MATHEMATICA Policy Research

REPORT

FINAL REPORT

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

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

September 17, 2015

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

Submitted to:

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

Submitted by:

Mathematica Policy Research
1100 1st Street NE, 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220

Project Director: Christina Clark Tuttle
Reference Number: 06826.750

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

ACKNOWLEDGMENTS

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

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

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

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

CONTENTS

TABLES

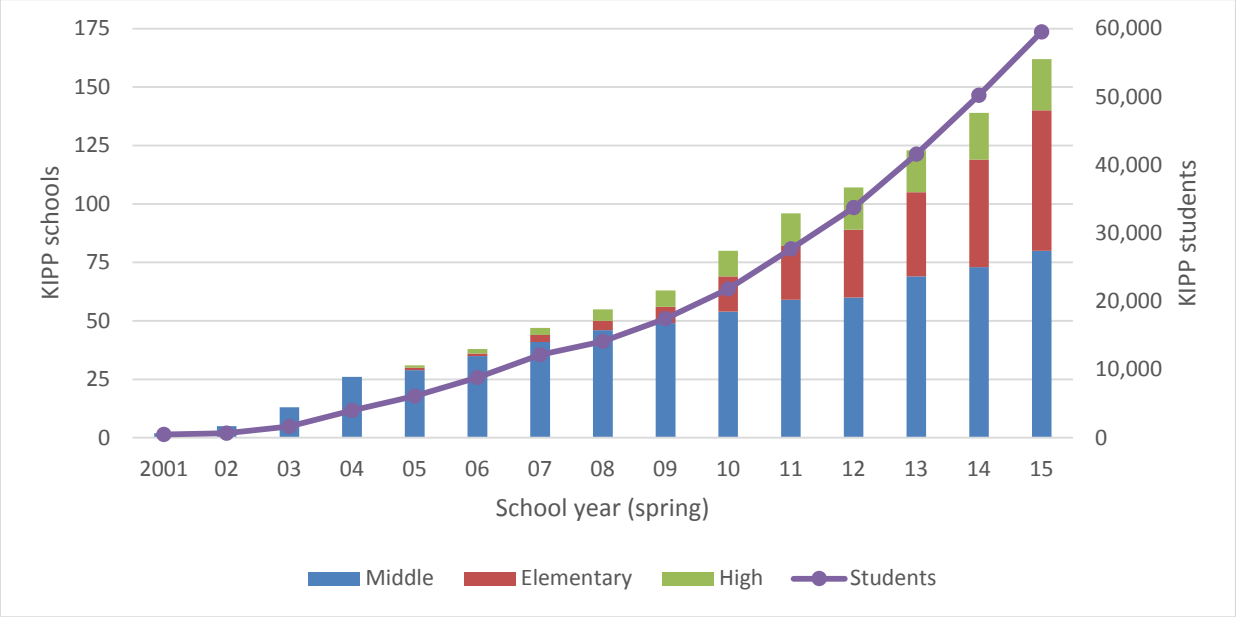
FIGURES

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

EXECUTIVE SUMMARY

KIPP is a national network of public charter schools whose stated mission is to help underserved students enroll in and graduate from college. KIPP began exclusively as a middle school program in 1994, but began expanding into the elementary and high school levels in 2004. By 2009–2010, KIPP was educating students in grades prekindergarten through 12, and as of 2014–2015 the network included 162 elementary, middle, and high schools serving 59,495 students (Figure ES.1). Prior studies (see Tuttle et al. 2013) have consistently found that attending a KIPP middle school positively affects student achievement, but few have addressed longer-term outcomes and no rigorous research exists on impacts of KIPP schools at levels other than middle school.

Figure ES.1. Number of KIPP schools and students, by year

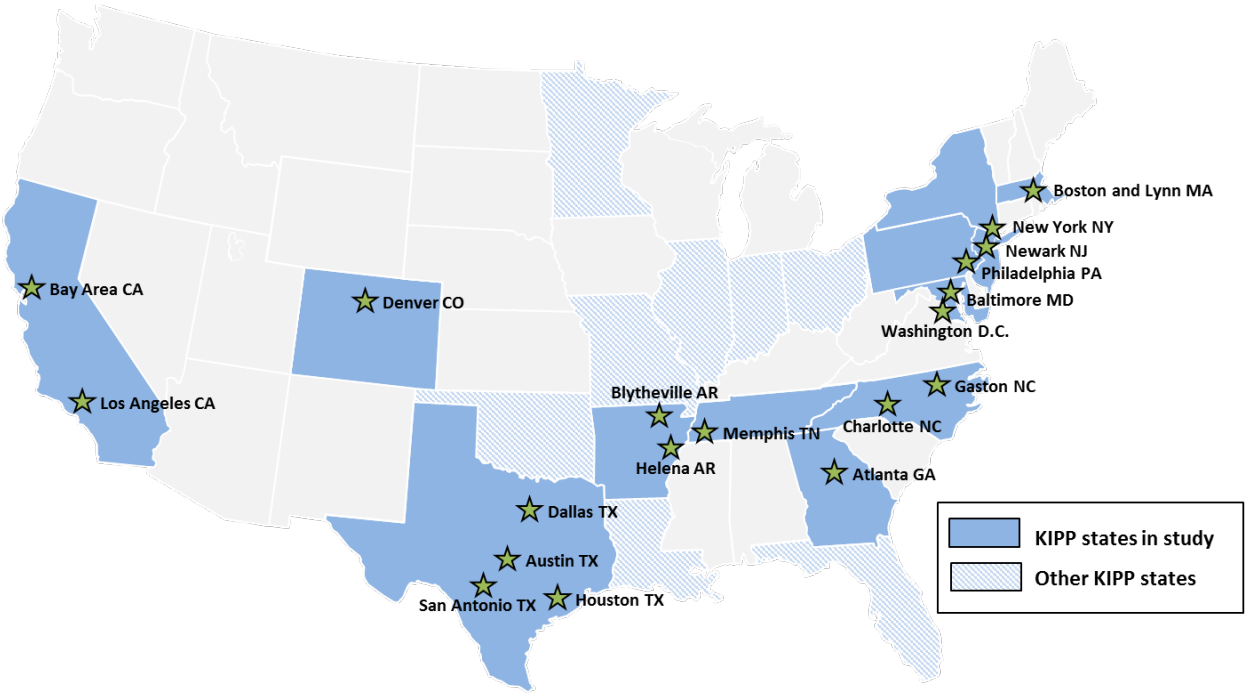


Source: KIPP Foundation.

As the KIPP network continues to grow, it faces the challenge of building a pipeline of leaders to effectively serve more students and schools. In 2010, the KIPP Foundation was awarded a five-year, $50 million Investing in Innovation (i3) scale-up grant by the U.S. Department of Education. The foundation used the i3 grant to scale up its network with the aim of sustaining KIPP's positive impacts—specifically by bolstering its leadership pipeline—while doubling the number of students served from 27,000 to over 55,000 by 2014–2015. The KIPP Foundation contracted with Mathematica Policy Research to conduct an independent evaluation of its success in improving student outcomes on a larger scale under the i3 scale-up grant. This study builds on two prior reports published by Mathematica (Tuttle et al. 2010, Tuttle et al. 2013), and is the first rigorous research to examine the impacts of KIPP schools at all three grade levels.

The key evaluation objective is to measure the impact of KIPP on student outcomes as the network scales up the number of schools, students, and grades served. To do this, we use a combination of lottery-based and quasi-experimental designs in a set of 8 elementary, 43 middle, and 18 high schools in 20 cities (Figure ES.2), employing the most rigorous study designs possible at each school level. Under different designs and samples, we measure KIPP's impacts on outcomes up to four years after students enter a KIPP school. The analysis uses data from study-administered student achievement tests; state assessments in math, English/language arts (ELA), science, and social studies; and student and parent surveys.

Figure ES.2. Location of KIPP schools in the study



Network-wide, KIPP schools have positive, statistically significant, and educationally meaningful impacts on student achievement, particularly at the elementary and middle school grades. We find that KIPP elementary schools have positive impacts on students' reading and math achievement. KIPP middle schools, meanwhile, have maintained a pattern of positive and significant impacts on reading and math over the last decade, even as the network has grown rapidly. While average impacts across the middle schools in the network declined somewhat since 2007, they stabilized during the i3 scale-up period. Moreover, the KIPP middle schools that have opened most recently—during the i3 scale-up period beginning in fall 2011—are producing positive impacts that are generally similar to those produced by older KIPP middle schools when they were in their first years of operation.

KIPP high schools have positive, statistically significant, and educationally meaningful impacts on achievement for new entrants to the network. For students continuing from KIPP middle schools, the marginal impacts of having the option to attend a KIPP high school were not statistically significant, on average (in comparison to students who did not have the option to attend a KIPP high school and instead attended a mix of other non-KIPP charter, private, and traditional public high schools). Among these continuing students, KIPP high schools have

positive impacts on several aspects of college preparation, including more discussions about college, increased likelihood of applying to college, and more advanced coursetaking.

Across grade levels, we generally find no impacts of KIPP schools on measures of students' motivation, engagement, educational aspirations, or behavior, but positive impacts on the satisfaction of parents with their child's school.

We describe these findings in more detail below.

What are the impacts of KIPP elementary schools on student achievement?

To measure impacts of KIPP elementary schools, we use a research design that uses school admissions lotteries as randomized experiments. This type of randomized design is the "gold standard" for research measuring the impacts of schools on student achievement. Students offered admission via the lottery are included in the treatment group; those not offered admission through the lottery (and enroll at other charter, private, or traditional public preschools or elementary schools) are included in the control group. This design ensures that treatment and control group students are similar at baseline (pre-KIPP) in terms of demographics and academic preparation as well as key factors such as motivation and parental support.

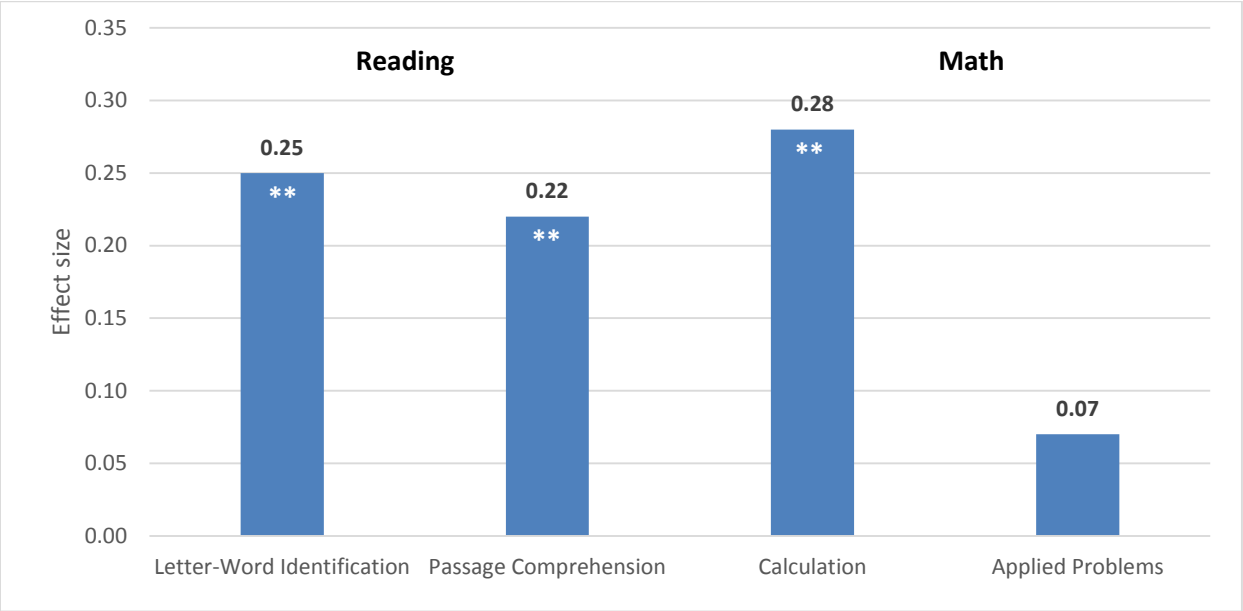
We used admissions lotteries for the 2011–2012 school year to measure the impacts of eight KIPP elementary schools. These schools make up 28 percent of KIPP elementary schools in operation in 2011–2012. To measure elementary school students' academic achievement in reading and math, we administered the Woodcock-Johnson III (WJ-III) assessment in the spring of the third follow-up year after the lottery, when most students who applied to pre-kindergarten at age 3 (PK3) were in kindergarten, and most who applied to kindergarten were in grade 2.

KIPP elementary schools have positive, statistically significant, and educationally meaningful impacts on three of four measures of students' reading and mathematics skills. On tests administered three years after entry, being offered admission to a KIPP elementary school leads to an increase of 0.25 standard deviation units on the Letter-Word Identification test and 0.22 on the Passage Comprehension test in reading (Figure ES.3). These impacts are equivalent to boosting a student's Letter-Word Identification score from the 78th percentile (the percentile corresponding to the control group students' mean score) to the 84th percentile, and boosting the Passage Comprehension score from the 48th to the 57th percentile. In math, being offered admission to a KIPP elementary school has a positive and statistically significant impact on students' Calculation score of 0.28, equivalent to an increase from the 58th to the 68th percentile. The impact on the Applied Problems score is smaller and not statistically significant.

What are the impacts of KIPP middle schools on student achievement?

We use two different approaches for measuring the impacts of KIPP middle school on student outcomes: the lottery-based design described above in sufficiently oversubscribed KIPP middle schools and a matched-student design in a broader set of KIPP middle schools. In the matched-student design, we identify a treatment group of students who enter KIPP middle schools in grade 5 or 6 and use propensity-score matching to define a comparison group of students—not attending KIPP—who most closely "match" the treatment group in terms of demographic characteristics and baseline test scores. This approach has been previously

Figure ES.3. KIPP elementary school achievement impacts



Notes: Model: Lottery-based design. Outcome: Woodcock-Johnson III Test. Sample size: eight schools; 654 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

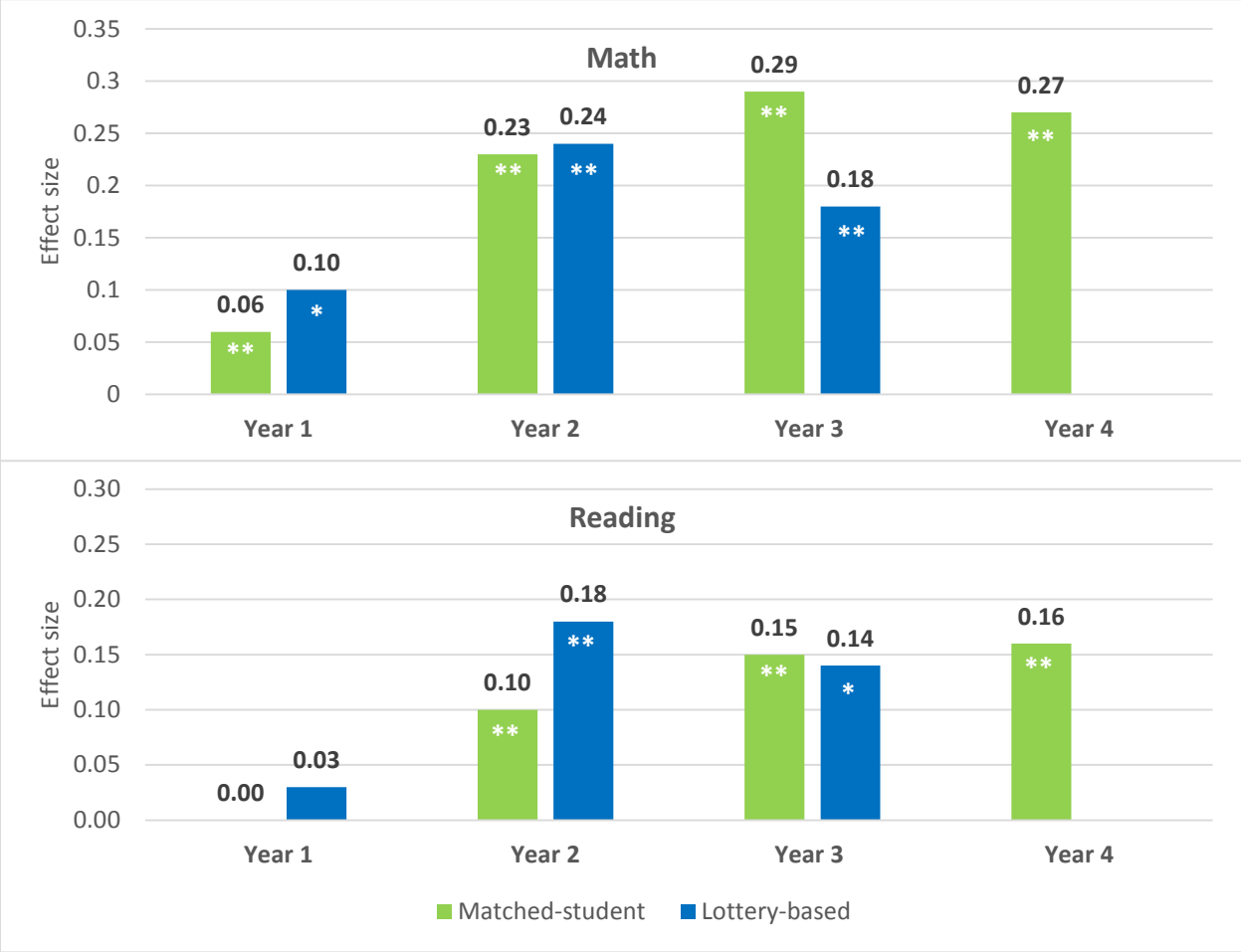
validated using lottery-based results (Tuttle et al. 2013; Fortson et al. 2015) and allows us to include as many KIPP middle schools as possible in our sample. In both designs, the comparison group comprises students attending other charter or traditional public schools.

Across the lottery-based and matched-student designs, our middle school sample includes 43 schools (53 percent of all KIPP middle schools in operation in 2014–2015), including 7 of 21 new KIPP middle schools that opened during the scale-up period. For both designs, we measured academic achievement scores on statewide assessments drawn from state- or district-provided administrative records. Students' scores were standardized (converted into z-scores) using statewide means and standard deviations, so scores represent students' achievement level relative to the typical student in the state at their grade level. We collected test score outcomes corresponding to the first three years after the lottery for the lottery-based sample, and the first four years after the treatment group entered KIPP for the matched-student sample.

Consistent with prior research, KIPP middle schools have positive, statistically significant, and educationally meaningful impacts in math, reading, science, and social studies. Based on both study designs, KIPP middle schools have positive and statistically significant impacts on students' state test scores in both math and reading, by the second year after students are admitted (Figure ES.4). For example, the lottery-based design suggests that being admitted to a KIPP middle school leads to an increase in students' average math score of 0.24 student standard deviation units after two years, equivalent to a student moving from the 40th to the 50th percentile in the state. The two-year reading impact of 0.18 is equivalent to a student moving from the 37th to the 44th percentile. The impact estimates from the matched-student design are similar for a larger sample of 37 schools, suggesting that KIPP middle schools lead to an increase in average math scores of 0.23 standard deviations and reading scores of 0.10

standard deviations. The matched-student design also suggests that, on average, KIPP middle schools have a positive and statistically significant impact of 0.25 standard deviations in both science and social studies (not shown in the figure), equivalent to moving the average student from the 48th percentile to the 58th percentile in science and from the 51st to the 61st percentile in social studies.

