29* N=84	-.17** N=28,207	-.09 N=42	-.11** N=14,949
-.02 N=287	.06 N=52,480	-.14 N=56	-.19** N=12,774	.13 N=19	.05 N=6,777
-.01 N=141	.06 N=289,537	-.32 N=13	-.23** N=35,151	.66* N=4	.01** N=12,798
.04 N=54	.00 N=11,669	-.53** N=67	-.35** N=843	.01 N=22	.24** N=1,514
.10 N=240	.08 N=46,362	-.24* N=47	-.01** N=5,112	-.07 N=25	-.28** N=6,336
.11 N=143	.10 N=51,128	-.23 N=43	-.04** N=13,105	-.40 N=23	-.20** N=9,108
.13 N=217	.10 N=150,067	-.26** N=35	-.11** N=19,301	-.01 N=24	-.26** N=16,366
.17 N=117	.04 N=903	-.17 N=22	-.31* N=26	-.24 N=17	.03 N=126
.17 N=231	.10 N=150,067	-.05 N=16	-.11** N=19,301	-.09 N=18	-.26** N=16,366
.25 N=245	.07 N=267,405	-.35* N=11	-.15** N=47,412	.66* N=17	.01** N=34,554
.30 N=235	.01 N=9,832	-.09** N=61	-.34** N=1,365	-.04* N=26	.18** N=2,765
.41 N=246	.00 N=12,875	.11** N=75	-.07** N=5,886	.20 N=52	.20** N=3,872

Note: Tests of statistical significance in this table compared the baseline scores of KIPP's within-district transfers (3) and out of district transfers (5) to the baseline scores of non-transferring students (1). Separate significance tests compared baseline scores of district schools' within-district transfers (4) and out of district transfers (6) to the baseline scores of non-transferring students (2).

* statistically significant at the five percent level
** statistically significant at the one percent level

Table B.7. Attrition Status and Baseline Reading Test Scores

Non-Transfers		Within-District Transfers		Out-of-District Transfers	
KIPP (1)	District (2)	KIPP (3)	District (4)	KIPP (5)	District (6)
-.56 N=246	.09 N=21,139	-.92** N=77	-.51** N=1,220	-.53 N=56	-.08** N=3,023
-.37 N=300	.09 N=21,139	-.76** N=55	-.51** N=1,220	-.58 N=46	-.08** N=3,023
-.31 N=270	.13 N=19,829	-.65** N=119	-.39** N=1,577	-.53 N=34	-.18** N=3,502
-.28 N=296	.06 N=8,700	-.55 N=28	-.16** N=1,254	-.23 N=10	.06 N=785
-.17 N=200	.10 N=33,000	-.34 N=49	-.12** N=3,182	-.40 N=24	-.31** N=4,229
-.16 N=165	.05 N=269,163	-.25 N=12	-.23** N=32,421	-.45 N=6	.05 N=11,478
-.09 N=278	.06 N=248,384	-.15 N=33	-.15** N=44,038	-.05 N=24	.04** N=31,675
-.08 N=277	.06 N=166,939	-.11 N=84	-.14** N=28,303	-.06 N=42	-.08** N=14,976
-.04 N=247	.08 N=47,396	-.28 N=47	.02** N=5,261	-.40 N=28	-.28** N=6,743
-.02 N=59	.13 N=195,114	-.41* N=75	-.04** N=31,979	-.55 N=5	.11* N=13,541
-.01 N=53	.01 N=11,638	-.42* N=67	-.37** N=843	-.16 N=22	.21** N=1,513
-.01 N=148	.00 N=14,294	-.03 N=26	-.09** N=5,133	.13 N=11	.15** N=1,388
.00 N=125	.05 N=5,617	-.17 N=72	-.21** N=1,069	-.20 N=11	.10 N=957
.01 N=290	.05 N=52,046	-.03 N=55	-.16** N=12,622	.10 N=16	.04 N=6,681
.10 N=145	.10 N=51,955	-.19 N=43	-.03** N=13,268	-.38 N=25	-.20** N=9,524
.13 N=221	.09 N=152,639	-.08 N=35	-.09** N=19,700	-.30 N=28	-.26** N=17,133
.15 N=231	.09 N=152,639	.10 N=16	-.09** N=19,700	.15 N=18	-.26** N=17,133
.18 N=141	.05 N=269,163	-.22 N=13	-.23** N=32,421	.36 N=4	.05 N=11,478
.24 N=117	.04 N=903	-.07 N=22	-.25 N=26	-.22 N=17	-.02 N=126
.31 N=242	.00 N=12,782	-.01** N=75	-.08** N=5,835	.21 N=51	.19** N=3,846
.31 N=236	.01 N=9,806	-.17** N=61	-.34** N=1,365	.06 N=26	.18** N=2,759
.35 N=236	.06 N=248,380	-.11* N=11	-.15** N=44,046	.29 N=17	.04** N=31,674

Note: Tests of statistical significance in this table compared the baseline scores of KIPP's within-district transfers (3) and out of district transfers (5) to the baseline scores of non-transferring students (1). Separate significance tests compared baseline scores of district schools' within-district transfers (4) and out of district transfers (6) to the baseline scores of non-transferring students (2).

* statistically significant at the five percent level
** statistically significant at the one percent level

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

APPENDIX C

ANALYTIC METHODS

Below, we provide details on our methods of estimating KIPP middle school impacts, including (1) our imputation strategy to address missing baseline data; (2) our propensity score techniques to identify a matched comparison group for our KIPP sample, and (3) our estimation model.

1. Imputation Strategy for Missing Baseline Data

Our benchmark estimates used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores separately by treatment status. This imputation process involves first estimating the following model:

(C1a) $$Yp\_math_{it} = \alpha + X_{it}\beta + \sum_r \varphi_r Yr\_math_{it} + \sum_{q=3}^{8} \gamma_q Yq\_reading_{it} + \varepsilon_{it}$$

(C1b) $$Yp\_reading_{it} = \alpha + X_{it}\beta + \sum_r \varphi_r Yr\_reading_{it} + \sum_{q=3}^{8} \gamma_q Yq\_math_{it} + \varepsilon_{it}$$

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_{it} 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 grade 3 through 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 grade 3 through 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 by treatment and control groups.

We first estimated equations (C1a) and (C1b) 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 then used that student's demographic characteristics and other non-missing test scores (in other words, values of the right hand side variables in equations (C1a) and (C1b)) 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 (C1a) and (C1b) above. This stochastic component is a randomly selected residual from equations (C1a) and (C1b) and is included to ensure that the variance of the imputed baseline test scores is the same as that of the observed values.