Figure ES.4. KIPP middle school achievement impacts



Notes: Model: Lottery-based and matched-student designs. Outcome: State test scores. Sample size: 15 schools, 608 students (lottery-based); 37 schools, 36,798 students (matched-student). Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

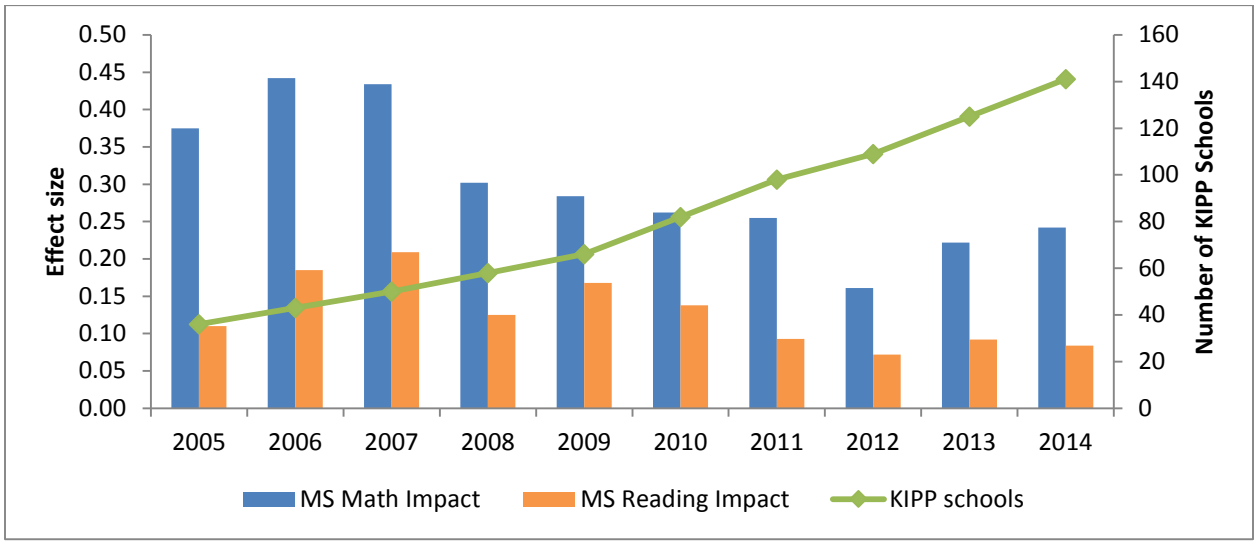
Are the previous positive impacts in KIPP middle schools maintained within the larger network as it scales?

Across the KIPP network, the average impacts of middle schools were positive and statistically significant throughout the 10-year period for which we have data, though higher in earlier years than recent years. To examine changes in the effectiveness of the network over time, we focus on trends in the impacts of KIPP middle schools, since the network has always included middle schools, and added elementary and high schools only in recent years. We calculate the average impact for each KIPP middle school in each school year, using the

results for students two years after KIPP entry. KIPP middle schools have positive and statistically significant impacts in both math and reading for all years from 2005 to 2014. Impacts were largest in 2007 and earlier, especially in math, ranging from 0.38 to 0.50 standard deviations, compared with 0.16 to 0.30 between 2008 and 2014. In 2013 and 2014, when these two-year impacts fully reflect the performance of KIPP schools during the scale-up period, math impacts are 0.22 and 0.24, respectively.

Several factors may explain the trends in KIPP middle school impacts, including changes in the number and composition of schools in the sample, the relative performance of newer versus older schools, and changes over time in the effectiveness of existing KIPP schools as the network has expanded. Overall, KIPP's student achievement impacts decreased during a time of high growth in the network, although they rebounded somewhat during the i3 scale-up period (Figure ES.5).

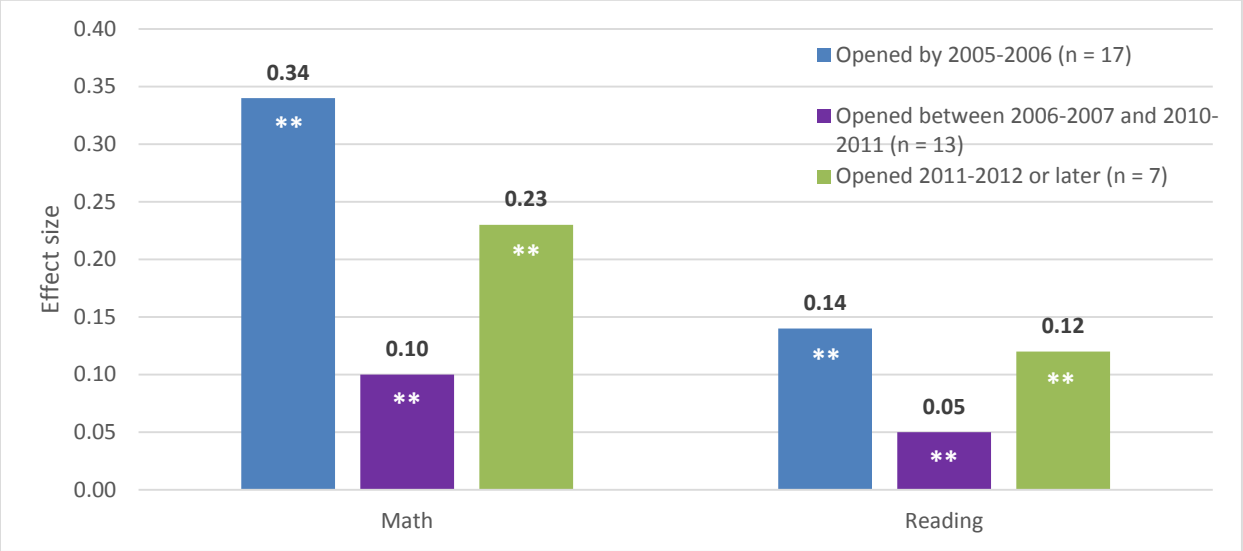
Figure ES.5. Change in the size of the KIPP network and middle school impacts over time



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. Impact estimates are calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. They are estimated separately by school year and plotted using the left-side y-axis. All impacts are statistically significant at the 0.01 level. The year refers to the spring semester of the school year when the achievement exams were taken. The size of the KIPP network is plotted against the right-side y-axis. MS = middle schools.

In fact, the newer KIPP middle schools in our matched student analysis—those opened during the i3 grant period (fall 2011 or later)—have positive impacts on math and reading achievement that are of a similar magnitude of those of the overall impacts for middle schools across the entire study period. When we compare the performance of schools opened during different periods in KIPP's history, we find that the schools opened during the scale up period have impacts that are not quite as large as the oldest KIPP schools (those opened by 2005), but larger than those opened during the period from 2006 to 2010 (Figure ES.6).

Figure ES.6. Impacts of KIPP middle schools on students two years after enrolling, by year opened



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study, based on the year the school opened. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

What are the impacts of KIPP high schools on student achievement?

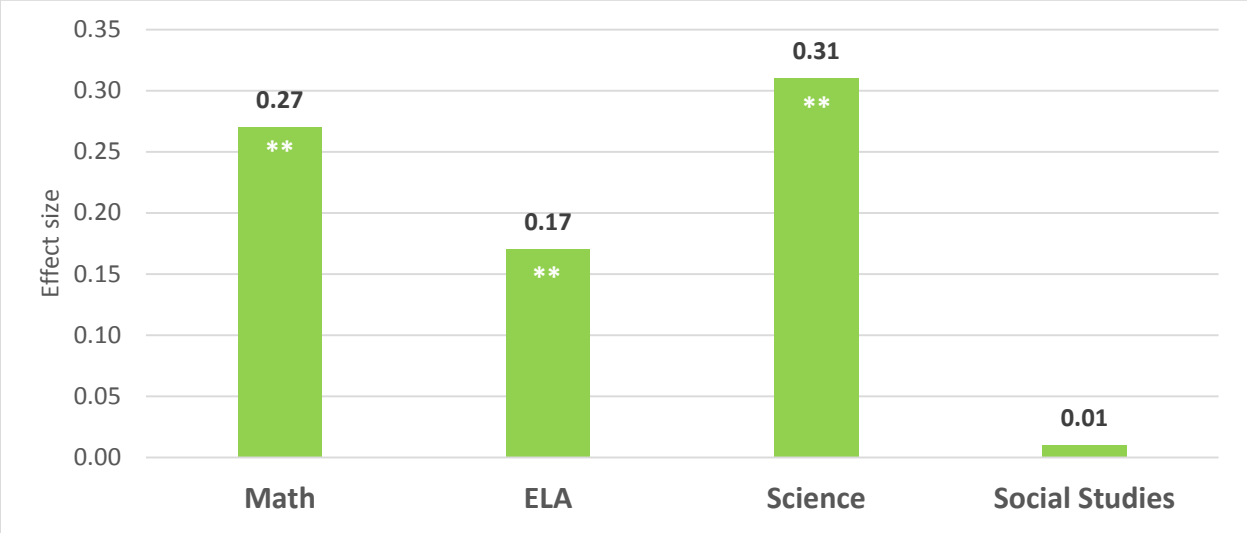
Since students enter KIPP high schools via two routes—from KIPP middle schools and non-KIPP middle schools—we use different quasi-experimental designs to measure impacts on the two groups of students. For the one-third of KIPP high-school students who entered the KIPP network for the first time in grade 9, we use a matched-student design similar to that described above for middle schools. We identify a comparison group for these new entrants, based on demographic characteristics and baseline test scores from grades 7 and 8, of students who attend other charter or traditional public high schools. For the two-thirds of KIPP high school students who also attended a KIPP middle school, we use a matched-school design, comparing outcomes for KIPP middle school students who had the option to attend a KIPP high school with outcomes for a similar set of KIPP middle school students who did not have this option. Whether or not students have the option to enter a KIPP high school depends on the location and timing of their enrollment in KIPP middle schools—in some places and years, the KIPP high school option is present and in others it is absent. Students who do not attend a KIPP high school enroll in a variety of other high schools, including other charter, private, magnet, or boarding schools, in addition to their traditional public school options. This design assumes that aside from the presence/absence of the KIPP high school option, the treatment and comparison groups are similar, on average.

We include 14 KIPP high schools in our matched-student analysis of new entrants and 8 high schools in our matched-school analysis of continuing KIPP students; 4 high schools are included in both designs. Across designs, the high school sample includes 82 percent of all KIPP high schools in operation in 2014–2015. We measure student achievement outcomes using state

assessments for the analysis of impacts on new entrants. For the analysis of impacts on continuing KIPP students, state test scores are less consistently available, because many in the comparison group were attending private high schools or public schools outside the jurisdictions providing data. We therefore measure achievement in the analysis of continuing KIPP students by administering a TerraNova assessment in the third follow-up year after high school entry (typically grade 11).

For new entrants to the network, KIPP high schools have positive, statistically significant, and educationally meaningful impacts on achievement in math, ELA, and science. Having the opportunity to attend a KIPP high school boosts new entrants' high school math scores by 0.27 standard deviation units, a statistically significant impact representing an increase from the 48th to the 59th percentile for the typical student (Figure ES.7). Impacts in ELA and science are 0.18 and 0.31 standard deviations, respectively, and are also significant. Relative to outcomes for the matched comparison group, these impacts are equivalent to an increase from the 47th to the 54th percentile in ELA and from the 42nd to the 54th percentile in science. The average impact in social studies (0.01) is close to zero and not statistically significant. The magnitude of the impact on graduation after four years is positive (four percentage points, not shown), but also not statistically significant.

Figure ES.7. KIPP high school impacts for new entrants



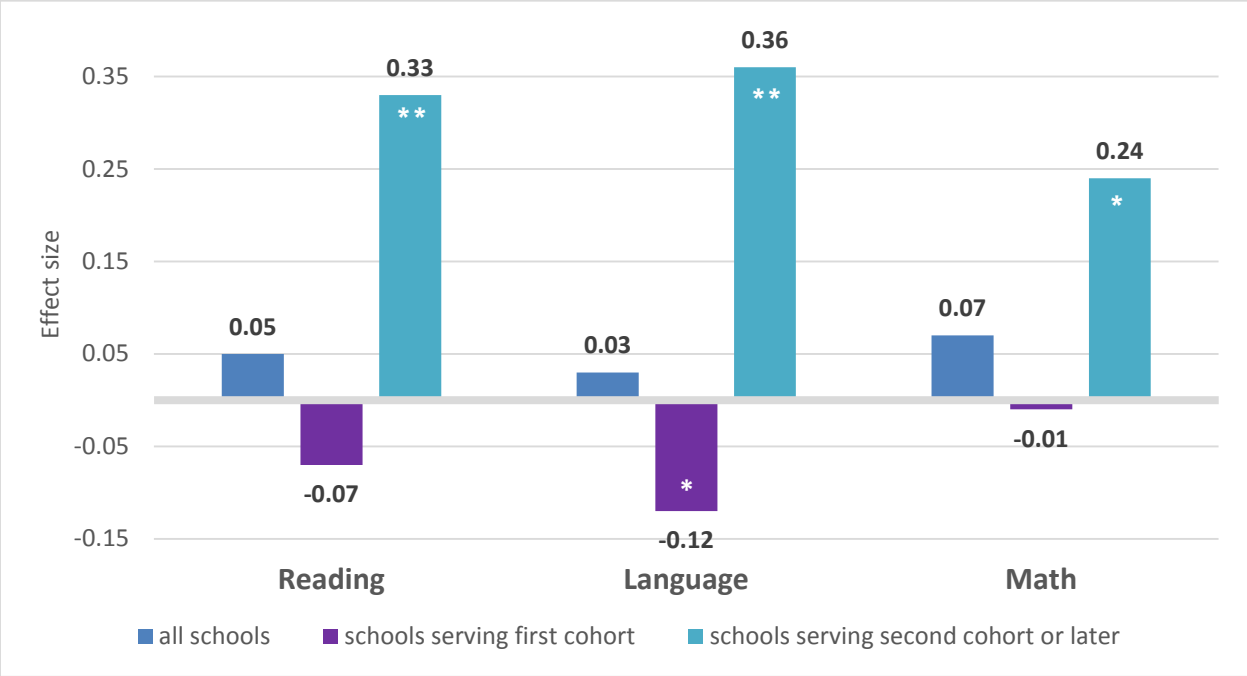
Notes: Model: Matched-student design. Outcome: State test scores. Maximum sample size: 14 schools; 1,861 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

For students continuing from KIPP middle schools, the achievement impacts of KIPP high schools are not statistically significant on average, but these impacts vary by school. For continuing students (compared with KIPP middle school graduates without access to a KIPP high school), the average impacts of KIPP high schools on TerraNova tests in reading, language, and math are positive, but small and not statistically significant (Figure ES.8). These results underestimate the full impact of actually attending a KIPP high school, because all students with the opportunity to attend are included in the treatment group, but not all of them in fact attended. Five of the eight KIPP high schools in this analysis were brand new and serving their first cohort

of students when we measured their impacts. There are significant differences between the impacts of these new KIPP high schools and those of more experienced KIPP high schools. For the five new schools, impacts are negative in all three subjects and statistically significant in language. For the three more experienced high schools, impacts are positive and statistically significant in all three subjects, with magnitudes ranging from 0.24 to 0.36. These more positive impacts for more experienced high schools could imply that KIPP high schools become more effective as they gain experience. Because we do not have data to measure impacts of high schools in multiple years under this design, we cannot determine if KIPP high schools increase their impacts on continuing students as the schools gain experience.

Continuing students with the option to attend a KIPP high school are less likely to drop out of high school. The overall dropout rate is very low, but is significantly lower for the treatment group—1 percent for those who had the chance to attend a KIPP high school and 3 percent for those who did not.

Figure ES.8. KIPP high school impacts for continuing students



Notes: Model: Matched-school design. Outcome: TerraNova test. Sample size: eight schools; 933 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

What are the impacts of KIPP schools on student outcomes other than achievement?

We administered surveys to students and parents to measure impacts on key outcomes other than achievement at all three grade levels.

KIPP elementary and middle schools have positive impacts on school satisfaction, particularly among parents. At both the elementary and middle school levels, being offered

admission to KIPP leads to increases in parents' satisfaction with their child's school. More than three-quarters of elementary parents in the treatment group rate their child's school as excellent, compared to about half of parents in the control group. At the middle school level, 56 percent of treatment group parents and 28 percent of control group parents rate the school as excellent. These findings are consistent with previous research on KIPP in particular and oversubscribed charter middle schools in general (Tuttle et al. 2013; Gleason et al. 2010). Similarly, KIPP has significant positive impacts on a parent-based index capturing satisfaction with school facilities, academics, safety, and discipline. KIPP also has significant positive impacts on several other satisfaction measures, including indices of school efforts to engage parents at both the elementary and middle school level and middle school students' perceptions of their schoolmates. Evidence of KIPP impacts on satisfaction do not extend to the high school level, however, as none of eight measures of impacts on student-reported satisfaction at that level were statistically significant.

KIPP high schools have positive effects on several aspects of college preparation, including discussions about college, applying to college, and coursetaking. KIPP high schools have positive and significant impacts on measures related to school assistance in planning for college, including the frequency of discussions about college at school, students being more likely to have in-depth discussions at school about how to pay for college, and teacher or counselor assistance with planning for college. In addition to assistance provided by the school, KIPP high schools have a positive and significant effect on college preparation activities undertaken by students, as well as on whether the student applied to at least one college or university by spring of senior year—93 percent of treatment students did so, compared with 88 percent of comparison students. Students with the opportunity to attend a KIPP high school enroll in schools more likely to offer advanced placement (AP) or international baccalaureate courses (97 percent versus 89 percent), and the number of AP courses and exams students have taken or intend to take is correspondingly higher.

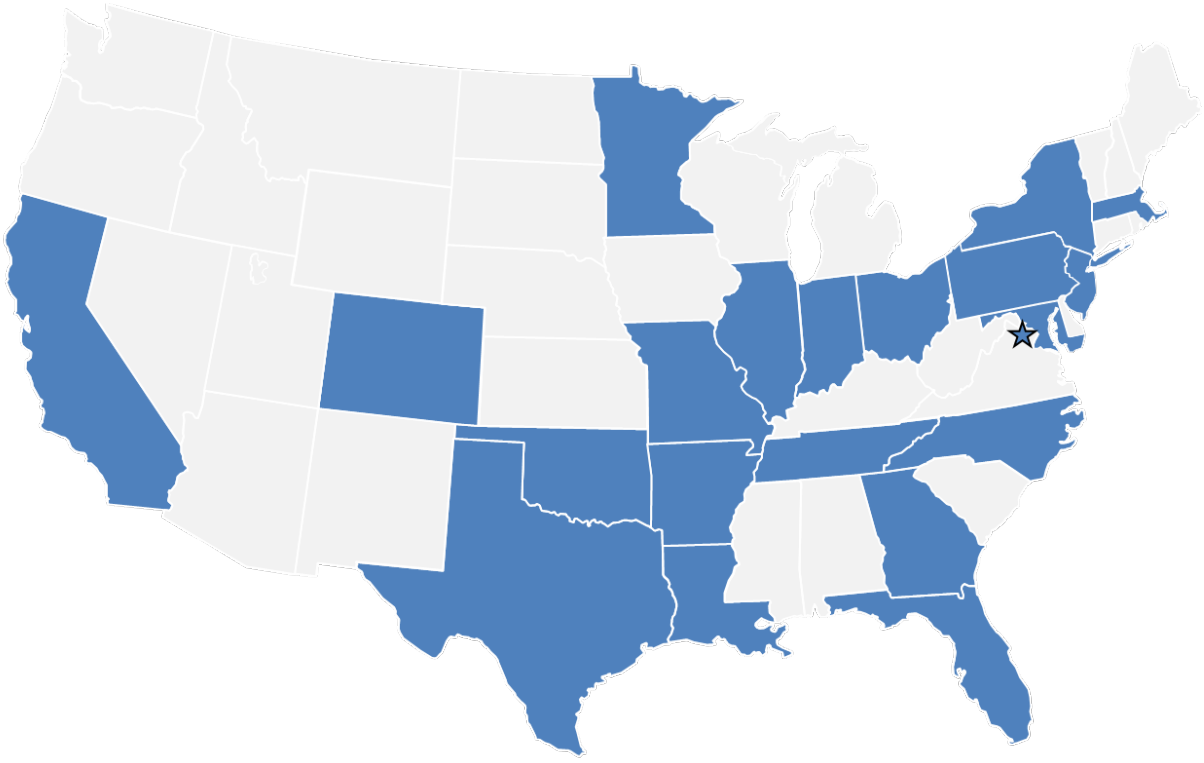
On average, across grade levels, KIPP schools have no statistically significant impact on most survey measures of student motivation and engagement, behavior, or educational aspirations. At all three grade levels, KIPP did not significantly affect measures of motivation and engagement related to student self-control, academic motivation, academic confidence, grit, school engagement, or effort in school, including student reports of the time spent on homework. Student behavior was measured only at the elementary and middle school levels; we find no evidence that KIPP schools affect behavior, including indices of positive behaviors, undesirable behaviors, peer pressure, illegal activities, parental concerns about their child, frequency of school disciplinary actions (according to the parent), and the extent to which the child is well-adjusted. We measure educational goals and aspirations using responses from both parents and the students themselves. In general, the educational goals and aspirations among these elementary, middle, and high school students are high in both the treatment (KIPP) and comparison (non-KIPP) groups. At the high school level, for example, 84 percent of students reported that they think they will graduate from college. For 12 of 13 outcomes, the estimated impact of KIPP is not statistically significant. The single exception is among parents of students at KIPP elementary schools, who are 10 percentage points more likely than the comparison group to believe their child is very likely to complete college (81 versus 71 percent).

I. BACKGROUND

A. The KIPP network of schools

KIPP is a national network of public charter schools with 162 elementary, middle, and high schools operating in the 2014–2015 school year, serving 59,495 students. Nearly all KIPP network schools are operated as part of one of 29 autonomous regional organizations in 20 states and the District of Columbia (Figure I.1).[1] KIPP regions oversee schools in a specific metropolitan or geographic area, providing support on leadership practices, human resources, business operations, technology, and development.

Figure I.1. States in the KIPP network



KIPP schools emphasize rigorous academics and character instruction with the ultimate goal of preparing students to succeed in college and beyond. The KIPP Approach is distinguished by seven key principles, which evolved from the Five Pillars, a set of operating principles which historically guided KIPP schools:[2]

1. A belief that all students can learn and achieve
2. A focus on college graduation as the ultimate goal

[1] Two middle schools are autonomous single sites.

[2] A full description of the KIPP Approach is available at www.kipp.org/our-approach.

3. An emphasis on providing rigorous academics while simultaneously developing student character
4. A belief that visionary, empowered leaders are central to the development and operation of successful schools
5. A belief that excellent teachers are critical to help students succeed in school and beyond
6. A belief that empowered leaders and teachers should leverage existing knowledge and resources when exercising their autonomy
7. A focus on continuous learning and improvement

All 162 KIPP schools in 2014-2015 are public charter schools, and nearly all have been charter schools since they opened. Thus, KIPP schools have greater autonomy to set their own policies than do most traditional public schools, but are accountable to their authorizers for achieving satisfactory performance. Over time, 15 KIPP schools have closed or lost their affiliation with KIPP— in several cases, for performance reasons. KIPP schools have open enrollment policies, but students must choose to apply to and enroll in a KIPP school, and may return to a district school at any time.

B. Findings from prior research

Prior research has consistently found that attending a KIPP school leads to positive effects on student achievement. These positive findings have, in part, fueled KIPP's rapid growth. Using quasi-experimental methods, Mathematica's prior study of 43 KIPP middle schools found positive impacts of KIPP on student achievement across four academic subjects in each of a student's first four years after enrollment and for all examined student subgroups (Gleason et al. 2014). For example, entering a KIPP middle school led to increases in student achievement in math and reading of 0.36 and 0.21 standard deviations after three years, respectively. These are large effects, equivalent to about 90 percent of a year of extra learning in math and about two-thirds of a year of extra learning in reading (Hill et al. 2008). Experimental impacts from this study based on randomized admission lotteries for a much smaller sample of schools and cohorts were consistent with these findings (Tuttle et al. 2013).

Other studies using strong research designs have also found positive impacts on student achievement that were educationally important and statistically significant. In a study of charter management organizations (CMOs), KIPP schools in Washington, D.C., were identified as demonstrating significantly positive two-year impacts in both math and reading, exceeding the impact of the average CMO (Furgeson et al. 2012; Lake et al. 2012). A propensity-score analysis of three KIPP Bay Area (California) middle schools found positive impacts of 0.16 to 0.86 standard deviations on student achievement (Woodworth et al. 2008). Finally, Angrist et al. (2010) used an experimental design based on a randomized admission lottery at KIPP Academy Lynn Middle School (Massachusetts) and found that a year of attendance significantly increased achievement scores by 0.35 standard deviations in math and 0.12 in reading.

Skeptics of KIPP argue that these schools rely on selective admission, attrition, and replacement of students to produce positive achievement results. However, data on student characteristics provide little evidence that KIPP "creams" or selectively enrolls higher-performing students at the middle school level (Tuttle et al. 2013). The typical KIPP student

scored at the 45th percentile within the district—that is, below the district average—in reading and math prior to entering KIPP.[3] Nearly all KIPP students (96 percent) are either black or Hispanic, and more than four-fifths (83 percent) are from households with incomes low enough to make them eligible for free or reduced-price school meals—percentages that exceed the averages at the (non-KIPP) elementary schools they attended prior to enrolling in KIPP middle schools. In contrast, KIPP students are somewhat less likely than students at these feeder elementary schools to have received special education services (9 versus 13 percent) or to have been classified as having limited English proficiency (10 versus 15 percent) when they were in elementary school. Patterns of student attrition from KIPP middle schools are similar to those at nearby non-KIPP public schools (Nichols-Barrer et al. 2015). However, unlike traditional public schools in surrounding districts, when students exit, KIPP schools tend to replace them with higher-achieving students, and fewer students are replaced in the later years of middle school. Still, KIPP's positive achievement impacts do not appear to be explained by advantages in the prior achievement of KIPP students, even when attrition and replacement patterns are taken into account (Nichols-Barrer et al. 2015).

In addition to affecting students' academic achievement, KIPP schools may influence student behaviors and attitudes related to long-term academic success. Using a lottery-based analysis of middle schools, Tuttle et al. (2013) found that KIPP middle schools have no statistically significant effect on a variety of measures of student attitudes that may be related to long-run academic success, including indices of student-reported self-control, academic self-concept, school engagement, effort/persistence in school, and educational aspirations. KIPP similarly does not have a significant effect on several measures of student behavior, including self-reported illegal activities, an index of good behavior, and parent reports of behavior problems. KIPP does have a significant effect on a student-reported measure of undesirable behavior, with KIPP students more likely to report behaviors such as losing their temper, arguing or lying to their parents, or giving their teachers a hard time. By contrast, KIPP leads to a significant increase in the amount of time students report spending on homework and has a positive effect on students' and parents' satisfaction with school.

Despite the growing body of research on the effects of KIPP, three important sets of questions remained prior to this report. First, little research has explored the long-term effects of KIPP middle schools. The KIPP Foundation found that, as of spring 2015, 45 percent of KIPP middle school graduates had earned a four-year college degree in 10 or more years, compared to the national average of less than 10 percent of students from low-income families (KIPP Foundation 2015). These findings are based on the two original KIPP academies in Houston and New York. In addition, these findings have not been supported by studies using more rigorous designs. Second, existing research does not tell us anything about impacts of KIPP schools at levels other than middle school. As KIPP expands up and down, will the model that has been successful in middle schools work at the elementary and high school levels as well? Third, we do not know much about whether KIPP middle schools have been able to maintain their positive impacts as the size of the overall network has grown. Tuttle et al. (2013) measured the impacts of a group of 41 KIPP middle schools that opened in or before 2009–2010, but the overall network

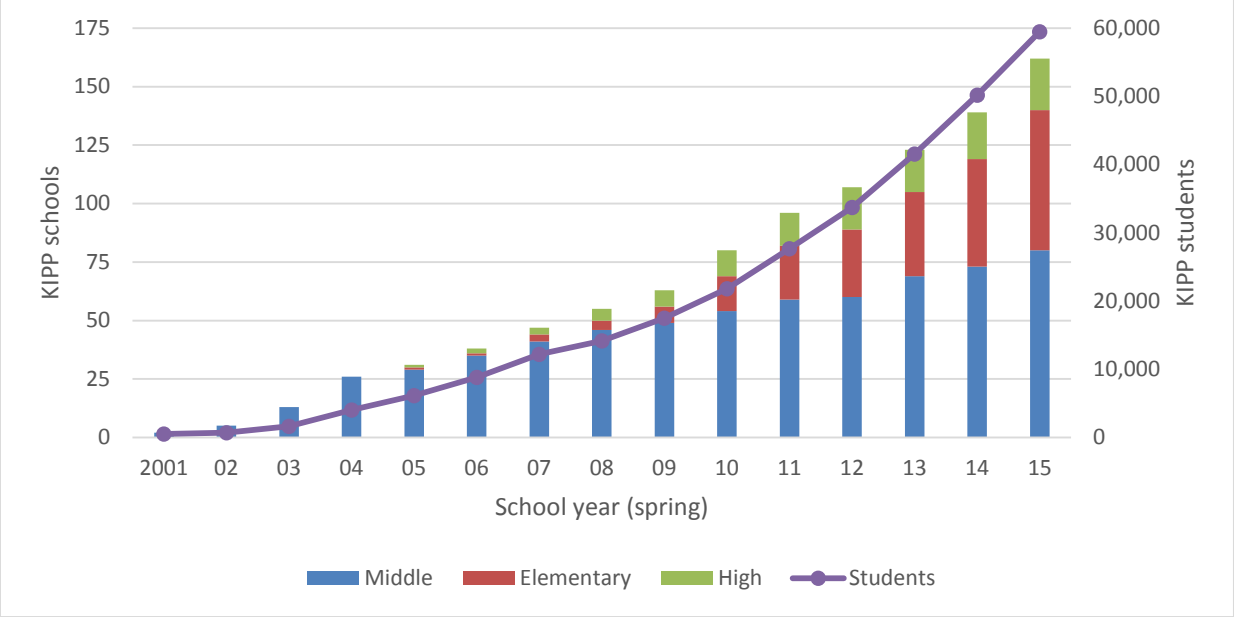
[3] The average achievement level of KIPP middle school students before they entered a KIPP school was also significantly lower than the average among the other students at the elementary schools they attended (Tuttle et al. 2013).

now includes four times that many schools. This study directly addresses the second and third questions.

C. The Investing in Innovation grant program

In 2000, KIPP founders Mike Feinberg and Dave Levin—in partnership with Doris and Don Fisher, founders of Gap Inc.—established the KIPP Foundation to support the expansion of the KIPP network from its first two schools (opened in 1994). The KIPP Foundation guides the network's growth by selecting and training school leaders, promoting the program model, and providing supports and services to KIPP schools and regions on legal, real estate, technology, finance, corporate governance, operations, communications, marketing, and development matters. With the support of the KIPP Foundation, the network has grown dramatically (Figure I.2). Initially, it included only middle schools serving grades 5 to 8 in an increasing number of cities across the country. KIPP expanded into the elementary and high school levels in 2004, with the first such schools opening in Houston. Since then, the majority of growth in the KIPP network has occurred through the opening of new elementary and high schools in cities where KIPP middle schools are already operating.

Figure I.2. Number of KIPP schools and students, by year



Source: KIPP Foundation.

Note: Fifteen schools that closed or left the KIPP network are not included. The network plans to open 21 new schools in fall 2015.

Given its success to date, a key question is whether KIPP can maintain its positive effects if it serves a larger set of students in a larger set of schools. To what extent is KIPP's early success related to the fact that it was a modestly sized organization and focused on the middle school grades? Can this success be maintained when the network has to recruit and retain a larger group of teachers and administrators, as well as attract and serve a larger number of students at different levels? More generally, stakeholders and policymakers wonder whether KIPP's

approach can be adopted more broadly in the charter school sector, or expand to traditional public schools. If the positive impacts of KIPP disappear as the network grows, this would suggest that the model may not work well if adopted by larger school systems. However, if impacts remain positive and strong as the KIPP network continues to expand, then the KIPP model may offer greater promise for public schools more generally.

In October 2010, the KIPP Foundation was awarded a five-year, $50 million Investing in Innovation (i3) scale-up grant by the U.S. Department of Education. The award was one of four i3 scale-up grants that year intended to fund expansion of programs demonstrating strong evidence of previous effectiveness in improving student achievement and attainment. The foundation planned to use the i3 grant to scale up its network while sustaining KIPP's positive impacts. The 2011–2012 school year represented the first full year of i3 grant implementation and marks the beginning of the "scale-up period," which ran through 2014–2015.

KIPP's goals for the proposed scale-up project were threefold. Activities during the scale-up period were focused on (1) increasing the pipeline of effective leaders prepared to lead KIPP schools, (2) placing these newly trained leaders into new or existing KIPP schools grounded in the KIPP Approach, and (3) equipping other public schools to adopt leadership practices that had been successful at KIPP. Table I.1 provides further details of these goals.

Table I.1. Summary of KIPP's project goals

Goal 1	Train 1,000 leaders, including approximately 250 principals, who will each open a new school or assume the leadership of an existing school during the grant period (includes approximately 60 principals outside the KIPP network), and 750 future leaders who will start on the path to school leadership.
Goal 2	Increase annual school openings by at least 50 percent, accelerating from opening an average of 10 schools per year in the last five years to 15 to 18 schools per year during the grant period. Accelerated growth will allow 50,000 students to be served in urban and rural KIPP schools by the end of the grant period, increasing to 66,000 students as those schools reach full enrollment.
Goal 3	Share proven KIPP leadership practices with non-KIPP schools (a) in the urban and rural school districts in which KIPP schools are located and (b) in other growing charter management organizations. By adopting these shared leadership practices, these non-KIPP schools will deepen and expand their own principal pipelines to benefit millions more students.

Source: KIPP i3 grant application; see http://www2.ed.gov/programs/innovation/2010/narratives/u396a100031.pdf.

D. Research questions

As the KIPP network continues to grow into new communities and grades, it faces the dual challenge of effectively serving more students while also building a solid pipeline of leaders to sustain its success. To evaluate its success in achieving these goals, the KIPP Foundation contracted with Mathematica Policy Research to conduct an independent evaluation of the KIPP network and its expansion under the i3 scale-up grant.

The key objective of the evaluation is to measure the impact of KIPP schools on student achievement as the network scales up the number of schools, students, and grades served. In particular, the evaluation addresses the following primary research questions:

- What are the impacts on student achievement of KIPP elementary schools and KIPP high schools?
- What are the impacts of the scaled-up network of KIPP middle schools on student achievement? Are the previous positive impacts in middle schools maintained within the larger network?
- How do impacts of new KIPP middle schools established during the i3 scale-up period compare with those of previously established KIPP middle schools, on average?

In addition, the evaluation addresses a series of exploratory questions of interest:

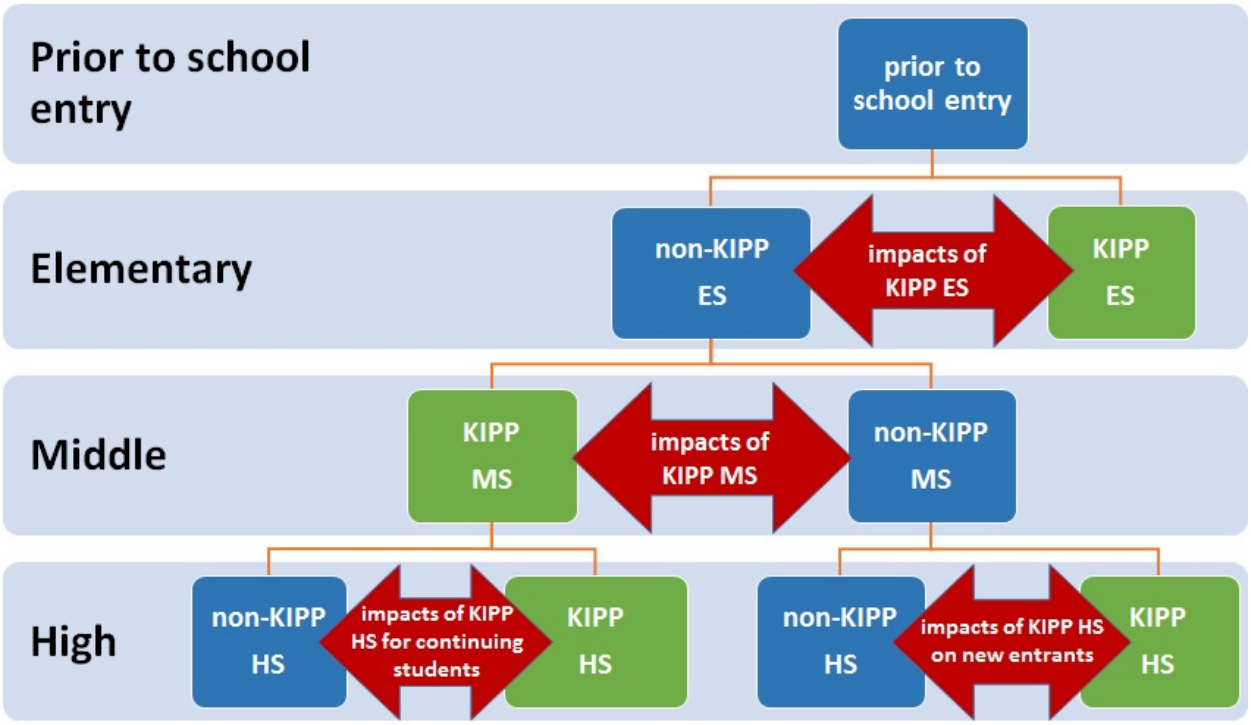
- Do the impacts of KIPP schools on student achievement vary across schools?
- What are the impacts of KIPP schools on student outcomes other than achievement, including:
  - Motivation and engagement?
  - School experiences and satisfaction?
  - Behavior?
  - Goals, aspirations, and college preparation?
  - Likelihood of dropping out of high school?

In the next chapter, we provide an overview of the research designs, data, and samples of schools and students that we employ to answer each of these questions. (More detail on our samples and methods may be found in the appendices to this volume). Chapter III discusses the impacts of KIPP elementary, middle, and high schools on the achievement of their students. Finally, in Chapter IV, we investigate the impacts of KIPP schools on outcomes outside of test scores.

II. STUDY DESIGN

To measure the effect of KIPP on student achievement, we employ the most rigorous possible study design at each school level. Figure II.1 summarizes the designs we use to estimate the effects of KIPP at the elementary, middle, and high school levels. The figure captures students' progression through their elementary and secondary school years, where they may have an opportunity to enter a KIPP school at various points. Among students following a similar schooling path up to each of these points, we compare those who enter KIPP with otherwise similar students who do not. Each of these comparisons serves as a basis for one of our study designs, and is represented in the figure by a red arrow. In each case, the treatment group (represented by green boxes) and the comparison group (represented by blue boxes) come from the same pool of students at the prior level. To measure the effects of KIPP middle schools, for example, we follow students who were in non-KIPP elementary schools and compare those who enter KIPP with those who remain in non-KIPP schools during their middle school years. In comparing outcomes for the KIPP and non-KIPP groups, we use either experimental or quasi-experimental methods to ensure that the groups are otherwise similar at baseline, so that outcome differences we later observe can be credibly attributed to attendance at a KIPP school.

Figure II.1. Overview of study designs, by grades served



An overview of each of these study designs, including information on the samples, outcome data, and years of follow-up, is presented in Table II.1. We discuss each of these in detail in the following pages. A complete list of the schools included in the study is found in Table II.2.