2. Matching Procedures

As mentioned in Chapter III, our matched comparison approach is a combination of first identifying a matched comparison group of students who are similar to the KIPP students, followed by applying an ordinary least squares (OLS) regression model to control for remaining differences.

The matching process consisted of three steps, described in greater detail below. First, we determined the covariates to be included in the matching model, and estimated the matching model. Second, we calculated propensity scores for sample members and selected a matched comparison group based on these propensity scores. Third, we tested the balance of baseline characteristics between our KIPP sample and matched comparison group.

As a first step in our matching procedure 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. We then performed an iterative propensity score estimation procedure on a combined data set of all "cohorts" by identifying a list of baseline demographic and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. (See Table C.1 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 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 we were testing had a p-value of 0.20 or lower, we retained it in the matching model. If its p-value exceeded 0.20, we dropped it. The dependent variable of this logistic model is a binary variable indicating whether a student ever enrolls in a KIPP school. We did not distinguish between students who enrolled for part of middle school or for the entire duration. We also did not distinguish between students who enrolled in a single KIPP school and those who enrolled in multiple KIPP schools.

Table C1. List of Potential Covariates for Inclusion in Propensity Score Model

Potential Covariate
Observed and imputed (when missing) math and reading baseline test scores from 1 year prior (always included)
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from 1 year prior
Imputed math and reading baseline test scores from 1 year prior
Observed (non-imputed) math and reading baseline test scores from 2 years prior
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from 1 year prior are imputed using method described in Appendix Section C1
Dummy variables indicating whether student repeated grades 1 and 2 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 1 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 using the variables identified by the abovementioned procedure and performed nearest neighbor matching (without replacement) of comparison group students to treatment group students, separately by cohort. We then tested for balance of the KIPP and matched comparison groups by conducting a test of the significance of differences between the two groups in their baseline test scores and other matching variables. We required baseline test scores in both subjects to be balanced across treatment and control groups; at least 90 percent of the other selected model covariates, higher order and interaction terms must also be balanced across treatment and control groups. We consider a covariate to be balanced when the means of this covariate by treatment group are not significantly different at the 5 percent level.[25]

3.   Estimation procedure

Our OLS regression model considers all math and reading test score data from grades 5 through 8 and incorporates baseline (4th grade) demographic controls, X_{it}, including indicators for gender, race/ethnicity, free/reduced-price lunch status, special education status, grade retention in a baseline year, and limited English proficiency; cohort (year by entry grade); outcome test grade level; and two years of baseline mathematics and reading test scores (3rd and 4th grade for 5th grade entry cohorts; 4th and 5th grade for 6th grade entry cohorts). (See Table C.2) The basic form of the model for each school is defined in equation 1.

$$(C2) \qquad y_{it} = \alpha + X_{it}\beta + \sum_{n=1}^{4} \delta_n\, Tn_{it} + grade\ dummies + year\ dummies + \varepsilon_{it},$$

where y_{it} is the outcome test score for student i at time t; a_{jt} is the intercept term; X_{it} is a vector of characteristics (demographic controls and two years of baseline test scores) of student i ; $\Sigma\ Tn_{it}$ is a set binary variable for treatment status in up to four years[26], indicating whether student i had first enrolled at KIPP n years previously, as of time 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. In this example, $T3$ would be equal to 0 otherwise. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; and β and δn 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.

---

[25] 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 they are used as control variables in the estimating equations. As shown in Table III.1, none of the baseline test scores in either subject in the matched comparison groups of each of the 22 schools differ by more than 0.25 of a standard deviation between treatment and control groups.

[26] Due to a combination of data availability and the year when the KIPP school first opened, treatment students received no more than 3 years of KIPP treatment in 5 out of our sample of 22 KIPP middle schools.

Table C2. List of Covariates Included in OLS Model

Covariate
Observed and imputed (when missing) math and reading baseline test scores from 1 and 2 years prior
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from 1 year prior
Imputed math and reading baseline test scores from 1 and 2 years prior
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 Section C1
Dummy variables indicating whether student repeated grades 1 and 2 years prior
Demographic variables (gender, race/ethnicity, special education status, free or reduced price lunch status and limited English proficiency status, where available)

APPENDIX D

ALTERNATIVE SPECIFICATIONS

In Appendix D, we provide supplementary tables for the analysis of KIPP middle school impacts on student achievement, including the sensitivity tests employing the alternate models discussed in Chapter III.

1. Falsification Tests for "Impacts" Prior to KIPP Enrollment

As discussed in chapter III, to help assess whether our model is capturing unobserved factors related to impacts, we estimated a model that predicts each student's test score in the year immediately before KIPP entry (fourth grade) based on the following equation:

$$(D1) \qquad y4_i = \alpha + X_{it}\beta + y3math_i + y3reading_i + y3repeat_i + \delta T_i + \varepsilon_i ,$$

where *y3math*, *y3reading*, and *y4*, are respectively the third grade mathematics and reading scores and fourth 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*, 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 fourth grade school level.

A KIPP school should not affect students' achievement trajectories before they enroll, so we should not find systematic KIPP effects the year prior to enrollment. For most of the KIPP schools, we see no evidence of a KIPP "effect" prior to entry. In 2 of 22 schools in math and 2 in reading, we find that—controlling for third grade scores and baseline characteristics—students who ultimately attend KIPP have higher fourth grade scores than students who will not attend KIPP (Table D.1). Conversely, in 2 of 22 schools in reading, students who ultimately attend KIPP do worse on fourth grade tests.

2. Alternative Sample: Districtwide Comparison Group

As discussed in Chapter III, in addition to our benchmark matched comparison approach, we estimated KIPP impacts using all students with available data in the district(s) as the comparison group. Tables D.2a and D.2b show the results from this estimation, which also incorporate our benchmark approach to grade repeaters and imputed baseline test scores. The pattern of results is similar to that of our benchmark model. Using the districtwide comparison group, effects are positive in at least one additional school each year in both reading and math, and in five additional schools in reading after the first year (13 versus 8 using the matched sample). Negative impacts are unchanged except in reading in year 4, at which point two schools show statistically significant negative effects when using the districtwide model (where none do using the matched sample).

3. Attrition Adjusted Estimates: An alternative way of defining treatment status

Our benchmark approach includes in the treatment group any student who ever attended a KIPP school, regardless of whether and for how many years he or she was enrolled. We estimated an alternate version of the model where, for each year's impact estimate, our treatment group only consisted of students who were enrolled in a KIPP middle school for at least part of that particular

school year.[27] In this case, a student who enrolled in a KIPP middle school in fifth grade in the 2002-03 school year but left KIPP after completing sixth grade at the end of the 2003-04 school year was only part of the treatment group during the 2002-03 and 2003-04 school years (for the year 1 and year 2 estimates).