Table II.1. Overview of study designs, by grade level

Grade level	Design	Sample				Outcome data (year collected)
		Schools	Treatment group	Comparison group	Number of cohorts and students	
Elementary (entry at grade PK3 or K)	Lottery-based	Oversubscribed elementary schools (n = 8)	Consenting lottery winners (n = 473)	Consenting lottery losers (n = 624)	1 cohort (1,097 students)	• Study-administered test (WJ-III; year 3) • Parent surveys (year 2)
Middle (entry at grade 5 or 6)	Lottery-based	Oversubscribed middle schools (n = 16)	Consenting lottery winners (n = 459)	Consenting lottery losers (n = 432)	1 cohort (891 students)	• State tests (years 1 through 3) • Student surveys (year 2) • Parent surveys (year 2)
	Matched-student	All middle schools with available data (n = 37)	KIPP MS students with valid baseline test scores (n = 20,312)	Matched comparison students never attending KIPP MS (n = 20,312)	2 to 13 cohorts (40,624 students)	• State tests (years 1 through 4)
High (entry at grade 9)	Matched-student (new entrants)	All high schools with available data (n = 14)	Students entering KIPP from a non-KIPP middle school with valid baseline test scores (in grade 8) (n = 1,380)	Matched comparison students never attending KIPP HS (n = 1,380)	1 to 9 cohorts (2,760 students)	• State tests (year 1, 2, or 3) • State graduation indicator (year 4)
	Matched-school (continuing students)	All high schools with an appropriate comparison cohort (n = 8)	Students from a KIPP middle school with the opportunity to attend a KIPP high school (n = 464)	Students from a KIPP middle school without the opportunity to attend a KIPP HS (n = 469)	1 cohort (933 students)	• Study-administered test (TerraNova; year 3) • Student surveys (year 4)

Notes: The year(s) reported for the outcome measures indicate the number of years after potential entry to a KIPP school during which the outcome data were collected. WJ-III refers to the Woodcock-Johnson III assessment.

We also conducted a matched-student analysis of the cumulative impacts of KIPP middle and high schools on students who attended both types of schools; these results are presented in Appendix C.

Table II.2. Study schools

Region	Level	School	Lottery -Based	Matched -Student	Matched-School
KIPP Atlanta	middle	KIPP South Fulton Academy	X		
KIPP Atlanta	middle	KIPP STRIVE Academy		X	
KIPP Atlanta	middle	KIPP Vision Academy		X	
KIPP Atlanta	middle	KIPP WAYS Academy	X	X	
KIPP Atlanta	high	KIPP Atlanta Collegiate		X	
KIPP Austin	middle	KIPP Austin Academy of Arts & Letters	X	X	
KIPP Austin	middle	KIPP Austin Beacon Prep		X	
KIPP Austin	middle	KIPP Austin College Prep	X	X	
KIPP Austin	middle	KIPP Austin Vista Middle School		X	
KIPP Austin	high	KIPP Austin Collegiate		X	X
KIPP Baltimore	middle	KIPP Ujima Village Academy	X		
KIPP Bay Area	middle	KIPP Summit Academy	X		
KIPP Bay Area	high	KIPP King Collegiate High School			X
KIPP Bay Area	high	KIPP San Jose Collegiate			X
KIPP Charlotte	middle	KIPP Charlotte		X	
KIPP Colorado	middle	KIPP Montbello College Prep		X	
KIPP Colorado	middle	KIPP Sunshine Peak Academy		X	
KIPP Colorado	high	KIPP Denver Collegiate High School		X	
KIPP Dallas-Fort Worth	middle	KIPP TRUTH Academy	X	X	
KIPP DC	elementary	KIPP DC: LEAP Academy	X		
KIPP DC	middle	KIPP DC: AIM Academy		X	
KIPP DC	middle	KIPP DC: KEY Academy		X	
KIPP DC	middle	KIPP DC: WILL Academy		X	
KIPP DC	high	KIPP DC: College Preparatory		X	X
KIPP Delta	middle	KIPP Blytheville College Prep. School		X	
KIPP Delta	middle	KIPP Delta College Preparatory School		X	
KIPP Delta	high	KIPP Delta Collegiate High School		X	
KIPP Eastern North Carolina	middle	KIPP Gaston College Preparatory	X	X	
KIPP Eastern North Carolina	high	KIPP Pride High School		X	
KIPP Houston	elementary	KIPP SHARP College Prep Lower School	X		
KIPP Houston	elementary	KIPP SHINE Prep	X		
KIPP Houston	middle	KIPP 3D Academy		X	
KIPP Houston	middle	KIPP Academy Middle School (Houston)		X	
KIPP Houston	middle	KIPP Courage College Prep		X	
KIPP Houston	middle	KIPP Intrepid Preparatory School		X	
KIPP Houston	middle	KIPP Liberation College Prep		X	
KIPP Houston	middle	KIPP Polaris Academy for Boys		X	
KIPP Houston	middle	KIPP Sharpstown College Prep	X	X	
KIPP Houston	middle	KIPP Spirit College Prep		X	
KIPP Houston	middle	KIPP Voyage Academy for Girls		X	
KIPP Houston	high	KIPP Generations Collegiate		X	

Table II.2 *(continued)*

Region	Level	School	Lottery-Based	Matched-Student	Matched-School
KIPP Houston	high	KIPP Houston High School		X	
KIPP Houston	high	KIPP Northeast College Preparatory		X	
KIPP Houston	high	KIPP Sunnyside High School		X	
KIPP L.A.	elementary	KIPP Raíces Academy	X		
KIPP L.A.	middle	KIPP LA College Preparatory School	X		
KIPP Massachusetts	middle	KIPP Academy Boston Middle School		X	
KIPP Massachusetts	middle	KIPP Academy Lynn Middle School	X	X	
KIPP Massachusetts	high	KIPP Academy Lynn Collegiate High School	X		
KIPP Memphis	middle	KIPP Memphis Academy Middle		X	
KIPP Memphis	middle	KIPP Memphis Collegiate Middle		X	
KIPP Memphis	high	KIPP Memphis Collegiate High		X	
KIPP New Jersey	elementary	SPARK Academy	X		
KIPP New Jersey	high	Newark Collegiate Academy			X
KIPP NYC	elementary	KIPP Academy Elementary School	X		
KIPP NYC	elementary	KIPP Infinity Elementary School	X		
KIPP NYC	middle	KIPP Academy Middle School (New York)	X	X	
KIPP NYC	middle	KIPP AMP Middle School		X	
KIPP NYC	middle	KIPP Infinity Middle School	X	X	
KIPP NYC	middle	KIPP STAR Harlem Middle School	X	X	
KIPP NYC	middle	KIPP Washington Heights Middle School		X	
KIPP NYC	high	KIPP NYC College Prep High School		X	X
KIPP Philadelphia	elementary	KIPP Philadelphia Elementary Academy	X		
KIPP Philadelphia	middle	KIPP Philadelphia Charter School	X		
KIPP Philadelphia	middle	KIPP West Philadelphia Preparatory	X		
KIPP Philadelphia	high	KIPP DuBois Collegiate Academy			X
KIPP San Antonio	middle	KIPP Aspire Academy		X	
KIPP San Antonio	middle	KIPP Camino Academy		X	
KIPP San Antonio	high	KIPP University Prep High School		X	X

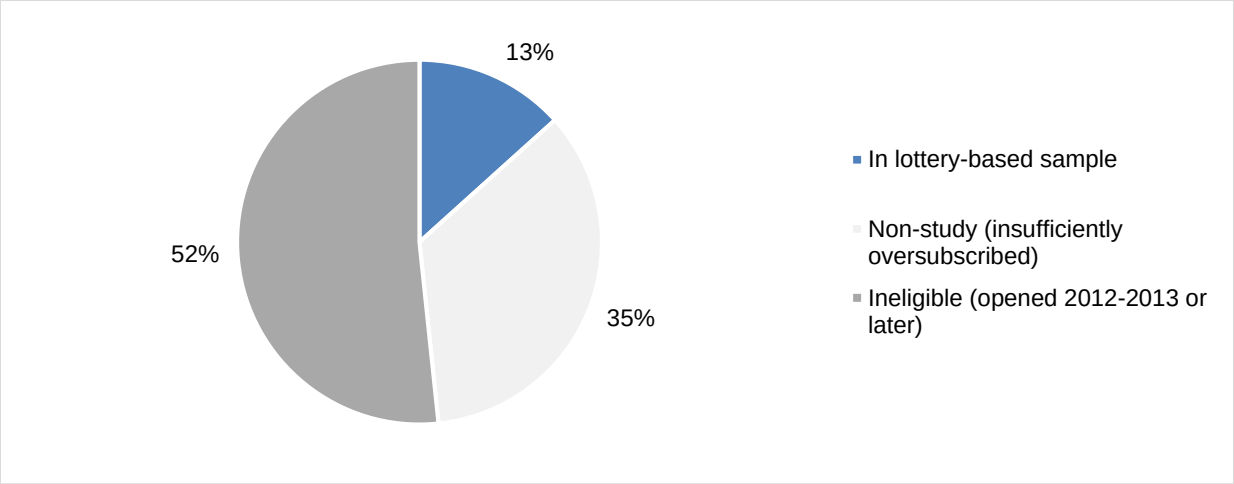
A. Elementary schools

At the elementary school level, we use our most rigorous approach: a **lottery-based design** capitalizing on school admissions lotteries that randomly determine which KIPP applicants are admitted, thereby creating a randomized experiment. Students winning admission to the KIPP school through the lottery form the study's treatment group; those with a poor lottery draw who are not offered admission form the control group. This design ensures that there are no systematic differences in treatment and control group students in terms of key baseline characteristics such as motivation and parental support, prior achievement, and demographics. At the time of admission to KIPP, treatment and control group students are distinguishable only by the luck of their lottery draws, which means that any subsequent differences in their outcomes can be attributed to the impact of attending KIPP.

1. Sample of schools and students

To be included in the lottery-based design, schools need to have substantially more applicants in an entry grade than they have seats to serve students (that is, they must be oversubscribed). In oversubscribed KIPP elementary schools, admissions lotteries are typically held at either pre-kindergarten for 3-year-old students (PK3) or kindergarten (K), depending on the school. Of the 29 KIPP elementary schools open in fall 2011, 8 were sufficiently oversubscribed to be in the analysis (3 with an entry grade of PK3, 5 with an entry grade of kindergarten).[4] Our elementary school sample thus includes 28 percent of KIPP elementary schools in operation during the 2011–2012 school year (the year for which the lotteries for admission occurred) and 13 percent of all KIPP elementary schools in operation by the end of the scale-up period in 2014–2015 (Figure III.2).

Figure II.2. Elementary schools in study sample (2014–2015)



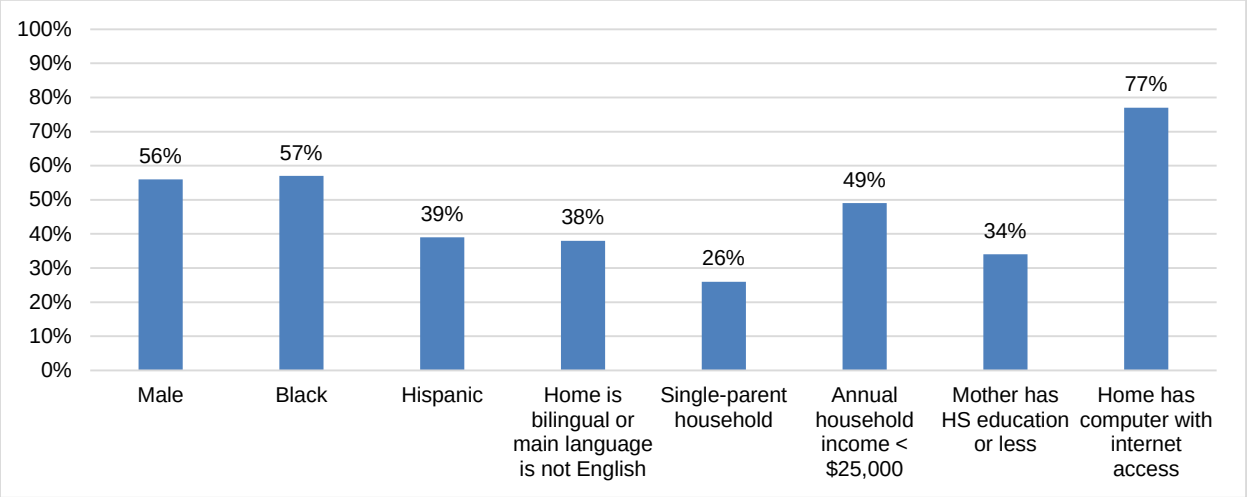
The treatment group comprises 473 students who participated in a lottery and won an offer of admission; the control group consists of 624 students who likewise participated in the lottery but did not receive an offer of admission. Not all students enrolling in a study school in the entry grade are included in the analysis, because some students were admitted outside the lottery (for example, if a student had a sibling already enrolled); in our sample of schools, 61 percent of open slots were filled via lottery. Lottery winners are included in the treatment group regardless of whether they ultimately enrolled in the KIPP school. Similarly, lottery losers are included in the control group regardless of whether they ultimately enrolled in a KIPP school. As a result, the lottery-based design produces estimates of the impact of an offer of admission to a KIPP school (typically referred to as an intent-to-treat, or ITT, estimate) rather than the impact of actually attending a KIPP school. Most lottery winners do attend a KIPP school, however, and most of those not offered admission never attend one. Among students in our sample, 79 percent of lottery winners (treatment group) attended the KIPP school to which they applied, while 6

[4] Of the remaining 21 schools, 11 also conducted lotteries for admission but either exhausted their waitlists (that is, made admissions offers to all lottery participants) or did not provide a sufficiently large treatment or control group. Another six opened in fall 2011 and did not conduct lotteries in spring 2011 (prior to opening).

percent of those who did not win an admissions lottery (control group) ended up attending a KIPP school.

Using information from a baseline survey of the parents of students applying to the KIPP elementary schools included in the lottery-based analysis, we can describe the average characteristics of the students those schools serve (Figure II.3). Most of the students in the elementary school sample are black or Hispanic (96 percent). Parents reported that a language other than English was the main language at home or that the home was bilingual for 38 percent of students, and 34 percent of students' mothers had completed a high school education or less. Approximately half of students are from households with annual incomes of $25,000 or less, and just over one quarter of students are from single-parent households. (Appendix D provides additional detail on sample members' characteristics, types of schools attended, and the baseline equivalence of our samples.)

Figure II.3. Characteristics of KIPP elementary school applicants



Notes: Data obtained from baseline survey of parents of applicants to KIPP elementary schools conducted in spring 2011. Sample includes data obtained for 422 students in the treatment group of the lottery-based baseline sample.

2. Data and methods

To measure elementary school students' academic achievement in reading and math, we administered the Woodcock-Johnson III (WJ-III) assessment in the spring of the third follow-up year after the lottery, when most students who applied at PK3 were in kindergarten and most who applied at kindergarten were in grade 2.[5] Students' WJ-III scores were standardized (into z-scores) using information on the performance of a nationally representative norming population.

[5] Based on grade-level appropriateness, we administered the Letter-Word Identification (Test 1) and Passage Comprehension (Test 9) tests in reading and the Calculation (Test 5) and Applied Problems (Test 10) tests in math. The Calculation test was administered only to the sample of students who participated in lotteries for schools with a kindergarten entry grade. We selected the WJ-III because, relative to other tests for this age-range, it (1) posed a low testing burden on young students in terms of the amount of time it takes to administer and (2) have a reliability for students ages 6 to 9 of over 0.90 for the reading tests and greater than 0.80 for the math tests (McGrew et al. 2007).

Thus, each student's score represents his or her achievement level relative to the national average for students at that grade level: scores greater than zero represent above-average achievement in the domain being tested and scores less than zero represent below-average achievement. Students' behavioral outcomes were measured using a survey of the parents of sampled students conducted in spring of the second follow-up year, when most students who applied at PK3 were in pre-kindergarten as 4-year-olds (PK4) and most students who applied at kindergarten were enrolled in grade 1.[6]

To calculate the impacts of KIPP elementary schools, we compare average outcomes for students in the treatment and control groups using a regression analysis that accounts for the baseline characteristics of sample members. Although a regression approach is not strictly necessary in a randomized experiment—we could simply compare the outcomes of the treatment and control groups—incorporating additional information on baseline characteristics can improve the precision of the results. For more detail on our lottery-based elementary school methodology, see Appendix D.

B. Middle schools

We use two different approaches for measuring the impacts of KIPP middle school on student outcomes. First, we implement the **lottery-based design** described above in sufficiently oversubscribed KIPP middle schools. Second, we employ a **matched-student design** in a broader set of KIPP middle schools, including some that are oversubscribed and some that are not. In this design, we identify a treatment group of students who enter KIPP middle schools in grade 5 or 6 and use propensity-score matching to define a comparison group of students—not attending a KIPP school—who most closely "match" the treatment group in terms of demographic characteristics and (most importantly) baseline test scores from grades prior to KIPP entry.[7] This approach explicitly accounts for differences between the two groups only to the extent that those differences are related to characteristics included in the matching process. Nonetheless, the approach has been validated, successfully replicating lottery-based results when baseline test scores are included among the characteristics used to select the comparison group (Tuttle et al. 2013; Fortson et al. 2015; Gill et al. forthcoming). Moreover, the matched-student design allows us to include many more KIPP middle schools in our sample.[8]

[6] In many cases, the surveys in our study included multiple items capturing the same underlying construct we wished to measure, so we created indices that summarize responses on related data items. We used principal component analysis to identify which group of related items to include in each index. Additional details on all indices are included in Appendix B.

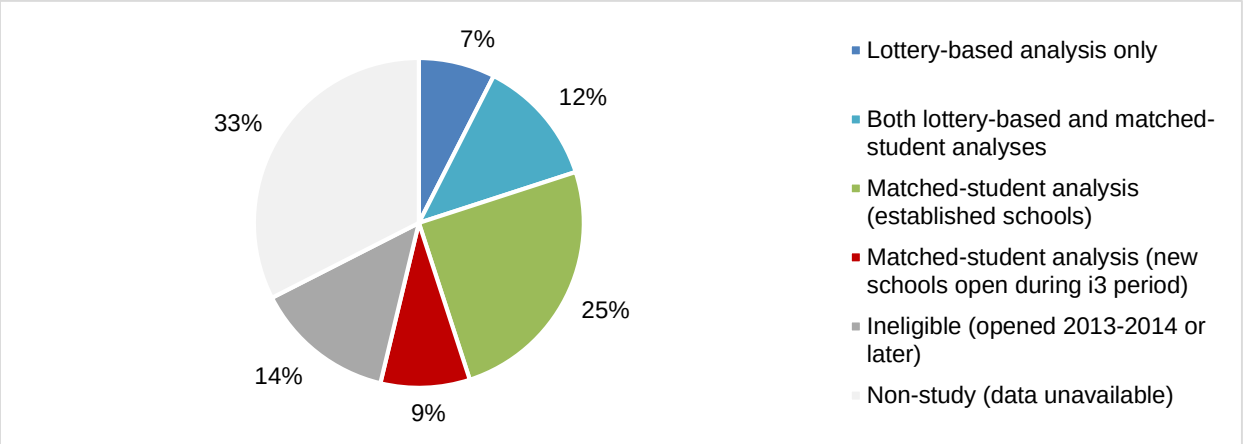
[7] This approach is not feasible for elementary schools in the KIPP network, since there are no available test scores for all students in a district prior to entry at kindergarten or pre-kindergarten.

[8] The matched-student sample includes a combination of schools that were included in our previous report on KIPP middle school impacts (Tuttle et al. 2013) and newer schools that could not be included in that report. Some middle schools that were included in the previous report could not be included here because we could not obtain current data for them. In total, 25 of 37 schools in this report were also included in the analyses for the prior report.

1. Sample of schools and students

Of the 60 KIPP middle schools open in 2011–2012, 16 (27 percent) were sufficiently oversubscribed to include in the lottery-based analysis.[9] We include 37 middle schools in the matched-student design, 10 of which are also in the lottery-based design. The matched-student design includes cohorts of students entering the KIPP middle schools in any year covered by our data.[10] In particular, this analysis includes between 2 and 13 cohorts of entrants, depending on the school.[11] Across the lottery-based and matched-student designs, our middle school sample therefore includes 53 percent of all KIPP middle schools in operation as of the end of the scale-up period in 2014–2015 (Figure II.4), including 7 of 21 new KIPP middle schools (33 percent) that opened during the scale-up period (the 2011–2012 school year or later). Of the other 14 new KIPP middle schools, 11 opened too late (fall 2013 or later) to be included in the matched-student analysis, which requires a minimum of two cohorts of students in each school.

Figure II.4. Middle schools in study sample (2014–2015)



The analytic sample in KIPP middle schools in the lottery-based design includes 891 students, comprising a treatment group of 459 students offered admission and a control group of 432 students not offered admission.[12] As in the elementary school lottery-based design, we measure the impact of admission to—rather than attendance at—a KIPP school, although admission and attendance are closely related. Among lottery participants, 72 percent of treatment group students and 5 percent of control group students attended a KIPP middle school. The matched-student analysis sample includes 40,624 students with valid test score outcome data.

[9] Of the remaining 43 schools, 19 also conducted lotteries for admission but either exhausted their waitlists or did not provide a sufficiently large treatment or control group for the analysis.

[10] Twenty-two KIPP middle schools open in spring 2011 were not included in the matched-student design because it was not possible to acquire the necessary data from the school districts or states in which those schools operated.

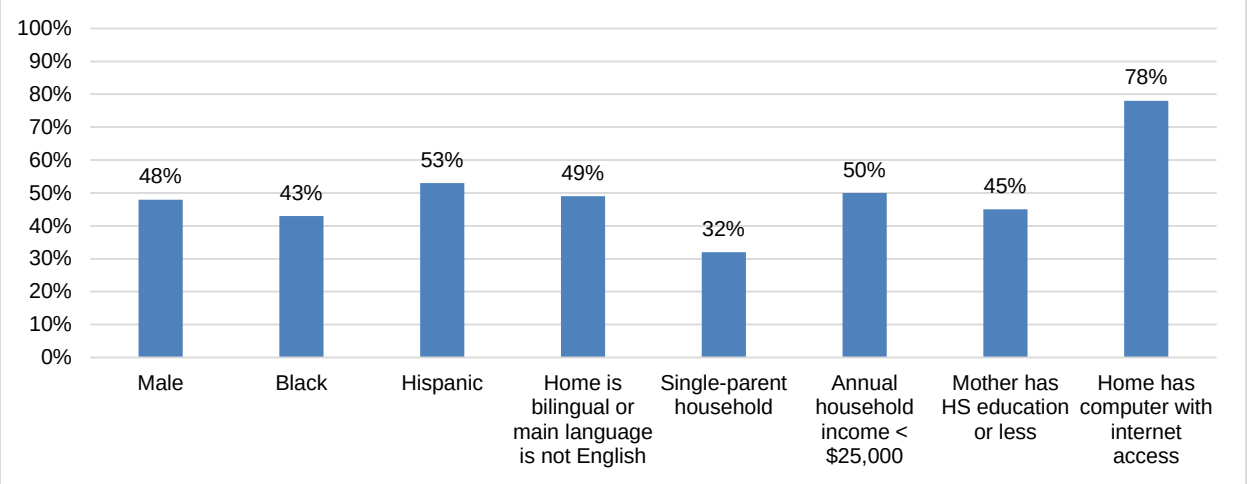
[11] Our outcome data represent the 2001–2002 through 2013–2014 school years, although the specific years and number of cohorts included vary by individual school (based on the years covered by the data provided by each jurisdiction as well as the year in which each school opened). Detail is provided, by school, in Appendix F.

[12] In the entry-grade lotteries for our sample schools, 68 percent of open slots were filled via the lottery; the remaining third of students in the entry grades in those schools were admitted outside the lottery and are therefore not included in the analytic sample.

The treatment group includes 20,312 students who entered a KIPP middle school included in the study at some point over the period 2001–2013; the comparison group of their matched counterparts who did not enter a KIPP school also includes 20,312 students, because we identified one comparison student for each KIPP student.

Applicants to KIPP middle schools in our study sample resemble applicants to KIPP elementary schools in our study sample based on observable characteristics (Figure II.5). Most students in the middle school lottery-based sample are Hispanic (53 percent) or black (43 percent), with less than 5 percent of students in other racial and ethnic categories. About half of students' parents reported that a language other than English was the main language at home or that the household was bilingual, and a similar proportion of students' mothers (45 percent) had completed a high school education or less. In addition, about half of students are from households with incomes of $25,000 or less, and roughly one quarter are from single-parent households. On average, students' math and reading scores in the year prior to the admission lottery were 0.12 and 0.21 standard deviations below their states' mean score, respectively. Appendix E provides additional detail on sample members' characteristics and evidence of baseline equivalence.

Figure II.5. Characteristics of KIPP middle school applicants



Notes: Data obtained from baseline survey of parents of applicants to KIPP middle schools conducted in spring 2011. Sample includes data obtained for 387 students in the treatment group of the lottery-based baseline sample.

Demographic characteristics of the middle school students in the treatment group of the matched-student sample are similar to those in the lottery-based sample; most students are Hispanic (47 percent) or black (51 percent). Most KIPP middle school students qualify for free or reduced-price lunch (89 percent), a proxy for having low family income. Almost 7 percent have special education needs and 10 percent are limited English proficiency students. On average, students' math and reading scores in the year prior to the admission lottery were 0.10 standard deviations below their states' mean score in both subjects. For a detailed discussion of the characteristics of KIPP middle school students relative to their surrounding districts, see Tuttle et al. (2013). Appendix F provides additional detail on sample members' characteristics and the baseline equivalence of our samples.

2. Data and methods

For middle schools in both designs (lottery-based and matched-student), we measured academic achievement scores on statewide assessments drawn from state- or district-provided administrative records. Students' scores on state tests were standardized (converted into z-scores) using statewide means and standard deviations, so scores represent students' achievement level relative to the typical student in the state at their grade level. We collected test score outcomes corresponding to the first, second, and third follow-up year after the lottery for the lottery-based sample and the first through the fourth year after the treatment group entered a KIPP school for the matched-student sample.

For the lottery-based design, we also measured a series of behavioral outcomes via surveys of students and their parents in spring of the second follow-up year, when most study students were in grade 6 (among those who applied for entry in grade 5) or grade 7 (among those who applied for grade 6). We did not measure survey-based outcomes for students in the matched-student design.

In both designs, the primary impact analysis compares average outcomes for the treatment and control groups using a regression approach that accounts for the baseline characteristics of sample members. In the lottery-based design, the main purpose of these control variables is to improve the statistical precision of the KIPP impact estimates. Because the matched-student analysis is quasi-experimental, these control variables play the more important role of accounting for any initial differences between the treatment and comparison groups before the former group entered a KIPP school. As described above, all students initially assigned to the treatment group in both the lottery-based and matched-student designs remained treatment group members throughout the analysis, regardless of whether they were enrolled at a KIPP school.

The specific regression analysis used for the matched-student design was established in our previous work and we also previously conducted a variety of sensitivity tests to check the robustness of the approach (Tuttle et al. 2013). These sensitivity tests were designed to address specific threats to the validity of our impact estimates; that is, possible reasons that our matched-student design could lead to biased or misleading estimates of the impact of KIPP middle schools, including:

- **Students leaving the district/state or moving to private schools (attrition from the sample).** Because the outcome measure is based on administrative records that cover only public school students in the district (or state, in some cases), students in our initial sample who moved to a private school or out of the jurisdiction are lost from our sample. If this occurs at a different rate among KIPP students than it does among comparison group students, this could lead to attrition bias in our impact estimates.
- **Students retained in grade.** When students are retained in grade, they take a different state test than others in their original cohort, making it difficult to compare the achievement levels of students who have been retained with those who have not.
- **Misspecification of the regression**. In the matched-student analysis, we rely on the covariates included in the regression to account for all of the relevant differences (aside

from KIPP attendance) that might influence student outcomes. Failing to properly specify this model could mean that the resulting impact estimates are biased.

The specification tests employed in Tuttle et al. (2013) provide evidence that our regression analysis addresses these threats to validity appropriately. The details of our approach to handling these issues, as well as other aspects of our methodology for the analyses of middle school impacts, are shown in Appendices E and F.

C. High schools

Unlike students in KIPP elementary and middle schools, most students do not enter a KIPP high school via a lottery. Among students enrolling in a KIPP high school in our sample, two-thirds (68 percent) came directly from attending a KIPP middle school in grade 8. As a result, we could not employ a lottery-based design at the high school level during the scale-up period, and instead use a quasi-experimental approach to measure the impacts of KIPP high schools. Moreover, since students can enter KIPP high schools via two routes—from KIPP middle schools and non-KIPP middle schools—we use two different quasi-experimental designs.

To measure the impact of KIPP high schools for the subset of students (32 percent) who enter the KIPP network for the first time in grade 9 (new entrants), we use a **matched-student design** similar to that described above for middle schools. In particular, we use propensity-score matching to identify a comparison group for these new entrants on the basis of demographic characteristics and baseline test scores from grades 7 and 8. As with the middle school matched-student design, we measure outcomes for all students in the initial sample for whom we have outcome data: treatment students who entered KIPP in grade 9 but ultimately moved back to a non-KIPP school are kept in the treatment group.[13]

To measure the marginal effect of KIPP high schools on student outcomes for those students who also attended a KIPP middle school ("continuing students"), we use a **matched-school design**. This design involves two different but complementary sets of comparisons, each comparing outcomes for a set of KIPP middle school students who had an option to attend a KIPP high school with outcomes for a similar set of KIPP middle school students who did not have an option to attend a KIPP high school. In each case, treatment and comparison group students are similar in that they all made the decision to enter a KIPP middle school, but they differ in that only the treatment group had the opportunity to enter a KIPP high school. Whether or not students had the option to enter a KIPP high school depends on the location and timing of their enrollment in KIPP middle schools—in some places and years, the KIPP high school option was present and in others it was absent. Specifically:

[13] For students who attended a KIPP high school after also attending a KIPP middle school, we can use a similar matched-student approach, but must conduct the matching at an earlier baseline (grade 4, prior to middle school entry and when both groups attended non-KIPP schools), as well as restricting the treatment group to students who entered a KIPP middle school and then persisted into a KIPP high school. This analysis provides a cumulative measure of the effects of attending both a KIPP middle and high school on outcomes at the high school level, but does not allow us to distinguish between the separate impacts of KIPP middle and high schools. For this reason, as well as other limitations of this approach, the results from this analysis and their limitations are not discussed in detail in Chapter III, but are included in Appendix C.

- For the first matched-school approach, we focus on a set of KIPP high schools in their first year of operation, and examine students graduating from the same KIPP middle school *in adjacent cohorts*. The treatment group in this design includes KIPP middle school students in grade 8 who had the option to attend the local KIPP high school in its first year of operation. The comparison group includes the previous cohort of KIPP students from the same middle school in grade 8 who did not have the option to attend the local KIPP high school because it had not yet opened. This design includes all five KIPP high schools that opened during the 2008–2009, 2009–2010, or 2010–2011 school years and that were served by KIPP middle schools with at least one cohort of students enrolling in grade 9 before the KIPP high school opened (that is, all schools with an eligible comparison group).
- For the second matched-school approach, the treatment group includes KIPP middle school students in grade 8 who had the option to attend the local KIPP high school. The comparison group includes KIPP students in grade 8 *from different middle schools* (in the same year) in regions with no KIPP high school open at the time. In other words, we compare students graduating from KIPP middle schools that fed into a KIPP high school with those graduating from KIPP middle schools that had no high school to feed into. To define a sample that was equivalent at baseline (grade 8), we identified the comparison KIPP middle schools that most resembled the treatment middle schools on the basis of average school-level characteristics. Then, within these matched sets of schools, we conducted student-level propensity score matching to identify the individual comparison students who most closely matched the treatment students.[14] For this design, we include all five KIPP high schools that opened prior to 2010 and were fed by KIPP middle schools with credible matched schools—two of which were also included in the approach using adjacent cohorts discussed above, for a total of eight unique schools in the matched-school designs.

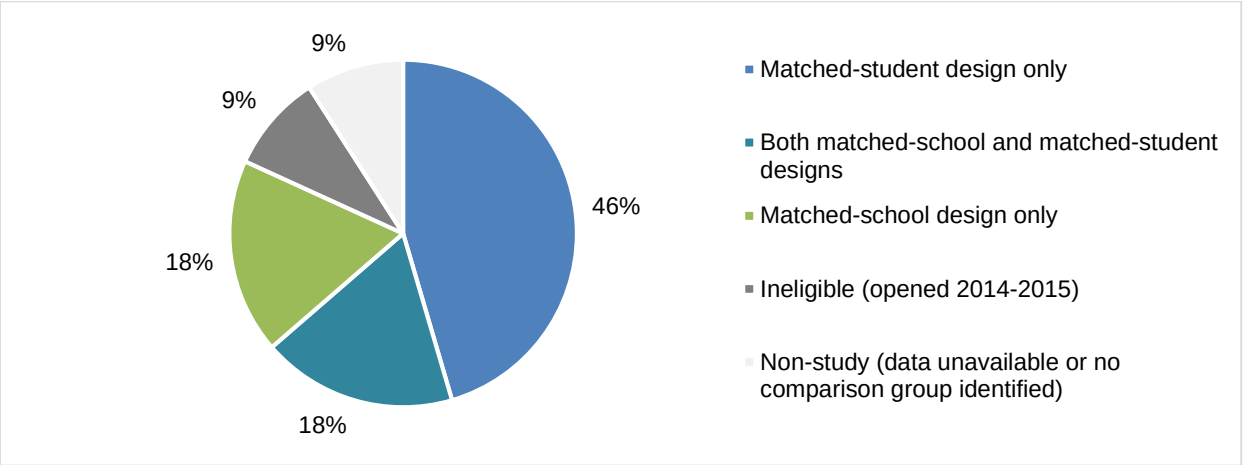
These designs assume that, aside from the presence/absence of the KIPP high school option, the treatment and control groups are similar to one another, on average. This assumption is based on the argument that the presence/absence of the high school option is largely beyond the control of individual students. The two matched-school designs ended up producing similar estimates of the impact of KIPP high schools for those high schools included in both approaches. Thus, we combine the results of the two matched-school designs for the purposes of this report. For more detail on this methodology, and results that are separate by model, see Appendix H.

1. Sample of schools and students

We include 14 KIPP high schools in our matched-student analysis of new entrants, and 8 high schools in our matched-school analysis of continuing KIPP students (including 4 that were also in the matched-student analysis). Across all the designs, our high school sample includes 82 percent of all KIPP high schools in operation as of the end of the scale-up period in 2014–2015 (Figure II.6). For information regarding the specific schools in these study samples, see Appendix A.

[14] Although we employed school-level matching to identify the eligible pool of comparison schools, we did not require a one-to-one match at the school level. Therefore, two different students from the same treatment school may each be matched to comparison students from different comparison schools, based on their individual characteristics.

Figure II.6. High schools in study sample (2014–2015)



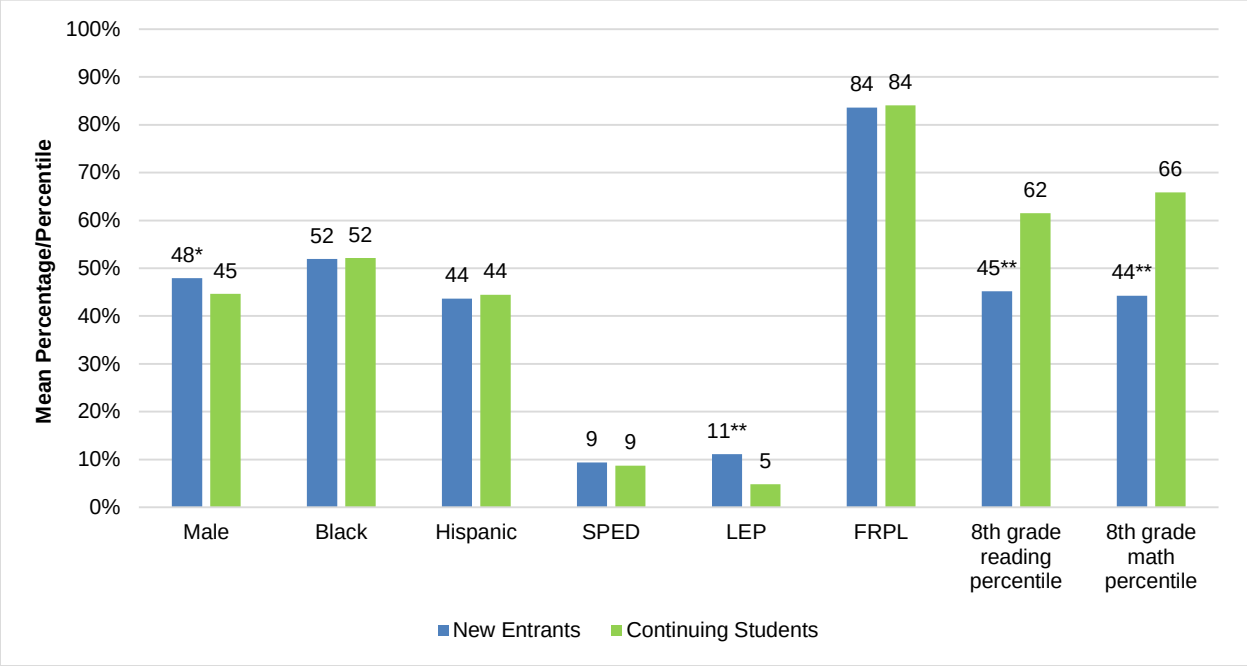
In the matched-student design of the impact of KIPP high schools for new entrants, we have a sample of 2,760 students for whom we have valid test score outcomes. The treatment group includes 1,380 students who entered one of the 14 KIPP high schools in the analysis for the first time in grade 9 at some point over the period 2005–2013. The comparison group of their matched counterparts includes 1,380 students who entered grade 9 at a non-KIPP school over this same period. In the matched-school design of the impact of KIPP high schools for continuing students, our sample includes 933 students with valid test score outcomes. The treatment group in this design includes 464 KIPP middle school students who had the opportunity to enter grade 9 in one of the 8 KIPP high schools in the analysis between fall 2007 and fall 2010. The comparison group in the matched-school design includes 469 KIPP middle school students who did not have the opportunity to enter a KIPP high school.[15] Although all treatment group students in the matched-student analysis entered a KIPP school initially, treatment group students in the matched-school analysis may or may not have entered the KIPP high school they had the option of entering. The matched-school analysis therefore measures the effect of the *availability* of a KIPP high school rather than *enrollment* in a KIPP high school. However, rates of enrollment in KIPP high schools among treatment group students are generally high, at 71 percent overall and ranging from 59 to 83 percent across the feeder middle schools for high schools in our sample. The comparison group attended a wide variety of non-KIPP high schools, including private, magnet, boarding, traditional public, or non-KIPP charter schools (see Appendix H for more detail).

The two different groups of KIPP high school students—new entrants and continuing students—are similar in several respects, but differ in important ways. Demographic characteristics of new entrants and continuing students are similar, with roughly equal proportions of black and Hispanic students, special education students, and students who qualify for free or reduced-price lunch (Figure II.8). New entrants are slightly more likely than continuing students to be male, and are twice as likely to have limited English proficiency. Most strikingly, the two groups differ with respect to their achievement level prior to high school

[15] In the matched-school sample, the size of the treatment group differs slightly from the size of the control group because students were not matched one-to-one in the matched-school design relying on adjacent cohorts.

entry. KIPP high school students who attended a KIPP middle school were much higher achieving prior to high school than those who are new entrants, with baseline (grade 8) scores 11 percentile points higher in reading and 16 points higher in math. As a result, although continuing KIPP high school students were higher achieving in grade 8 than students at non-KIPP high schools, new entrants into KIPP high schools had lower grade 8 achievement scores than non-KIPP high school students. Appendix G provides additional detail on sample members' characteristics and evidence of baseline equivalence for the matched-student analysis of new entrants, and detail on sample members' characteristics and evidence of baseline equivalence for the matched-school analysis of continuing students is included in Appendix H.