We obtained these attrition adjusted estimates (AAE) by first 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 benchmark impact estimates by dividing them by a Bloom adjustment factor, p, which is equal to the number of students currently enrolled at KIPP at year t divided by the number of students who are either currently enrolled in a KIPP school at year t or were previously enrolled in a KIPP middle school prior to year t:

(D2)   *Marginal AAE= Marginal BE/p*

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

(D3a)   *Cumulative AAE, year 1 = Marginal AAE, year 1*

(D3b)   *Cumulative AAE, year 2 = Cumulative AAE, year 1 + Marginal AAE, year 2*

(D3c)   *Cumulative AAE, year 3 = Cumulative AAE, year 2 + Marginal AAE, year 3*

(D3d)   *Cumulative AAE, year 4 = Cumulative AAE, year 3 + Marginal AAE, year 4*

Note that this procedure makes a very strong assumption, namely, that students who withdraw from KIPP schools experience no continuing effect of their prior enrollment at KIPP. If this is not true—as seems likely—then the adjusted estimates will typically overestimate KIPP's full effect on students who remain enrolled. These attrition-adjusted estimates for mathematics and reading are presented in Tables D.3a and D.3b respectively. By definition, the number of statistically significant estimates does not change from our benchmark approach, but the magnitude of the impacts is larger.

4.   Grade Repetition: Assigned scores at the 5th percentile

We considered several alternate approaches to addressing the relatively high rate of grade repetition in KIPP, especially compared to the prevalence of grade repetition in comparison schools. In Tables D.4a and D.4b we present results from an alternate, 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 at the fifth percentile of the analysis sample for that school in the grades they would have attended under a "normal" grade progression. As we might expect, the magnitude and number of statistically significant positive impacts tends to be smaller

---

[27] This approach is analogous to the "treatment on treated" approach used in the experimental analysis. As before, we refrain from using the "treatment on treated" language in the context of the nonexperimental analysis for the sake of accuracy, since all the students in the treatment group attended KIPP for some length of time.

than under our benchmark approach; in reading, more schools also show a negative impact. Overall, however, the substantive story does not change, as the majority of KIPP schools still show positive impacts by two years after enrollment.

5.   Alternative Sample: Students with Non-missing Baseline Test Score Data

To test the sensitivity of our results to our imputation strategy to address missing baseline data, we re-did our matching procedure and estimated our benchmark model using the subsample of students with complete baseline test score data—that is, students for whom we did not impute scores (Tables D.5a and D5.b). Again, the pattern of positive results persists, although effects become statistically insignificant in between one and four schools in most years. In year 1 in reading, for example, 8 schools show positive impacts using the benchmark sample compared to 6 using the non-imputed sample; conversely, 2 schools show negative impacts using the benchmark sample compared to none using the non-imputed sample.

Table D.1. Fourth Grade Falsification Impacts

School	Mathematics Falsification Impacts	Reading Falsification Impacts
School D[a]	-.17 (.09)	-.13 (.07)
School A	-.07 (.04)	-.10** (.03)
School Q	-.05 (.05)	.01 (.05)
School O	-.04 (.08)	-.04 (.08)
School H	-.04 (.05)	-.03 (.04)
School C	-.03 (.03)	-.03 (.03)
School B	.00 (.04)	-.01 (.03)
School K	.00 (.03)	-.03 (.03)
School J	.01 (.04)	-.01 (.04)
School L	.01 (.03)	-.04** (.01)
School E	.02 (.04)	.10 (.06)
School T	.02 (.06)	.10 (.05)
School G[a]	.02 (.10)	-.10 (.10)
School P	.02 (.08)	-.01 (.07)
School N	.03 (.03)	.02 (.04)
School I	.06 (.04)	.03 (.03)
School U	.06 (.05)	.12** (.05)
School S	.08 (.07)	.06 (.03)
School V	.08 (.05)	.11* (.05)
School M	.08** (.03)	.04 (.03)
School R	.12** (.04)	.06 (.04)
School F	.14 (.08)	-.07 (.06)

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D2.a. Effects on Test Scores in Mathematics, by Number of Years After Enrollment, Using District-wide Comparison Group

School	Year 1	Year 2	Year 3	Year 4
School P	-.06** (.02)	.16** (.03)	.19** (.05)	.17** (.06)
School S	-.01 (.02)	.25** (.03)	.32** (.05)	.24** (.06)
School B	.01 (.02)	.13** (.03)	.12* (.05)	.22** (.07)
School G[a]	.01 (.07)	-.10 (.08)	-.04 (.08)	.02 (.16)
School D[a]	.03 (.05)	.09 (.06)	.12* (.05)	.11 (.06)
School U	.03 (.04)	.22** (.05)	.10 (.07)	.03 (.13)
School M	.10* (.05)	.06 (.07)	.10 (.10)	-.17 (.12)
School L	.16** (.03)	.38** (.04)	.27** (.06)	.20* (.09)
School K	.17** (.02)	.32** (.03)	.45** (.05)	.39** (.06)
School O	.19** (.05)	.34** (.09)	.56** (.14)	N/A
School N	.21** (.03)	.14** (.04)	.25** (.05)	.24** (.06)
School I	.23** (.03)	.26** (.05)	.26** (.08)	-.07 (.16)
School A	.32** (.03)	.35** (.03)	.31** (.04)	.27** (.06)
School J	.36** (.03)	.58** (.03)	.57** (.03)	.51** (.05)
School F	.37** (.04)	.57** (.05)	.74** (.07)	.50** (.10)
School E	.40** (.05)	.37** (.07)	.55** (.13)	N/A
School V	.52** (.03)	.63** (.04)	.47** (.04)	.45** (.05)
School T	.52** (.05)	.39** (.05)	.65** (.10)	N/A
School Q	.54** (.04)	.60** (.05)	.69** (.06)	N/A
School C	.57** (.03)	.55** (.03)	.64** (.05)	.58** (.07)
School H	.77** (.03)	.80** (.04)	.85** (.05)	N/A
School R	.79** (.03)	.69** (.04)	.70** (.04)	.57** (.07)

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D2.b. Effects on Test Scores in Reading, by Number of Years After Enrollment, Using District-wide Comparison Group