Figure II.8. Characteristics of continuing students and new entrants to KIPP high schools



Notes: Bars represent mean characteristics for two different types of KIPP high school students—continuing students and new entrants. Differences between the two KIPP samples are statistically significant at the 0.05 level (*) or the 0.01 level (**), two-tailed test, as noted by the asterisks in the graph. Each KIPP high school is given equal weight to calculate the overall average and statistical significance.

2. Data and methods

We used administrative data from jurisdictions (states or districts) hosting at least one KIPP high school to measure student achievement outcomes based on state assessments for the matched-student analysis of new entrants to KIPP high schools. State assessments tend to be administered differently in high schools than in middle schools. A majority of the high school tests are end-of-course exams in a single subject such as algebra I or biology that students may take at different points during their high school careers.[16] We obtained data on students' test scores from grade 9, 10, or 11, and the analysis is based on the first end-of-course test score in a

[16] A small number of jurisdictions provided data from end-of-grade exams administered to all students in a given grade.

given subject observed for each student (that is, we disregard retests for students who took the same subject in multiple years). We use as outcomes only those subject tests that were taken by a majority of students and by roughly equal proportions of KIPP and non-KIPP students, and where the timing of the test did not differ greatly between the treatment and matched comparison group.[17]

Our analytic approach for the matched-student analysis of new entrants to KIPP high schools mirrors our approach to the middle school matched-student analysis, with a few exceptions. When conducting propensity score matching, we matched the treatment group of new entrants to a comparison group of students in the same grade and district who did not go on to attend a KIPP high school but instead attended a non-KIPP public high school. We matched these groups on grade 7 and 8 test scores and demographic characteristics. As in the middle school analysis, KIPP students transferring out of KIPP high schools remained in the treatment group. For a detailed discussion of our analytic methods, including additional details about the tests administered, see Appendix G.

For the matched-school design of continuing KIPP students, we used a different data source to measure the student achievement outcome, because many students in this sample attended either private schools or schools in other states, such that our district or state administrative data did not include achievement test scores for these students. Thus, we measured achievement levels for students in this sample by administering a TerraNova assessment in the third follow-up year after high school entry (typically grade 11) to these students.[18] To measure outcomes reflecting student behavior and attitudes for students in the matched-school sample, we administered a student survey conducted in spring of the fourth follow-up year after high school entry (typically grade 12).

To estimate the impact of KIPP high schools on achievement and non-achievement outcomes for our matched-school designs, we compare average outcomes for the treatment and comparison groups using a regression that controls for baseline characteristics of matched sample members. See Appendix H for a detailed discussion of these methods.

[17] Because of the characteristics of high school assessments, it is important to remember that the impact estimates for each of these high school outcomes may not be directly comparable to the impact estimates for KIPP middle schools. For instance, there may be differences between the types of knowledge tested in high school end-of-course exams (which tend to be offered in multiple different grade levels) relative to the middle school exams (which have a separate test designed for each grade level). Similarly, high school exams tend to measure knowledge in academic areas (such as geometry or biology) that are more specialized than the general types of mathematics and literacy knowledge measured in statewide middle school tests.

[18] We administered Form G, Level 21/22 in reading, language, and math to the students in the sample, calculating z-scores that were standardized to capture student achievement relative to that of a nationally representative norming population.

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

III. KIPP'S IMPACTS ON ACHIEVEMENT AND GRADUATION

Network-wide, KIPP schools have positive, statistically significant, and educationally meaningful impacts on student achievement, particularly at the elementary and middle school grades. KIPP elementary schools have positive impacts on students' reading and math achievement. KIPP middle schools also have positive impacts on reading and math, as well as on achievement in science and social studies. For KIPP high schools, impacts on achievement are positive for some schools and groups of students, including new entrants to the network, but not statistically significant for others. We do not find evidence that KIPP high schools have statistically significant impacts on the probability that a student graduates in four years, but they do reduce the likelihood of dropping out of high school by a statistically significant amount.

Across the network as a whole, average impacts of KIPP at the middle school level (where the longest trends can be examined) have been consistently positive in both reading and math from 2005 through 2014, even as the system has grown rapidly. Average impacts have declined since 2007, but newly opened KIPP middle schools are producing positive impacts that are generally similar to those achieved by older KIPP middle schools when they were in their first years of operation. Older KIPP middle schools have shown some small declines in impacts, though they remain significantly positive.

A. Elementary school achievement

The KIPP network initially included only middle schools. The first KIPP elementary school opened in 2004–2005; by 2010–2011 when the i3 grant was awarded, there were 23 KIPP elementary schools. During the scale-up period, the number of KIPP elementary schools almost tripled, to 60 in operation as of the 2014–2015 school year, making them a much more important part of the KIPP network. No previous research has provided evidence on the effects of KIPP elementary schools.

Based on the randomized design described in Chapter II, we estimate the impact of being offered admission to a KIPP elementary school on students' early achievement in reading and mathematics, measured three years after the schools' admissions lotteries. As described previously, these outcomes are measured at different points in students' elementary school careers—at kindergarten for students who applied to enter KIPP at PK3 and in grade 2 for those who applied to enter KIPP in kindergarten. We find the following:

KIPP elementary schools have positive, statistically significant, and educationally meaningful impacts on three of four measures of students' reading and mathematics skills. On tests administered three years after entry, being offered admission to a KIPP elementary school has positive and statistically significant impacts on two measures of students' reading achievement (WJ-III Letter-Word Identification and Passage Comprehension) three years after random assignment (Table III.1). The Letter-Word Identification assessment requires students to orally name letters and words; in the Passage Comprehension assessment, students are required to read a printed passage and orally name a missing key word that makes sense in the context of

that passage.[19] Being offered admission to a KIPP elementary school is estimated to lead to an increase of 0.25 standard deviation units on the Letter-Word Identification test and 0.22 on the Passage Comprehension test.[20] The impact of an offer of admission on Letter-Word Identification is approximately equivalent to a student moving from the 78th percentile (the percentile corresponding to the control group students' mean score) to the 84th percentile; the impact on Passage Comprehension scores is equivalent to a move from the 48th percentile to the 57th percentile.

Table III.1. Impacts of offer of admission to KIPP elementary school

Outcome (Z-Scores)	Mean, lottery winner	Mean, lottery loser	Impact estimate	Number of schools	Number of students
Math achievement					
Calculation [a]	0.48	0.20	0.28** (0.11)	5	371
Applied Problems	0.04	-0.03	0.07 (0.05)	8	652
Reading achievement					
Letter-Word Identification	1.01	0.76	0.25** (0.07)	8	651
Passage Comprehension	0.18	-0.04	0.22** (0.07)	8	648

Notes: Outcomes are measured on Woodcock-Johnson III, administered in the spring of the third follow-up year. All impacts in this table are intent-to-treat (ITT), based on regression models that pool all lottery elementary schools and that control for baseline covariates. Standard errors are in parentheses. Measures of the complier average causal effect (CACE, sometimes referred to as a treatment-on-treated or TOT estimate) for each outcome are provided in Appendix D. Means for lottery losers are unadjusted; means for the lottery winners are equal to lottery losers' mean plus the estimated impact.

[a] Subtest administered only to students in grade 2 at the time of assessment.

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

In math, being offered admission to a KIPP elementary school has a positive and statistically significant impact on students' Calculation score (0.28 standard deviation units), but the impact on the Applied Problems score is smaller and not statistically significant (Table III.1). In the Calculation test, administered only to students in the grade 2 sample, students are required to solve printed mathematical calculations (for example, "3 - 1 = ?"). In the Applied Problems test, students are required to perform math calculations in response to orally presented problems.[21] The magnitude of the offer of admission to a KIPP elementary school on students' Calculation

[19] For kindergarteners, the Passage Comprehension test begins with a series of questions requiring the student to choose the appropriate rebus (representation of a word or phrase using pictures or symbols that suggest its syllables) based on an image, or choose the correct image based on a printed phrase.

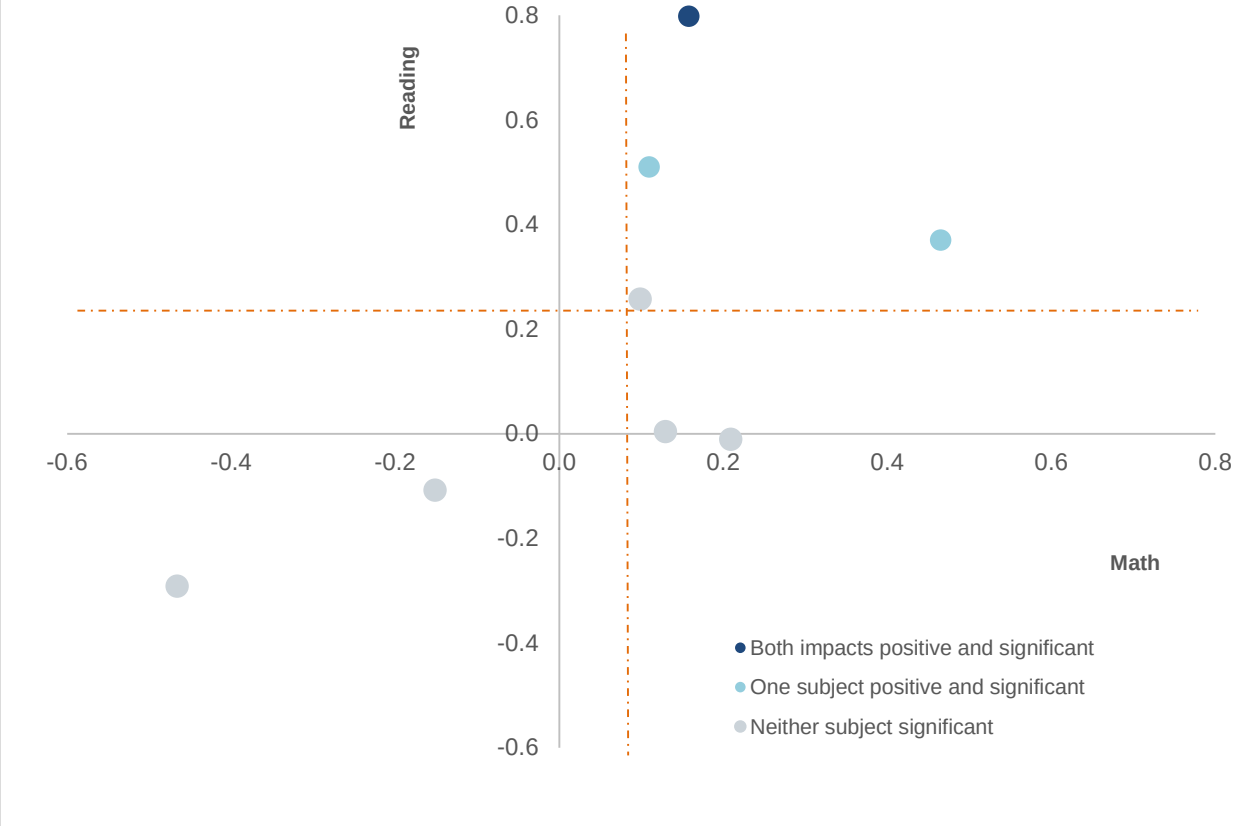
[20] Recall that these are intent-to-treat (ITT) estimates of the effect of an offer of an admission to a KIPP elementary school, regardless of whether the student ultimately enrolled in a KIPP elementary school. In other words, the treatment group includes students that did not actually attend KIPP schools.

[21] For kindergarteners, the test begins with a series of questions requiring the student to identify the number of instances an object appears in an image containing other objects.

scores is 0.28 standard deviation units, approximately equivalent to moving a student from the 58th percentile to the 68th percentile.

The impact of KIPP at the elementary school level varies by school. Among the eight KIPP elementary schools we examine, three have positive and statistically significant impacts on at least one measure of academic achievement (Figure III.1). Two additional schools have marginally significant positive impacts on one outcome ($p < 0.10$). There are no statistically significant impacts in the other three schools. In these schools, the point estimates for one school are positive and the estimates for two schools are negative in both subjects.

Figure III.1. Distribution of elementary school reading and math impact estimates after three years



Note: Each circle represents the average math impact estimate (across tests) and average reading impact estimate (across tests) for one KIPP elementary school. Dark blue circles indicate that at least one impact in both subjects is statistically significant and positive at the 0.05 level, two-tailed test. Light-blue circles indicate that at least one impact in only one of the two test subjects is statistically significant and positive. Grey indicates that all impacts are statistically indistinguishable from zero. No school had a statistically significant and negative impact in either subject in any test. The orange lines represent the average impacts across KIPP elementary schools.

B. Middle school achievement

Previous research finds that KIPP middle schools have positive impacts on student achievement, boosting the test scores of students above the levels they would have achieved had they attended non-KIPP schools (Gleason et al. 2013; Angrist et al. 2010; Woodworth et al. 2008). However, these studies measured KIPP impacts at a time when the network was limited in size, prior to the recent increase during the i3 scale-up period. The number of middle schools increased 33 percent between 2011–12 and 2014–15, while the number of schools across the entire network increased 51 percent over the same period.

The growth in the number of KIPP schools could present challenges for the network, making it difficult to maintain the positive impacts on student achievement found in prior studies of KIPP middle schools. When multiple KIPP middle schools serve the same district region, the network must recruit a larger number of students to attend KIPP middle schools; identify, recruit, and train a larger number of effective leaders; bring on board and retain a larger number of qualified teachers; and address bureaucratic issues that arise when a group of related schools must operate in the same educational market. Failing to adequately meet these challenges could result in a watering down of positive KIPP impacts.

A key objective of this report is to determine whether the KIPP network has maintained positive impacts among its middle schools. Do the existing schools, which have been successful in the past, produce equally positive impacts when faced with the challenges caused by a growing network? Are newly opened KIPP middle schools as successful as the schools KIPP has previously opened? Using both lottery-based and matched-student designs for measuring the impacts of KIPP middle schools, we find the following:

Consistent with prior research, KIPP middle schools have positive, statistically significant, and educationally meaningful impacts in math, reading, science, and social studies. Based on both study designs, KIPP middle schools have positive and statistically significant impacts on students' state test scores in math and reading, beginning two years after students are admitted (Table III.2).[22] For example, the lottery-based design suggests that being admitted to a KIPP middle school leads to an increase in students' average score in math of 0.24 student standard deviation units after two years, equivalent to a student moving from the 40th percentile to the 50th percentile in the distribution of his or her state math test. The two-year impact in reading of 0.18 standard deviation units is equivalent to a student moving from the 37th percentile to the 44th percentile in the distribution. The impact estimates are similar for the matched-student analysis based on a larger sample of 37 schools and multiple cohorts per school, suggesting that KIPP middle schools lead to an increase in average math scores of 0.23 standard deviations and reading scores of 0.10 standard deviations two years after admission.[23]

[22] See Appendices E and F for additional evidence on the impacts of KIPP middle schools on student achievement, including sensitivity tests that examine whether the impact estimates change with different modeling assumptions and methods.

[23] Note that our matched-student analyses weight students equally across cohorts for a given school.

Table III.2. Middle school impacts on state tests in math and reading

Outcome (Z-scores)	Years after admission 1	2	3	4
Middle schools (lottery-based analysis)				
Math achievement	0.10*	0.24**	0.18**	na
	(0.05)	(0.06)	(0.07)	
Reading achievement	0.03	0.18**	0.14*	na
	(0.05)	(0.05)	(0.06)	
Number of schools	15	15	14	
Number of students	608	563	458	
Middle schools (matched-student analysis)				
Math achievement	0.06**	0.23**	0.29**	0.27**
	(0.01)	(0.01)	(0.01)	(0.02)
Reading achievement	0.00	0.10**	0.15**	0.16**
	(0.01)	(0.01)	(0.01)	(0.01)
Number of schools	37	37	31	30
Number of students	36,798	29,386	23,433	14,425
New middle schools (matched-student analysis)[a]				
Math achievement	0.04	0.23**	na	na
	(0.02)	(0.04)		
Reading achievement	0.04	0.12**	na	na
	(0.02)	(0.03)		
Number of schools	7	7		
Number of students	2,471	1,205		

Source: State and district administrative records data

Note: Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Lottery-based estimates measure the impact of an offer of admission to a KIPP middle school (the ITT estimate) and are based on regression models that pool all lottery schools and control for baseline covariates. The impact estimates from the matched-student design use a similar regression model. A matched-student impact estimate was calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. Standard errors are in parentheses. na=not available.

[a] Estimates report the impact of KIPP on students who enrolled in a KIPP middle school that was founded in 2011–2012 or later.

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

The newer KIPP middle schools in our matched-student analysis—those that opened during the i3 grant period (fall 2011 or later)—have positive impacts on math and reading achievement that are of a similar magnitude as those of the overall middle school impact estimates. In year 2, for example, impacts on math test scores are 0.23 standard deviations among both newer KIPP middle schools and in the overall sample (Table III.2).[24] The magnitude

[24] The estimated impact in the overall sample of the matched-student design reflects the average estimated impact of the KIPP middle schools included in the sample, with each school weighted equally. An alternative approach for measuring impacts would have given greater weight to KIPP middle schools established earlier and that served a greater number of KIPP students (in other words, an approach that would have weighted schools according to the number of cohorts or number of students from the school in our data). Under that weighted scheme, the estimated impacts of KIPP middle schools based on the matched-student design would have been greater than 0.23.

of the impacts in reading in year 2 are also similar across the two samples (0.12 for newer schools compared to 0.10 for all middle schools). These results provide evidence that the new KIPP middle schools in our data appear to be just as effective in increasing student achievement as the typical middle school in the study, even though the new schools are in their first years of operation.

Because science and social studies tests are not administered annually during middle school, we based estimates of the impacts of KIPP middle schools on science and social studies achievement on a single score in each subject—the highest middle school grade where science or social studies scores could be observed for more than one cohort of KIPP students in our sample. On average, KIPP middle schools have a positive and statistically significant impact of 0.25 standard deviations in both science and social studies, equivalent to moving the average student from the 48th percentile to the 58th percentile in science and from the 51st to the 61st percentile in social studies (Table III.3).

Table III.3. Middle school impacts on state tests in science and social studies

Outcome (Z-scores)	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
Science achievement	0.07	-0.18	0.25** (0.01)	32	18,433
Social studies achievement	0.10	-0.15	0.25** (0.02)	18	10,440

Source: State and district administrative records data.

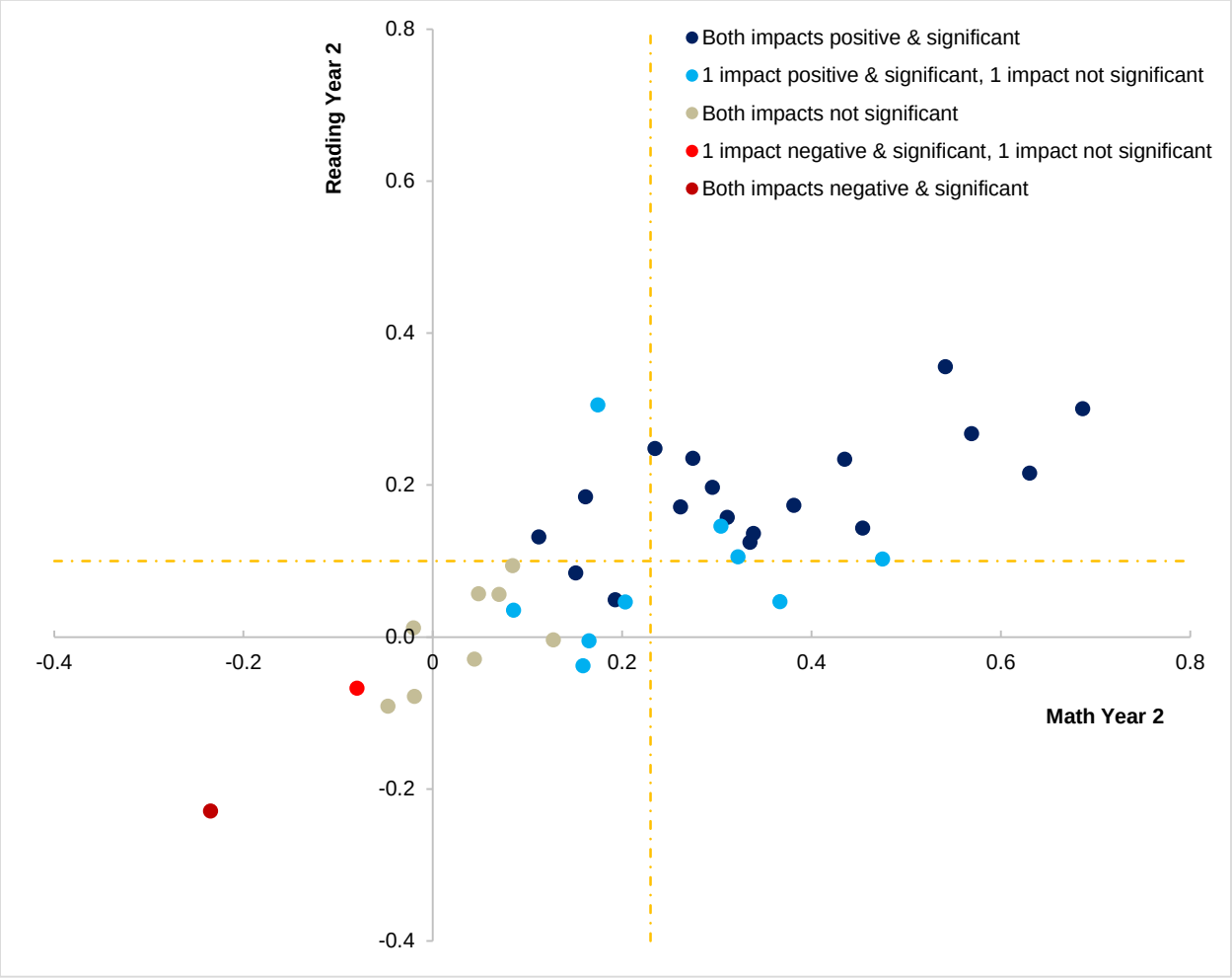
Notes: Impacts represent the cumulative effect of attending KIPP, not the marginal effect of an additional year of treatment. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. An impact estimate was calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. Standard errors are shown in parentheses. The grade level of middle school exams used for this analysis varied by jurisdiction. For each subject and site, we selected the highest middle school grade level where science or social studies was observed for more than one cohort of KIPP students.

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

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

We also examined the distribution of school-level impacts for our matched-student sample of 37 schools (Figure III.2). Each point plots one KIPP middle school's math and reading impact estimates two years after students enroll, based on all years covered by our data. Overall, 18 schools have positive and statistically significant impacts in both math and reading and nine schools have a positive and statistically significant impact in either math or reading (but not both). Eight schools have no statistically significant impacts, one school has a negative and statistically significant impact in only one subject, and one school has statistically significant negative impacts in both math and reading.

Figure III.2. Distribution of middle school reading and math impact estimates after two years



Notes: Impacts calculated separately for each middle school; each circle represents one middle school with two-year cumulative impacts plotted for reading (y-axis) and math (x-axis). Impact estimates are noted as statistically significant at the 0.05 level, two-tailed test. The dashed orange lines represent the average impact in math (vertical line) and reading (horizontal line).

C. Patterns of KIPP middle school impacts over time

Across the KIPP network, the average impacts of middle schools were positive and statistically significant throughout the 10-year period for which we have data, though higher in earlier years than recent years. To examine changes in the effectiveness of the network over time, we focused on trends at the middle school level, since the KIPP network has always included middle schools but added elementary and high schools only in recent years. Table III.4 shows estimates of the average impact of all operating KIPP middle schools where we have data *in each school year*, as measured by the schools' cumulative effect on students two years after initial KIPP enrollment. The number of KIPP middle schools included in the average impact increases from 8 schools in 2005 to 34 schools by 2014. KIPP middle schools have had positive and statistically significant impacts in both math and reading for all years from 2005 to 2014. KIPP-wide average

impacts were largest in 2007 and earlier (when no more than 17 schools were included in the sample), especially in math, ranging from 0.38 to 0.50 standard deviations, compared with 0.16 to 0.30 standard deviations between 2008 and 2014. In 2013 and 2014, when these two-year impacts fully reflect the performance of KIPP schools during the scale-up period and 30–34 schools are included in the sample, math impacts are 0.22 and 0.24 standard deviations. Differences over time in the average impacts of KIPP middle schools on reading achievement are less pronounced.

Table III.4. Impacts of KIPP middle schools on students two years after enrolling, by calendar year

Year	Math impact estimate	Reading impact estimate	Number of schools	Number of students	Average age of schools (years)
2005	0.38** (0.04)	0.11** (0.04)	8	848	2.9
2006	0.44** (0.03)	0.19** (0.03)	13	1,420	3.9
2007	0.43** (0.02)	0.21** (0.02)	17	1,908	4.1
2008	0.30** (0.02)	0.13** (0.02)	20	2,383	4.5
2009	0.28** (0.02)	0.17** (0.02)	23	2,974	4.9
2010	0.26** (0.02)	0.14** (0.02)	23	3,250	5.9
2011	0.26** (0.02)	0.09** (0.02)	20	2,933	6.0
2012	0.16** (0.02)	0.07** (0.02)	23	3,307	6.2
2013	0.22** (0.02)	0.09** (0.02)	30	4,298	6.9
2014	0.24** (0.01)	0.08** (0.01)	34	5,118	6.7

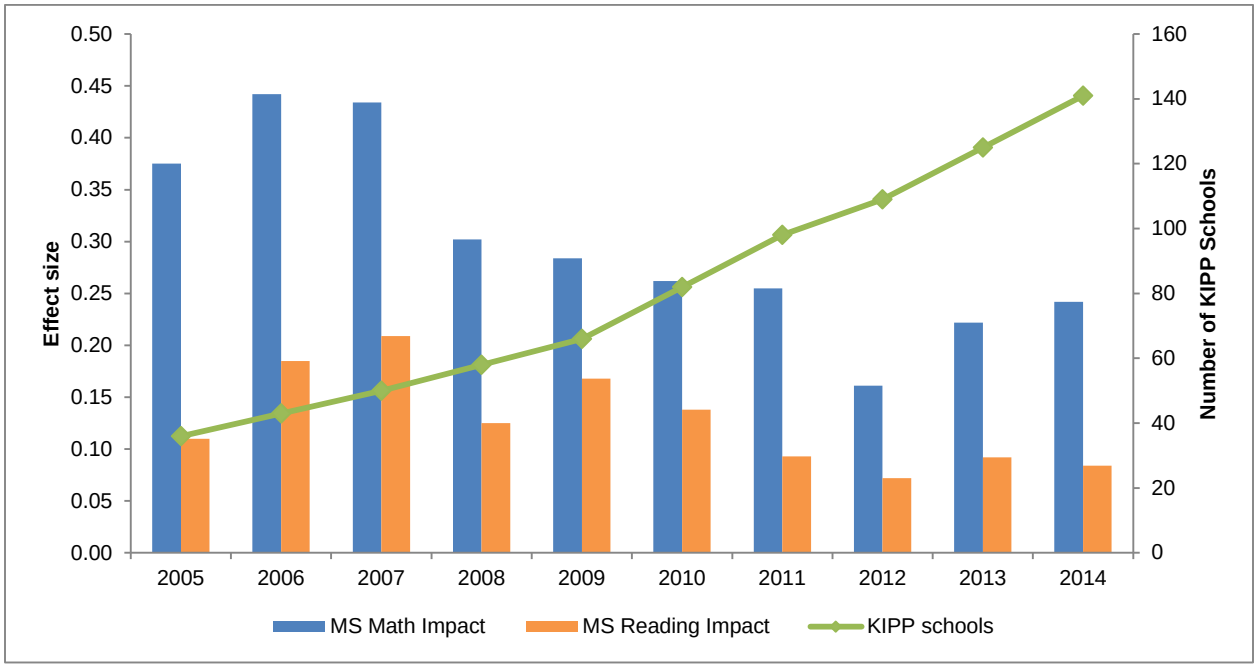
Notes: Impact estimates are estimated separately by school year; each table row includes an impact estimate based solely on the test scores of students affiliated with a KIPP middle school in that school year as compared to the scores of those students' matched comparison group in that school year. Impact estimates are calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. Standard errors are shown in parentheses. The year refers to the spring semester of the school year when the achievement exams were taken. Schools were omitted from the sample when fewer than 20 treatment students could be observed in a given year.

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

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

Several factors may explain these trends in impacts, including changes in the number and composition of schools in the sample, the relative performance of newer versus older schools, changes over time in the effectiveness of existing KIPP schools as the network has expanded, and changes in the effectiveness of schools attended by comparison students. Overall, KIPP's student achievement impacts moderated during a time of high growth in the network, although this moderation in impacts did not accelerate during the i3 scale-up period (Figure III.3). Although middle schools were part of the growth story, the largest expansion occurred in the number of KIPP elementary and high schools.

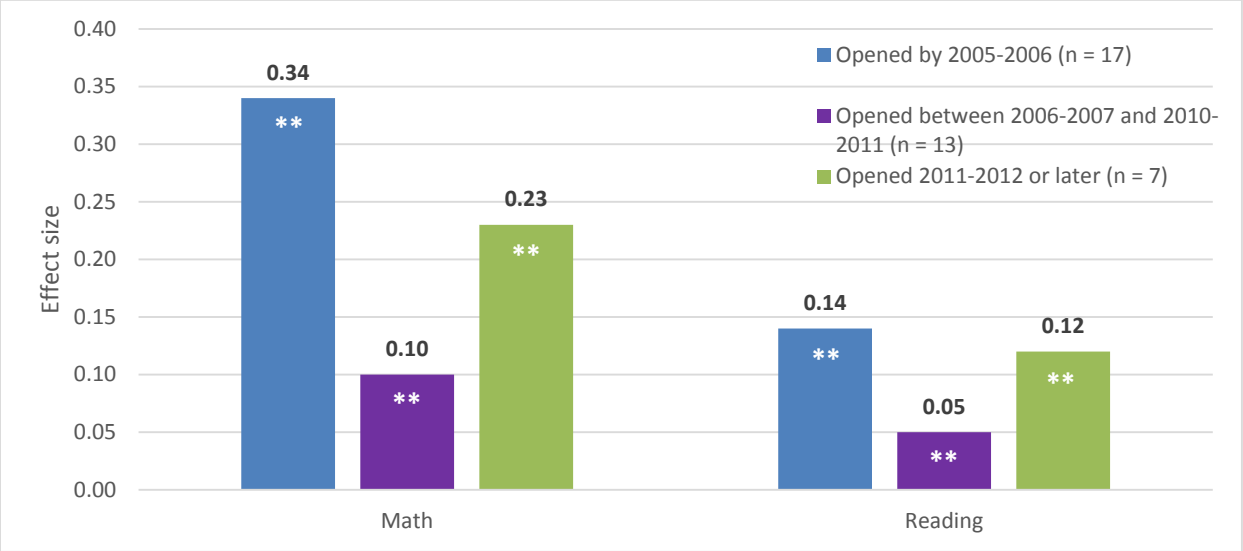
Figure III.3. Change in the size of the KIPP network and middle school impacts over time



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. Impact estimates are calculated separately for each KIPP school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. They are estimated separately by school year and plotted using the left side y-axis. All impacts are statistically significant at the 0.01 level. The year refers to the spring semester of the school year when the achievement exams were taken. The size of the KIPP network is plotted against the right-side y-axis.

To further explore the overall trends in the network-wide estimates of KIPP impacts over time, we examined the relative performance of KIPP middle schools that opened at different points in the network's history. Figure III.4 presents impacts of KIPP schools (using all years of data to measure each school's effectiveness) separately for groups of schools opened during different periods. Average impacts are highest for the group of schools that opened first—in the 2005–2006 school year and earlier, and lowest for the group of schools that began operating between 2006–2007 and 2010–2011, when the network started significant expansion into the elementary and high school levels (by adding more than one school at each level per year beginning in 2006–2007) as well as into new cities. Schools that opened during the scale-up period—fall 2011 or later—fall in between, but the magnitude of their impacts is closer to the earliest KIPP schools.

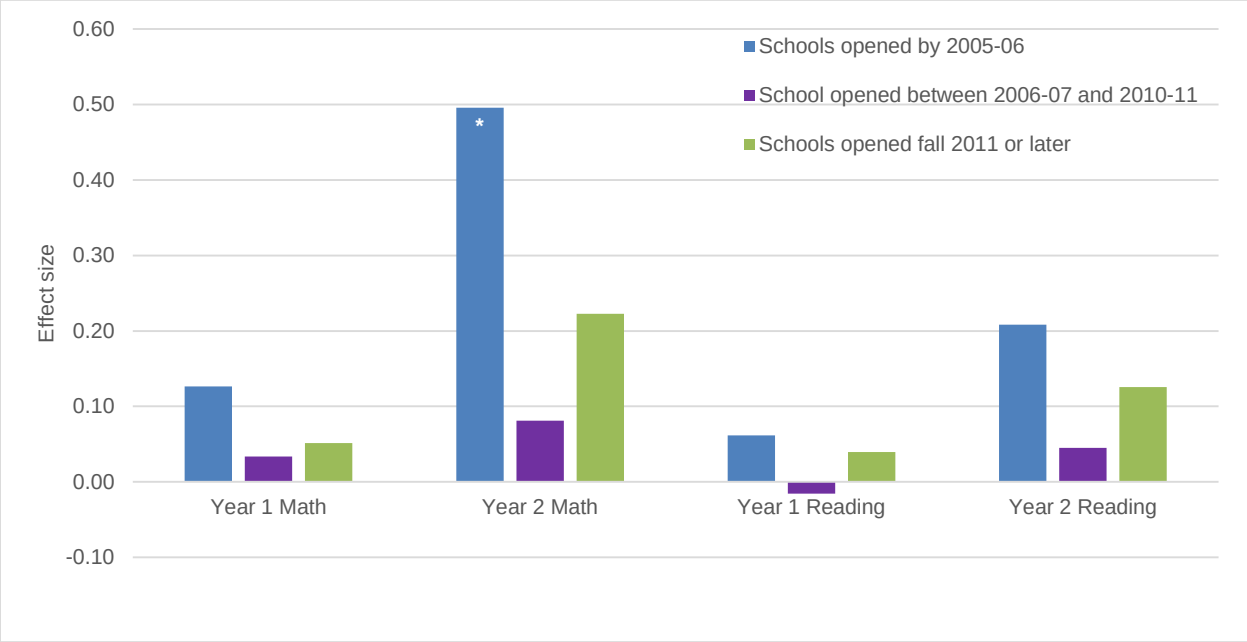
Figure III.4. Impacts of KIPP middle schools on students two years after enrolling, by year opened



Notes: Impact estimates are the cumulative two-year impact of KIPP on students who enrolled in any of the KIPP middle schools in the school records data provided to the study, based on the year the school opened. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

KIPP middle schools that opened in fall 2011 or later—during the scale-up period—performed about the same in their first two years of operation as KIPP middle schools that opened prior to 2011, on average. The comparison shown in Figure III.4 is based on all years of data available for a given school, but a fairer way to compare more recently opened schools with established KIPP middle schools is to estimate their impacts in their first years of operation. Figure III.5 shows impacts in the first two years of operation for the oldest KIPP schools (opened by fall 2005), more recent KIPP schools (opened between fall 2006 and fall 2010) and schools that opened in the last few years (in fall 2011 or later). This analysis sheds light on whether schools opening during the period of KIPP network expansion facilitated by the i3 scale-up grant were any more or less effective from the outset than schools opening during earlier periods of KIPP growth. KIPP middle schools opened during the scale-up period had higher average impacts in math and reading than schools that opened between fall 2006 and fall 2010, but the differences were not statistically significant. The earliest KIPP schools had larger average student achievement impacts during their first two years of operation than new KIPP middle schools. However, only the year 2 math impact estimate is a statistically significant difference.

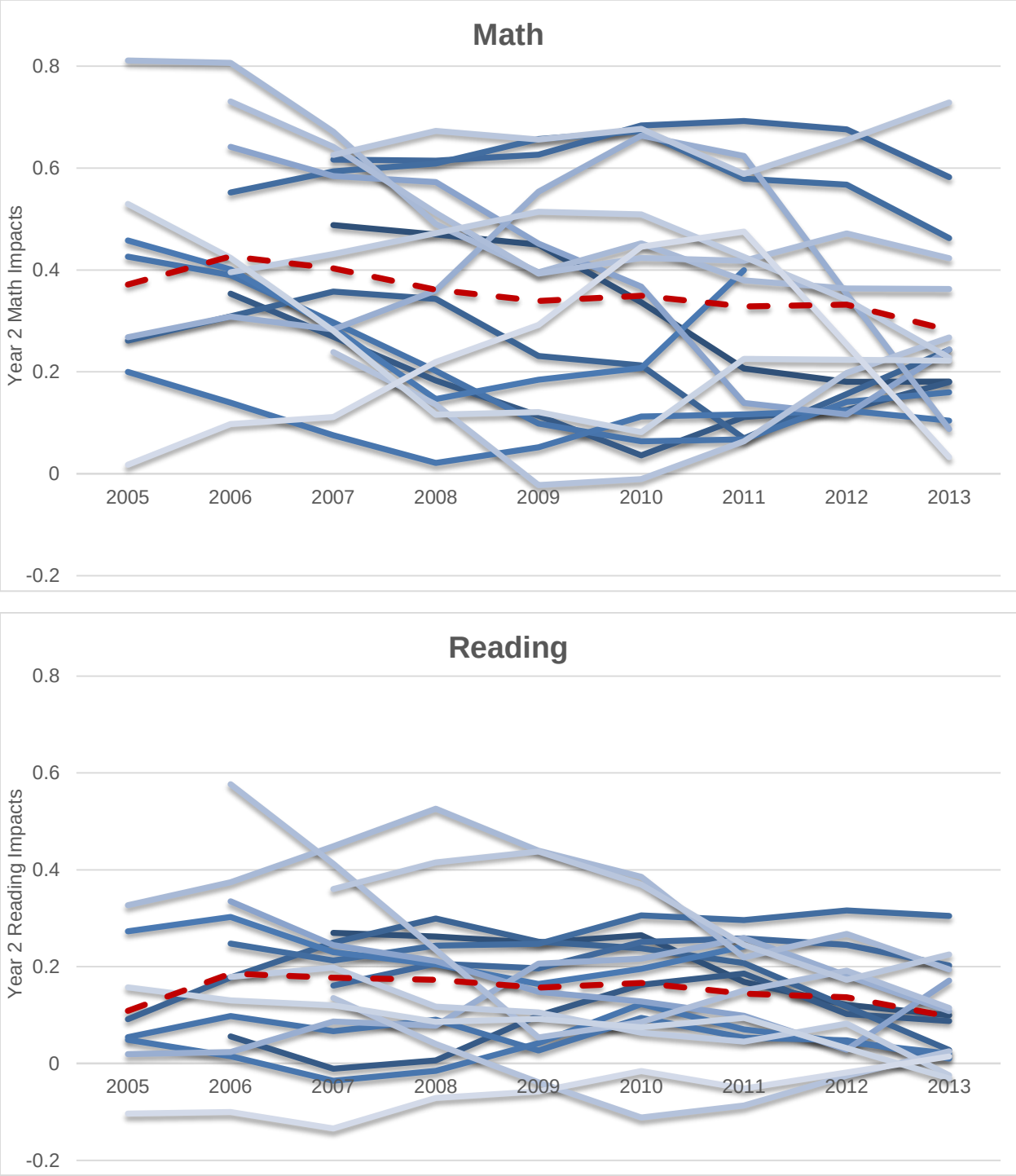
Figure III.5. Differences in achievement outcomes within first two years of operation



Notes: This figure reports the one-year and cumulative two-year impacts of KIPP on students who enrolled in a KIPP middle school during the first two years of that school's operation. Results are divided between schools opened by fall 2005, schools opened from fall 2006 through fall 2010, and schools opened since fall 2011 that are part of the study sample. Impacts are calculated by comparing the outcomes of these treatment students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. The impact estimates are estimated separately by school and school year. The average impact estimates reported here assign an equal weight to each of the school-level impact estimates. The single statistically significant difference in impacts with the post-2011 schools (in year 2 math) at the .05 level, as measured by a two-tailed t-test, is noted with an asterisk (*).