School	Year 1	Year 2	Year 3	Year 4
School U	-.12** (.05)	-.11* (.04)	-.10 (.06)	-.20* (.09)
School B	-.09** (.03)	-.01 (.04)	-.02 (.05)	.03 (.08)
School M	-.01 (.04)	-.11 (.06)	-.09 (.08)	-.24* (.12)
School P	.01 (.02)	.08** (.03)	.02 (.04)	.06 (.05)
School D[a]	.03 (.05)	.05 (.06)	.20** (.06)	.08 (.06)
School S	.05* (.02)	.13** (.03)	.21** (.03)	.20** (.05)
School N	.05 (.03)	.10** (.04)	.27** (.05)	.07 (.07)
School K	.05* (.03)	.004 (.03)	.06 (.04)	.13** (.05)
School F	.06 (.04)	.21** (.05)	.37** (.06)	.42** (.09)
School T	.07 (.05)	.23** (.06)	.21* (.09)	N/A
School G[a]	.08 (.08)	-.08 (.08)	.01 (.08)	.09 (.17)
School L	.09** (.03)	.22** (.04)	.18** (.05)	.07 (.10)
School V	.11** (.04)	.36** (.05)	.37** (.06)	.28** (.07)
School Q	.12** (.04)	.24** (.05)	.31** (.07)	N/A
School A	.14** (.03)	.26** (.03)	.25** (.04)	.25** (.05)
School R	.18** (.03)	.21** (.03)	.22** (.04)	.33** (.07)
School J	.18** (.03)	.28** (.03)	.35** (.04)	.18** (.05)
School I	.20** (.03)	.22** (.05)	.19** (.06)	.19 (.15)
School O	.25** (.04)	.27** (.08)	.44** (.13)	N/A
School C	.27** (.03)	.31** (.03)	.42** (.04)	.39** (.06)
School H	.30** (.03)	.57** (.05)	.92** (.11)	N/A
School E	.45** (.04)	.30** (.06)	.47** (.11)	N/A

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D3.a. Effects on Test Scores in Mathematics, by Number of Years After Enrollment, Attrition-Adjusted

School	Year 1	Year 2	Year 3	Year 4
School P	-.13**	.11	.40**	.12
School D[a]	-.05	.09	N/A	N/A
School B	-.01	.18**	.29**	.50**
School G[a]	-.001	N/A	N/A	N/A
School M	.01	-.03	.13	-.70
School U	.01	.22**	.17	-.14
School S	.04	.30**	.33**	.24*
School L	.11**	.32**	.22**	.10
School K	.14**	.39**	.61**	.56**
School I	.15**	.33**	.34**	-.51
School N	.18**	.17**	.34**	.22*
School O	.27**	.40**	1.01**	N/A
School J	.29**	.53**	.58**	.65**
School A	.34**	.42**	.33**	.18**
School F	.35**	.59**	.79**	.53**
School E	.42**	.37**	.84**	N/A
School V	.49**	.60**	.48**	.47**
School T	.52**	.48**	.62**	N/A
School C	.58**	.60**	.68**	.68**
School Q	.58**	.67**	.66**	N/A
School H	.69**	.75**	.82**	N/A
School R	.76**	.70**	.77**	.51**

Note: This model adjusts the marginal yearly KIPP effect according to the number of early transfers from KIPP during middle school. Cumulative yearly effects are derived from adjusted marginal effects. Significance levels of each year's cumulative effect are derived directly from the study's benchmark model.

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D3.b. Effects on Test Scores in Reading, by Number of Years After Enrollment, Attrition-Adjusted

School	Year 1	Year 2	Year 3	Year 4
School P	-.15**	.01	.16	.29
School U	-.13*	-.14*	-.05	-.12
School B	-.05	.01	.03	.04
School M	-.03	-.09	-.09	-.30
School D[a]	-.03	-.003	N/A	N/A
School G[a]	.03	N/A	N/A	N/A
School L	.04	.19**	.14*	.00
School Q	.04	.27**	.29**	N/A
School N	.04	.15**	.34**	.13
School S	.05	.20**	.30**	.17*
School K	.06	-.03	.02	.21
School F	.08	.25**	.43**	.40**
School T	.10	.18*	.27	N/A
School V	.11	.35**	.34**	.26**
School J	.14**	.24**	.30**	.19**
School A	.16**	.31**	.31**	.39**
School I	.17**	.23**	.17*	.16
School R	.17**	.21**	.35**	.44**
School O	.25**	.24**	.63**	N/A
School C	.30**	.34**	.46**	.39**
School H	.30**	.44**	.96**	N/A
School E	.42**	.29**	.43**	N/A

Note: This model adjusts the marginal yearly KIPP effect according to the number of early transfers from KIPP during middle school. Cumulative yearly effects are derived from adjusted marginal effects. Significance levels of each year's cumulative effect are derived directly from the study's benchmark model.

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D4.a. Effects on Test Scores in Mathematics, by Number of Years After Enrollment, Conservative Approach to Grade Repetition

School	Year 1	Year 2	Year 3	Year 4
School P	-.15** (.05)	.02 (.07)	.22* (.09)	.12 (.10)
School M	-.11 (.06)	-.10 (.11)	.22 (.12)	-.05 (.11)
School G[a]	-.03 (.09)	.02 (.10)	-.07 (.10)	-.17 (.18)
School D[a]	-.03 (.07)	-.09 (.07)	-.01 (.07)	-.12 (.08)
School B	-.01 (.03)	.11** (.04)	.16** (.06)	.20* (.10)
School L	.004 (.04)	.23** (.05)	.16* (.07)	.12 (.10)
School U	.004 (.05)	.14* (.06)	.05 (.09)	-.26 (.16)
School S	.01 (.04)	.26** (.06)	.29** (.08)	.24* (.10)
School I	.08 (.05)	.22** (.06)	.28** (.08)	.21 (.12)
School K	.12** (.03)	.30** (.05)	.38** (.07)	.26** (.09)
School N	.17** (.04)	.11* (.05)	.20** (.07)	.004 (.11)
School O	.22** (.06)	.29** (.10)	.54** (.14)	N/A
School J	.29** (.04)	.47** (.04)	.50** (.06)	.54** (.09)
School A	.31** (.03)	.33** (.04)	.23** (.06)	.16 (.08)
School F	.33** (.05)	.50** (.06)	.58** (.10)	.38** (.14)
School E	.45** (.07)	.39** (.10)	.71** (.23)	N/A
School V	.49** (.05)	.54** (.06)	.39** (.06)	.33** (.08)
School T	.49** (.06)	.38** (.08)	.52** (.13)	N/A
School C	.55** (.04)	.50** (.05)	.56** (.06)	.50** (.09)
School Q	.55** (.06)	.63** (.08)	.57** (.10)	N/A
School H	.69** (.05)	.66** (.07)	.64** (.10)	N/A
School R	.73** (.04)	.65** (.06)	.64** (.07)	.42** (.11)

Note: After grade repetition, students were assigned the fifth percentile standardized test score observed in the grade and year where they would have appeared in the absence of grade retention. In all other respects, this model is the same as the study's benchmark model.

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D4.b. Effects on Test Scores in Reading, by Number of Years After Enrollment, Conservative Approach to Grade Repetition

School	Year 1	Year 2	Year 3	Year 4
School P	-.16** (.05)	-.09 (.06)	.03 (.08)	.17 (.13)
School U	-.16** (.06)	-.14* (.06)	-.16* (.08)	-.31* (.14)
School M	-.13** (.05)	-.15 (.10)	-.01 (.10)	-.27 (.18)
School L	-.06 (.04)	.13* (.05)	.15* (.07)	.11 (.12)
School B	-.04 (.04)	-.02 (.05)	-.03 (.06)	-.05 (.10)
School D[a]	-.02 (.08)	-.14 (.08)	-.11 (.09)	-.25* (.10)
School K	.01 (.04)	-.09 (.05)	-.10 (.07)	-.02 (.09)
School N	.01 (.04)	.04 (.05)	.15* (.08)	-.03 (.12)
School S	.04 (.04)	.17** (.05)	.22** (.05)	.15 (.08)
School G[a]	.04 (.10)	-.04 (.09)	-.03 (.11)	.002 (.21)
School Q	.05 (.06)	.25** (.07)	.24* (.10)	N/A
School F	.07 (.05)	.17* (.07)	.20* (.09)	.22 (.14)
School T	.08 (.06)	.15* (.08)	.29* (.13)	N/A
School V	.11 (.06)	.33** (.06)	.32** (.08)	.25** (.09)
School A	.13** (.03)	.21** (.05)	.16* (.06)	.20* (.08)
School I	.14** (.04)	.18** (.06)	.19** (.07)	.27* (.12)
School J	.14** (.04)	.20** (.04)	.26** (.06)	.16* (.07)
School R	.15** (.04)	.22** (.05)	.23** (.08)	.23* (.10)
School O	.23** (.05)	.19* (.09)	.38** (.13)	N/A
School C	.27** (.04)	.29** (.05)	.35** (.06)	.29** (.09)
School H	.30** (.05)	.38** (.06)	.82** (.13)	N/A
School E	.44** (.06)	.31** (.08)	.35* (.17)	N/A

Note: After grade repetition, students were assigned the fifth percentile standardized test score observed in the grade and year where they would have appeared in the absence of grade retention. In all other respects, this model is the same as the study's benchmark model.

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D5.a. Effects on Test Scores in Mathematics, by Number of Years After Enrollment, Students with Non-Missing Baselines Test Scores Only

School	Year 1	Year 2	Year 3	Year 4
School P	-.15* (.06)	.05 (.08)	.32** (.10)	.17 (.13)
School S	-.02 (.07)	.20* (.09)	.03 (.12)	-.08 (.18)
School D[a]	.01 (.07)	.05 (.08)	.16 (.08)	.12 (.09)
School B	.03 (.03)	.22** (.05)	.33** (.08)	.47** (.13)
School L	.04 (.05)	.29** (.09)	.32** (.12)	.32* (.15)
School U	.06 (.07)	.12 (.08)	.09 (.12)	N/A
School G[a]	.09 (.10)	.13 (.12)	.18 (.15)	N/A
School M	.09 (.08)	.16 (.16)	N/A	N/A
School K	.10** (.03)	.26** (.05)	.39** (.06)	.37** (.08)
School N	.20** (.04)	.19** (.05)	.30** (.07)	.26** (.09)
School I	.23** (.05)	.36** (.08)	.49** (.13)	N/A
School O	.25** (.08)	.51** (.18)	N/A	N/A
School J	.32** (.05)	.49** (.06)	.57** (.07)	.69** (.11)
School A	.35** (.03)	.40** (.04)	.32** (.05)	.23** (.08)
School T	.40** (.06)	.24** (.07)	.37** (.14)	N/A
School F	.41** (.08)	N/A	N/A	N/A
School E	.42** (.07)	.32** (.12)	N/A	N/A
School Q	.47** (.07)	.56** (.08)	.59** (.11)	N/A
School V	.51** (.05)	.55** (.06)	.50** (.08)	.30* (.12)
School C	.58** (.04)	.62** (.04)	.71** (.06)	.69** (.09)
School R	.61** (.05)	.63** (.06)	.72** (.08)	.61** (.14)
School H	.71** (.06)	.86** (.08)	.89** (.09)	N/A

Note: This model does not include imputed baseline test scores. In all other respects, this model is the same as the study's benchmark model.

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

Table D5.b. Effects on Test Scores in Reading, by Number of Years After Enrollment, Students with Non-Missing Baselines Test Scores Only

School	Year 1	Year 2	Year 3	Year 4
School U	-.14 (.07)	-.15* (.08)	-.17 (.11)	N/A
School P	-.10 (.06)	-.06 (.07)	.19* (.09)	.17 (.12)
School F	-.08 (.09)	N/A	N/A	N/A
School B	-.08 (.04)	.04 (.06)	.17* (.08)	.32* (.13)
School D[a]	-.06 (.08)	-.05 (.08)	.09 (.10)	.08 (.11)
School T	-.06 (.06)	.10 (.08)	.13 (.13)	N/A
School M	-.05 (.06)	-.13 (.12)	N/A	N/A
School G[a]	-.05 (.13)	.08 (.11)	.15 (.14)	N/A
School K	-.01 (.03)	-.02 (.05)	.01 (.06)	.09 (.07)
School L	.01 (.05)	.11 (.09)	.30* (.12)	-.23 (.28)
School S	.02 (.05)	.16* (.07)	.11 (.08)	-.01 (.13)
School N	.04 (.04)	.16** (.05)	.30** (.07)	.14 (.10)
School R	.04 (.05)	.14* (.06)	.13 (.07)	.38** (.13)
School J	.07 (.04)	.14** (.05)	.25** (.07)	.11 (.09)
School Q	.08 (.07)	.22* (.09)	.33** (.11)	N/A
School V	.12 (.06)	.18* (.08)	.31** (.11)	.04 (.17)
School A	.15** (.03)	.28** (.04)	.31** (.05)	.32** (.07)
School I	.21** (.04)	.34** (.06)	.31** (.07)	N/A
School H	.23** (.05)	.51** (.07)	.87** (.13)	N/A
School O	.26** (.07)	.58** (.15)	N/A	N/A
School C	.29** (.03)	.35** (.04)	.48** (.05)	.53** (.08)
School E	.34** (.06)	.07 (.11)	N/A	N/A

Note: This model does not include imputed baseline test scores. In all other respects, this model is the same as the study's benchmark model.

* statistically significant at the five percent level
** statistically significant at the one percent level
[a] KIPP affiliation removed

APPENDIX E

SUBGROUP ANALYSES

In this appendix, we present estimates we derived to identify whether KIPP had differential impacts on particular subgroups of students. In general, our strategy to identify potential subgroup differences involved adding interaction terms between the treatment indicators and each subgroup of interest to our benchmark impact models. 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. We discuss findings for each subgroup in turn below.

1. Limited English Proficiency

Due to a variety of factors related both to school populations and data availability, it was only possible to estimate the differential effect of KIPP on students with limited English proficiency (LEP) in 4 of the 22 KIPP schools in the sample (Tables E.1a and E1.b). One school showed a pattern of positive marginal effects for LEP students in mathematics above and beyond the average effect in that school. At this school, in other words, KIPP effects were significantly more positive for LEP students than for non-LEP students. No similar pattern was found in reading, where KIPP effects were statistically similar for LEP and non-LEP students. In the other three schools, there were no significant differences in effects.

2. Race/Ethnicity

As described in chapter II, virtually all of the students attending KIPP are racial and/or ethnic minorities. Over 90 percent of students are black or Hispanic in 19 of 22 schools. The majority of these schools (12 of 22) are also racially and ethnically homogenous, such that it is impossible to estimate subgroup differences based on race/ethnicity in those schools.[28] We do estimate the differential effects of KIPP on Hispanic students in 10 schools with a measurable proportion of black and/or non-Hispanic white students (Tables E.2a and E2.b) and on black students in 10 schools with a measurable proportion of Hispanic and/or non-Hispanic white students (Tables E.3a and E3.b).

We do not find systematic patterns of differential effects for Hispanic or black students across our sample of KIPP schools, though we do observe possible trends in a subset of schools that we will investigate in more detail with larger samples in future reports. We find that KIPP effects for Hispanics are significantly more positive than for non-Hispanics in at least one year in 4 schools in mathematics and 4 schools in reading, but also find effects that are significantly less positive in a single year in mathematics in 3 schools. Conversely, we find that KIPP effects for black students are significantly less positive than for non-black students in 2 schools in math and 4 schools in reading

---

[28] For similar reasons, we did not estimate subgroups differences comparing students eligible for free or reduced-price lunches (FRPL) to those ineligible, since the prevalence of FRPL eligibility was universally high in KIPP schools.

in at least one year. There were no significant differences in KIPP effects for black versus non-black students in 6 schools.

3.   Gender

As Tables E.4a and E4.b demonstrate, there is no differential pattern of KIPP effects for males versus females. Among 148 separate year by school by subject models that we estimated, we found no significant differences between males and females in 92 percent of cases. The 8 percent of cases in which we did find a significant difference is nearly as low as we would expect to occur by chance, and these estimates are not consistently positive or negative.

4.   Minority Males

Given the body of evidence that males from racial and ethnic minority groups may be particularly academically at risk, we investigated whether KIPP has a differential effect on Hispanic males or black males. Again, these estimates were feasible only where a given school's study body was not racially homogenous. These results mirror somewhat those seen for the racial and ethnic subgroups: there are no systematic patterns, but the estimated effect for Hispanic males is significantly more positive than for other students in 3 schools in mathematics and 2 schools in reading in at least one year (Tables E.5a and E5.b). The marginal effect of KIPP for black males is significantly less positive than for other students in 4 schools each in math and reading, but significantly more positive in one school (Tables E.6a and E6.b).

5.   Baseline Achievement

Finally, prior research has suggested that impacts of KIPP and other high-performing charter schools may be higher for more academically disadvantaged students: that is, those who are performing less well on standardized tests of achievement prior to entry (e.g. Angrist et al. 2010). To investigate whether this was potentially true in our sample of KIPP schools, we interacted treatment status with the baseline achievement test score in our benchmark model. Although the coefficients on these interaction terms are less readily interpretable than those for the other subgroups, the statistical significance and direction of the coefficient provides information on whether there is a possible relationship between baseline achievement and impacts. In this way, if the coefficients were generally positive and significant, we would deduce that the effects of KIPP are highest for students who were higher performers before they entered KIPP. Conversely, if the coefficients were generally negative and significant, we deduce that the effects of KIPP are highest for students who performed less well at baseline.

As with the other subgroups, we find no systematic pattern or differential effects for higher or lower performing students at baseline. In most cases, the estimated interaction term is not statistically significant. When there is a significant relationship, approximately equal numbers of schools have statistically significant positive (3) and negative (4) marginal effects in mathematics (Table E.7a). In reading, coefficients are generally negative in 5 schools (indicating that the KIPP effect is higher for lower performers) and positive in 3 schools (indicating that the KIPP effect is higher for better performers). Again, we will explore this relationship in more detail and with more schools in future reports.

Table E.1a. Marginal Effects for Limited English Proficiency Students in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.02 (.10)	-.11 (.12)	-.03 (.15)	N/A
.10 (.12)	.26* (.11)	.32 (.20)	.80** (.26)
.11 (.12)	.11 (.18)	.04 (.22)	.45 (.31)
.12 (.08)	.15 (.09)	.17 (.12)	.20 (.14)
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.1b. Marginal Effects for Limited English Proficiency Students in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.10 (.08)	.06 (.08)	-.03 (.10)	.04 (.13)
-.03 (.11)	.10 (.13)	.21 (.18)	.14 (.13)
.02 (.12)	.10 (.16)	-.20 (.28)	N/A
.08 (.14)	.07 (.17)	.16 (.21)	.26 (.29)
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.2a. Marginal Effects for Hispanic Students in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.14 (.12)	-.02 (.17)	.22 (.25)	N/A
-.05 (.09)	-.21* (.10)	-.08 (.14)	-.23 (.20)
.01 (.12)	.22 (.19)	.39 (.27)	N/A
.02 (.08)	-.08 (.10)	-.20* (.09)	-.02 (.12)
.12 (.10)	.08 (.11)	.04 (.14)	-.04 (.16)
.13 (.08)	.13 (.09)	.20* (.10)	.41** (.12)
.16 (.12)	.24 (.12)	.36* (.15)	.75** (.22)
.27** (.08)	.26** (.08)	.22* (.11)	.36* (.15)
.27* (.11)	.31* (.12)	.07 (.13)	N/A
.64 (.36)	-.41** (.15)	.04 (.17)	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.2b. Marginal Effects for Hispanic Students in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.08 (.08)	.04 (.08)	.13 (.09)	.12 (.12)
-.01 (.10)	.06 (.11)	-.04 (.13)	-.14 (.17)
.03 (.10)	.04 (.14)	.09 (.19)	N/A
.06 (.10)	.25* (.11)	.20 (.13)	.38** (.14)
.06 (.10)	.03 (.11)	-.19 (.13)	-.29 (.15)
.08 (.07)	.17* (.08)	-.01 (.10)	.17 (.13)
.10 (.13)	.05 (.10)	.08 (.13)	N/A
.13 (.11)	.23 (.14)	.19 (.15)	.06 (.11)
.16 (.10)	.33** (.12)	.28 (.20)	N/A
.77* (.35)	-.07 (.20)	.13 (.20)	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.3a. Marginal Effects for Black Students in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.64 (.36)	.41** (.15)	-.04 (.17)	N/A
-.21** (.07)	-.22** (.08)	-.18 (.10)	-.21 (.14)
-.18 (.11)	-.23 (.13)	.01 (.14)	N/A
-.14 (.08)	-.14 (.09)	-.22* (.09)	-.41** (.12)
-.12 (.10)	-.08 (.11)	-.03 (.14)	.05 (.16)
-.06 (.10)	.16 (.12)	-.06 (.17)	-.11 (.24)
-.05 (.09)	.05 (.10)	.14 (.09)	.00 (.12)
-.01 (.12)	-.22 (.19)	-.39 (.27)	N/A
.08 (.08)	-.03 (.08)	-.02 (.11)	-.29 (.16)
.14 (.12)	.02 (.17)	-.22 (.25)	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.3b. Marginal Effects for Black Students in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.77* (.35)	.07 (.20)	-.13 (.20)	N/A
-.10 (.13)	-.05 (.10)	-.08 (.13)	N/A
-.08 (.11)	-.28* (.14)	-.47** (.18)	N/A
-.06 (.10)	-.22* (.11)	-.17 (.13)	-.37** (.14)
-.06 (.10)	-.03 (.11)	.19 (.13)	.29 (.15)
-.03 (.10)	-.04 (.14)	-.09 (.19)	N/A
-.03 (.07)	-.14 (.08)	-.03 (.09)	-.13 (.12)
-.01 (.07)	-.08 (.09)	.01 (.10)	.07 (.11)
.04 (.08)	-.11 (.08)	-.18* (.09)	-.18 (.12)
.17 (.11)	.10 (.14)	.12 (.15)	.21 (.15)
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.4a. Marginal Effects for Male Students in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.23 (.16)	-.21 (.18)	-.05 (.19)	N/A
-.20* (.10)	-.19 (.11)	.03 (.13)	N/A
-.14 (.15)	-.11 (.17)	.07 (.27)	N/A
-.13 (.08)	-.06 (.12)	-.33 (.17)	N/A
-.08 (.12)	-.29 (.19)	-.04 (.23)	N/A
-.06 (.07)	-.04 (.08)	-.04 (.10)	.26 (.14)
-.04 (.09)	-.04 (.10)	-.07 (.12)	N/A
-.03 (.11)	-.11 (.12)	-.47* (.20)	N/A
.00 (.07)	.00 (.09)	-.30** (.11)	-.06 (.14)
.00 (.07)	-.01 (.08)	-.04 (.09)	.12 (.12)
.01 (.07)	.01 (.07)	-.05 (.09)	.07 (.13)
.02 (.08)	.01 (.09)	-.15 (.11)	-.48** (.14)
.03 (.10)	.04 (.12)	-.15 (.14)	N/A
.04 (.09)	.00 (.10)	.00 (.10)	.13 (.11)
.04 (.10)	-.01 (.12)	-.19 (.14)	N/A
.06 (.13)	.10 (.13)	-.12 (.13)	-.13 (.13)
.06 (.12)	-.10 (.16)	-.27 (.24)	N/A
.07 (.09)	.01 (.11)	.07 (.15)	.41 (.22)
.12 (.08)	.10 (.09)	.13 (.12)	N/A
.12 (.09)	.26* (.11)	.21 (.13)	N/A
.13* (.06)	.05 (.07)	-.08 (.10)	N/A
.15* (.06)	.09 (.08)	-.01 (.10)	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.4b. Marginal Effects for Male Students in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.20 (.20)	-.16 (.17)	-.22 (.19)	N/A
-.13 (.11)	-.12 (.11)	.02 (.11)	N/A
-.10 (.08)	-.06 (.12)	-.27 (.15)	N/A
-.10 (.13)	-.23 (.15)	-.26 (.20)	N/A
-.05 (.11)	-.05 (.14)	-.06 (.22)	N/A
-.02 (.10)	-.13 (.13)	-.38 (.19)	N/A
.01 (.10)	-.04 (.15)	-.14 (.20)	N/A
.05 (.08)	-.11 (.09)	-.26* (.11)	-.13 (.15)
.06 (.10)	-.07 (.11)	.00 (.13)	-.06 (.15)
.06 (.10)	.05 (.11)	.05 (.14)	.19 (.21)
.07 (.07)	.03 (.08)	.02 (.09)	.08 (.11)
.09 (.07)	.06 (.08)	.07 (.09)	.18 (.12)
.09 (.08)	.05 (.10)	.02 (.11)	N/A
.09 (.07)	.06 (.08)	.11 (.09)	N/A
.13 (.08)	.12 (.09)	.00 (.10)	-.11 (.14)
.14 (.11)	.01 (.11)	-.10 (.14)	N/A
.14 (.11)	.08 (.10)	.15 (.12)	N/A
.15 (.09)	.20 (.11)	.35 (.23)	N/A
.15* (.07)	.02 (.09)	-.02 (.11)	N/A
.17* (.07)	.03 (.08)	-.05 (.09)	.06 (.11)
.19 (.14)	.18 (.14)	.09 (.14)	.04 (.15)
.24** (.07)	.16* (.08)	.15 (.09)	N/A

Note:　　Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.5a. Marginal Effects for Hispanic Male Students in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.23 (.17)	-.13 (.20)	.14 (.29)	N/A
-.02 (.11)	-.14 (.13)	.02 (.19)	.05 (.30)
.00 (.10)	-.16 (.11)	.07 (.14)	N/A
.04 (.10)	-.12 (.11)	-.17 (.11)	.05 (.12)
.07 (.18)	.12 (.17)	.24 (.23)	.62* (.29)
.07 (.09)	.06 (.11)	.04 (.14)	N/A
.12 (.10)	.09 (.11)	.14 (.16)	.41 (.29)
.12 (.14)	.04 (.15)	-.13 (.19)	-.28 (.19)
.13 (.09)	.24* (.11)	.23 (.13)	N/A
.17 (.12)	.09 (.14)	.10 (.13)	.38* (.19)
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.5b. Marginal Effects for Hispanic Male Students in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.02 (.11)	.04 (.10)	.06 (.12)	.20 (.14)
.06 (.08)	.09 (.10)	.00 (.14)	.22 (.21)
.08 (.10)	-.06 (.11)	.08 (.11)	N/A
.09 (.13)	-.05 (.14)	-.36* (.16)	-.36* (.17)
.10 (.12)	.04 (.13)	.02 (.18)	.05 (.30)
.11 (.15)	.19 (.21)	.14 (.27)	.03 (.16)
.14 (.14)	.17 (.16)	.07 (.20)	N/A
.17 (.10)	.06 (.12)	.19 (.17)	.20 (.15)
.18 (.10)	.28* (.11)	.51* (.26)	N/A
.23** (.07)	.16 (.08)	.15 (.09)	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.6a. Marginal Effects for Black Male Students in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-2.75** (.32)	.13 (.21)	.05 (.35)	N/A
-.18* (.09)	-.17 (.10)	-.13 (.11)	.08 (.14)
-.14 (.15)	-.12 (.16)	-.11 (.16)	N/A
-.09 (.08)	-.07 (.08)	-.12 (.10)	-.07 (.13)
-.04 (.08)	-.01 (.09)	-.25* (.12)	.10 (.14)
.00 (.18)	.39 (.61)	-.21 (.20)	N/A
.01 (.11)	.11 (.13)	.17 (.11)	.13 (.18)
.05 (.07)	-.08 (.08)	-.08 (.09)	-.05 (.13)
.06 (.14)	.12 (.17)	-.06 (.25)	.60** (.14)
.20 (.13)	-.04 (.18)	-.53* (.25)	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.6b. Marginal Effects for Black Male Students in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-2.66**	-.98**	-.09	N/A
(.26)	(.25)	(.52)	
-.10	-.14	-.19	-.37*
(.11)	(.13)	(.14)	(.17)
-.09	-.24	-.55*	N/A
(.11)	(.14)	(.24)	
.02	-.10	-.11	.05
(.08)	(.09)	(.12)	(.15)
.08	-.05	.01	.13
(.07)	(.09)	(.09)	(.12)
.08	-.08	-.42*	N/A
(.15)	(.13)	(.20)	
.10	.03	.12	.14
(.09)	(.10)	(.11)	(.13)
.15	.09	-.03	N/A
(.18)	(.10)	(.12)	
.16*	-.04	-.16	-.09
(.08)	(.09)	(.11)	(.12)
.23	.20	.20	.21
(.17)	(.19)	(.23)	(.17)
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A
N/A	N/A	N/A	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.7a. Marginal Effects for Baseline Achievement in Mathematics, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.23** (.08)	-.08 (.08)	.10 (.10)	N/A
-.18** (.06)	-.19** (.07)	-.29** (.09)	N/A
-.13 (.07)	-.11 (.08)	-.16 (.10)	-.06 (.10)
-.12 (.07)	-.11 (.08)	-.10 (.09)	N/A
-.11 (.07)	-.11 (.08)	-.16 (.20)	N/A
-.11* (.05)	-.13** (.05)	-.08 (.05)	-.19** (.07)
-.08 (.06)	.04 (.07)	-.04 (.09)	N/A
-.07 (.06)	-.07 (.06)	-.03 (.08)	.06 (.08)
-.06 (.05)	-.15* (.06)	-.09 (.08)	-.07 (.11)
-.01 (.09)	.01 (.13)	.45** (.16)	N/A
.01 (.04)	.03 (.05)	.03 (.07)	N/A
.01 (.08)	.09 (.11)	.09 (.15)	N/A
.01 (.05)	-.06 (.05)	-.07 (.07)	-.03 (.12)
.02 (.11)	.17 (.10)	.10 (.16)	N/A
.02 (.07)	.02 (.11)	.09 (.11)	N/A
.02 (.09)	.16 (.10)	.23 (.12)	N/A
.04 (.07)	.00 (.07)	-.07 (.07)	-.02 (.08)
.05 (.04)	-.03 (.04)	.00 (.06)	N/A
.10* (.04)	.09 (.05)	.02 (.06)	-.12 (.10)
.13 (.10)	.05 (.10)	.20 (.12)	N/A
.14 (.08)	-.10 (.08)	-.02 (.09)	-.01 (.10)
.29** (.08)	.11 (.11)	.06 (.11)	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

Table E.7b. Marginal Effects for Baseline Achievement in Reading, by Number of Years after KIPP Enrollment

Year 1	Year 2	Year 3	Year 4
-.21** (.06)	-.24** (.08)	-.20* (.08)	N/A
-.18** (.06)	-.10 (.08)	-.09 (.12)	N/A
-.17* (.07)	-.15* (.07)	-.17* (.07)	N/A
-.16* (.06)	-.07 (.07)	-.11 (.09)	-.05 (.10)
-.14 (.07)	-.09 (.09)	.07 (.09)	N/A
-.12* (.05)	-.20** (.06)	-.21** (.07)	-.19 (.10)
-.09 (.07)	-.01 (.07)	.08 (.09)	N/A
-.07 (.05)	-.06 (.05)	.04 (.07)	.09 (.11)
-.06 (.08)	-.09 (.07)	-.07 (.08)	-.04 (.10)
-.04 (.05)	-.08 (.05)	-.11 (.06)	-.13 (.08)
-.03 (.04)	-.10* (.05)	-.05 (.06)	N/A
-.02 (.04)	.02 (.05)	-.03 (.06)	N/A
-.01 (.07)	-.15 (.09)	.31 (.23)	N/A
.00 (.13)	-.04 (.10)	.03 (.09)	N/A
.00 (.05)	-.08 (.06)	-.01 (.11)	N/A
.02 (.05)	.01 (.05)	.00 (.06)	-.04 (.08)
.03 (.05)	-.01 (.05)	-.01 (.06)	-.09 (.07)
.03 (.08)	-.04 (.12)	-.04 (.14)	N/A
.05 (.08)	.12 (.12)	.06 (.11)	N/A
.06 (.08)	.21** (.07)	.11 (.07)	.05 (.08)
.09 (.07)	.31** (.08)	.23* (.12)	N/A
.28** (.10)	.15 (.12)	.15 (.13)	N/A

Note: Each table row represents a different KIPP school. Table values do not represent effect sizes. Instead, each interaction effect estimate represents the difference in KIPP's average impact when comparing members of the subgroup to those who are not members of the subgroup. A positive and significant interaction effect estimate signals that KIPP's average impact is higher for the examined subgroup. Table cells with "N/A" entries represent effect years where the sample size for the subgroup of interest/comparison was too small to generate a meaningful interaction effect estimate. Table regressions are based on the study's benchmark matching model.

* statistically significant at the five percent level
** statistically significant at the one percent level

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

MATHEMATICA
Policy Research, Inc.

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