Finally, we examined the school-level trends in our sample to see how the performance of individual schools affects the average annual impacts described above, focusing on the set of 17 schools open by 2005–2006 that we observe for several years. Because the sample size for a given school in a given year is small, we "smoothed" the trends in these figures by averaging the estimate in a given year with the school's impact estimate in the prior year and subsequent year (Figure III.6). Each shaded blue line represents the smoothed impact trend for a particular KIPP middle school (lines were shaded differently to make it easier to follow an individual line over time; different shades of blue do not signify differences in impacts). The dotted red line represents the average of all the individual smoothed school-level estimates. As shown in these figures, the overall trend in the average impacts of the network is relatively stable over time. There is a slight decline in both reading and math impacts as the schools (and the network) age.

Figure III.6. Achievement impacts of KIPP middle schools over time, two years after enrollment, by school (schools opened by fall 2005)



Notes: This figure reports school-level estimates of the cumulative two-year impact of KIPP on math (top) and reading (bottom) test scores for schools in our sample that were opened by the 2005-2006 school year.. The impact estimates are estimated separately by school year for each school in the study sample. Each blue line represents a different KIPP middle school and the dotted red line represents the average of all the blue lines. The smoothed impact estimates average the estimate in a given year with the school's impact estimate in the prior year and subsequent year.

D. High school achievement

KIPP opened its first high school in 2004–2005; as of the 2014–2015 school year, the network included 22 high schools. Unlike the first KIPP middle and elementary schools that opened, KIPP high schools were designed from the outset to serve students continuing within the network, so that students completing grade 8 at a KIPP middle school would be able to remain within the network throughout their high school years. In addition to these continuing students, the typical KIPP high school also serves some students entering the network from non-KIPP middle schools.

As was the case with KIPP elementary schools, KIPP high schools could not simply adopt the educational approach used by KIPP middle schools and apply it to older students. The curriculum at the high school level is different, student issues differ, and college preparation activities are especially important. Moreover, schools face the challenge of serving the two groups of students described above, continuing students and new entrants, who differ in terms of their familiarity with the KIPP culture and approach and may differ in terms of their initial level of academic preparation.

We measure the impacts of KIPP high schools on student achievement using two approaches, both of which are quasi-experimental. To estimate impacts of these schools on continuing students, we compare two groups of KIPP middle school graduates—those with and those without an available KIPP high school to enter upon completing middle school. To estimate impacts for new entrants to the network, we use a matched-student design, much like the matched-student design used to estimate impacts of KIPP middle schools.

The estimated impacts of KIPP high schools may also be influenced by the counterfactual condition—the set of non-KIPP students included in the analysis to represent what would have happened to the treatment group of KIPP students had they not been able to attend a KIPP high school. In the absence of a KIPP high school, network middle schools try to connect graduates with high schools most likely to encourage college preparation activities, and many graduates choose options other than traditional public high schools. In the matched-school comparison group, for example, 25 percent of students attend traditional public high schools, 38 percent attend non-KIPP charter high schools, 14 percent attend magnet schools, and 14 percent attend private or boarding schools. By comparison, in regions with KIPP high schools in our sample, an average of 70 percent of KIPP middle school students (the treatment group in our matched-school design) enroll in a KIPP high school, 5 percent attend other charter high schools, 3 percent attend private or boarding schools, 5 percent attend magnet schools, and 13 percent attend traditional public high schools. In the new entrant analysis, since the treatment group of new entrants come from non-KIPP middle schools in the district, we define the comparison group to be other students at non-KIPP middle schools who remain at non-KIPP public schools in their high school years—either traditional, charter, or magnet.

We find that the impacts of KIPP high schools differ for different groups of students and for different schools:

For new entrants to the network, KIPP high schools have positive, statistically significant, and educationally meaningful impacts on achievement in math, ELA, and science. Attending a KIPP high school boosts new entrants' high school math scores by 0.27 standard deviation units, a statistically significant impact representing an increase from the 48th to the 59th percentile for the typical student (Table III.5). Impacts in ELA and science are 0.18 and 0.31 standard deviations, respectively, and are also significant. Relative to outcomes for the matched comparison group, these impacts are equivalent to an increase from the 47th to 54th percentile in ELA and from the 42nd to 54th percentile in science. The average impact in social studies (0.01) is close to zero and not statistically significant.

Table III.5. Impacts of KIPP high school on achievement for new entrants

	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
Mathematics achievement	0.22	-0.04	0.27** (0.04)	12	1,489
ELA achievement	0.11	-0.07	0.17** (0.04)	14	1,861
Science achievement	0.11	-0.20	0.31** (0.04)	12	1,383
Social studies achievement	-0.14	-0.15	0.01 (0.06)	8	643

Source: State and district administrative records data.

Notes: Impacts were calculated separately for each KIPP high school. In a given high school, the outcome may be either an end-of-course exam (e.g., algebra), or an end-of-grade exam (e.g., grade 10 mathematics). Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Standard errors are shown in parentheses.

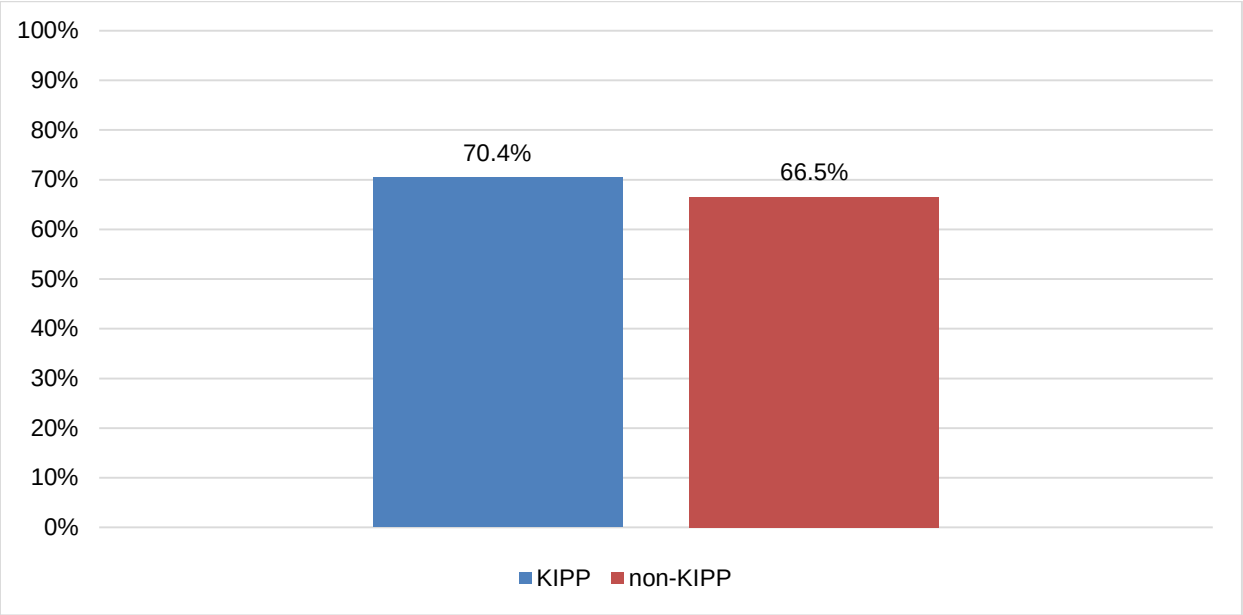
* Impact estimate is statistically significant at the 0.05 level, two-tailed test.

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

For new entrants to KIPP high schools, we also examine the probability of graduating within four years of entry. We find that this group of KIPP high schools did not significantly affect four-year graduation rates among new entrants (Figure III.7). The estimated impact on graduation rates is positive but small (4 percentage points) and statistically indistinguishable from zero.

There are two important limitations to the study's graduation analyses. First, the graduation indicator we use cannot distinguish between dropouts (who did not graduate) and students who left the data for some other reason such as transferring to private school or to a different school district (and who may or may not have graduated from high school). Both of these groups of students are classified as "non-graduates" in the analysis. Second, our propensity score matching approach relies on the assumption that the pre-KIPP characteristics of sample members observed in our data (middle school test scores and demographic attributes) fully capture attributes that are associated both with selection into KIPP and the outcomes of interest. In our graduation analyses, unlike our analyses of achievement, we have no baseline measure of the outcome being examined (graduation).

Figure III.7. Impact of KIPP high school on graduation among new entrants

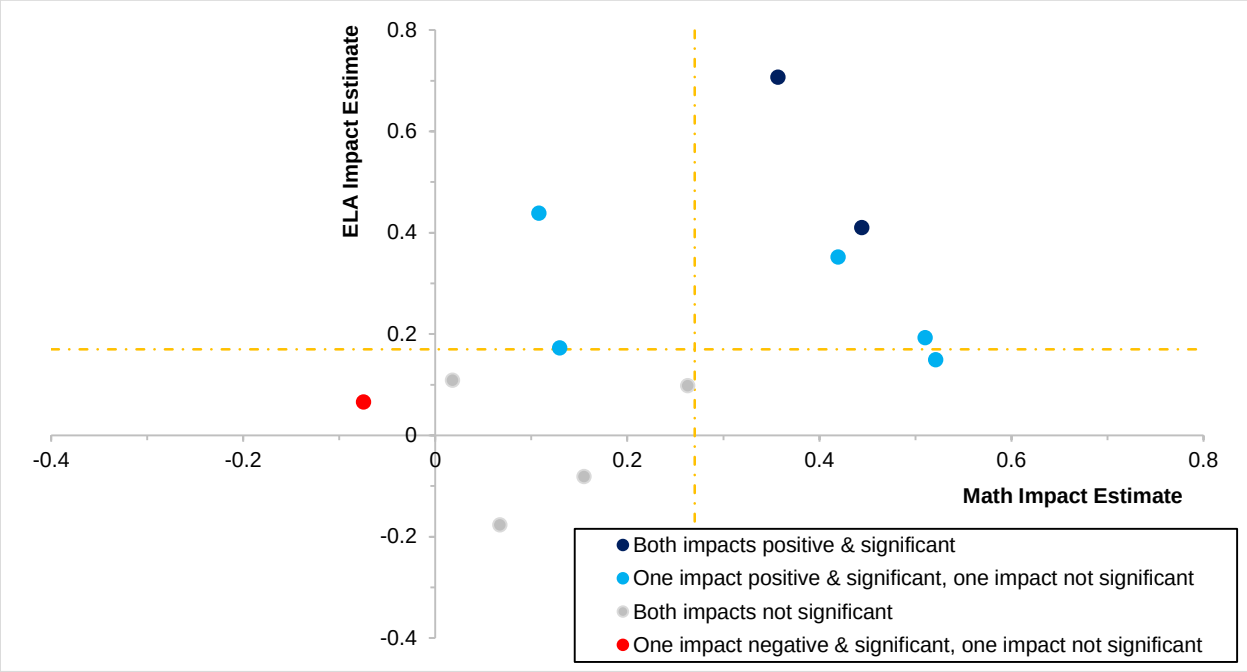


Notes: Impacts were calculated separately for the 7 KIPP high schools with available data for this outcome. Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Difference is not statistically significant at the 0.05 level, two-tailed test.

As is the case with both KIPP elementary and middle schools, the impact estimates for new entrants vary across the different KIPP high schools in our sample (Figure III.8). Of the 12 schools with both math and ELA impact estimates, two high schools have positive and statistically significant impacts on new entrants in both math and ELA, five schools have a positive and significant impact on new entrants in one subject but not the other, four schools have no significant impacts on new entrants in either subject, and one school has a negative and significant impact on new entrants in one subject.[25]

[25]The sample size for each school varies greatly, since different KIPP high schools have different numbers of cohorts and different numbers of students entering KIPP for the first time in grade 9 in our data. Thus, in some schools, impacts are estimated relatively imprecisely.

Figure III.8. Distribution of KIPP high school impacts on new entrants in ELA and math



Note: Each circle represents the math and ELA impact estimate for one KIPP high school. Dark blue circles indicate that impacts in both subjects are statistically significant and positive at the 0.05 level, two-tailed test. Light blue circles indicate that the impact in only one of the two test subjects is statistically significant and positive. Grey indicates that both impacts are statistically indistinguishable from zero. Light red circles indicate that the impact in only one of the two test subjects is statistically significant and negative. No school has a statistically significant and negative impact in both subjects. The dashed orange lines represent the average impact in math (vertical line) and ELA (horizontal line).

For students continuing from KIPP middle schools, the achievement impacts of KIPP high schools are not statistically significant on average, but these impacts vary by school. For continuing students, we measure KIPP high schools' impacts on student achievement in reading, language, and math using their scores on the TerraNova standardized assessment. These impacts are positive but small (none is larger than 0.07 standard deviation units) and not statistically significant (Table III.6).

Table III.6. Impacts of KIPP high schools on achievement for continuing students

	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
TerraNova Reading achievement	0.21	0.16	0.05 (0.05)	8	933
TerraNova Language achievement	0.04	0.01	0.03 (0.06)	8	933
TerraNova Math achievement	0.01	-0.06	0.07 (0.06)	8	933

Source: Study-administered test.

Notes: Impacts were calculated separately for each KIPP high school. Marginal impacts of opportunity to attend a KIPP high school are ITT estimates, measured on the TerraNova assessment. Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Standard errors are shown in parentheses.

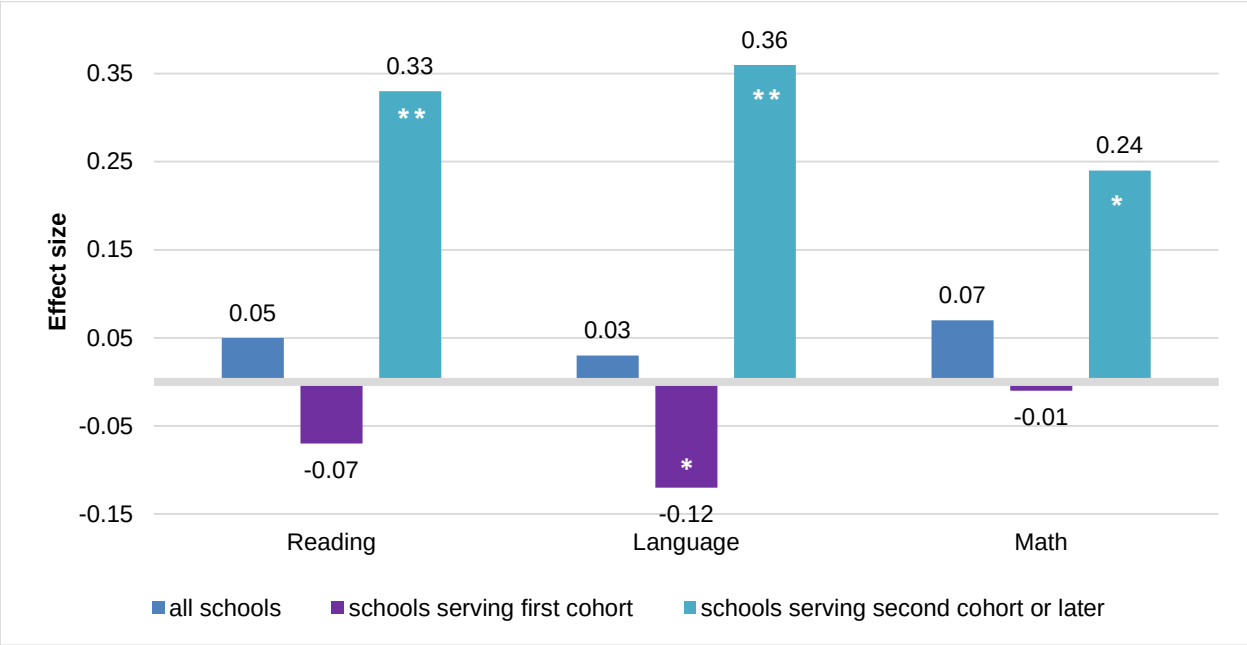
* Impact estimate is statistically significant at the 0.05 level, two-tailed test.

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

Five of the eight KIPP high schools included in this analysis were brand new and serving their first cohort of students when we measured their impacts. It turns out that there are significant differences between the estimated impacts of these new KIPP high schools versus those with more experience that are serving a later cohort of students. For the set of five new high schools where we observe only the first cohort of students, impacts are negative in all three subjects, and statistically significant in language (Figure III.9). For the three high schools where we observe only a later cohort, by contrast, impacts are positive and statistically significant in all three subjects, with magnitudes ranging from 0.24 to 0.36 standard deviation units. These more positive impacts for the more experienced high schools could imply that KIPP high schools become more effective as they gain experience. We examined this possibility in the high schools in the analysis of new entrants, but did not find clear evidence that impacts improved after the first cohort of students: impacts were somewhat lower for the first cohort than for later cohorts, but the difference was not statistically significant. The more positive impacts could result from some other characteristic of the more experienced KIPP high schools that distinguishes them from the new schools. Because we do not observe impacts of high schools in multiple years under this design, we cannot distinguish among these hypotheses.

Continuing students with the option to attend a KIPP high school are less likely to drop out of high school. The overall dropout rate in the entire sample is very low (only 15 students reported dropping out), but is significantly lower for the treatment group—1 percent for those who had the chance to attend a KIPP high school and 3 percent for those who did not. The magnitude of this impact (two percentage points) is similar to the magnitude of the impact of KIPP high schools on the likelihood of graduating from high school after four years among new entrants (four percentage points), but is measured more precisely.

Figure III.9. High school impacts for continuing students, by cohort examined



Notes: Model: Matched-school design. Outcome: TerraNova test. Sample size: eight schools; 933 students. Statistically significant at the 0.05 level (*) or 0.01 level (**), two-tailed test.

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

As we have shown, the matched-student analysis finds that KIPP high schools have positive impacts for new entrants to the network, whereas the matched-school analysis finds that, on average, impacts are not statistically significant for continuing students. There are four potential explanations for these differences, related to differences in the KIPP high schools included in the analyses, differences in the test-score measures, differences in the students, and differences in the experiences of the comparison groups.

First, the new entrant analysis includes impacts for a different set of high schools than the matched-school analysis (see Appendix A). Differences between the two sets of impact estimates could simply reflect differences between the effectiveness and age of various KIPP schools included in the analyses at the time outcomes were measured at each school. Although we did not find clear evidence that KIPP high schools improve with experience, data are not sufficient to fully assess whether high school age or other characteristics of the high schools included in the analyses might be driving the results.

Second, the assessments measuring student achievement differ in the two analyses. The type of learning measured by the TerraNova assessment (the outcome for continuing students in the matched-school analysis) may differ from that measured by standardized end-of-course subject tests (the outcome for new entrants in the matched-student analysis). In addition, we administered the TerraNova assessment to students three years after entering high school (usually in grade 11), whereas the timing of the state exams varied by jurisdiction and exam but were administered as early as grade 9 in some cases.

Third, the fact that two analyses are measuring impacts for different sets of students—new entrants to KIPP versus continuing students—may explain these differences. KIPP high schools may have more positive impacts for students with the kinds of characteristics typical of new entrants than for students with characteristics typical of continuing students. As discussed in Chapter II, new entrants have much lower baseline (grade 8) test scores than continuing students. If KIPP high schools are delivering a model that is more effective for students with lower prior achievement, this could help to explain why the impact estimates for new grade 9 entrants are more substantial than the impacts for students who already had extensive exposure to KIPP in middle school. A related explanation is that continuing students may have already benefited substantially from their time in a KIPP middle school, so attending KIPP during high school represents a smaller change for them than it does for new entrants.[26]

Fourth, estimated impacts may differ due to differences in the experiences of the non-KIPP comparison students in the two analyses. In the matched-student analysis, new entrants to KIPP are being compared to students who attended non-KIPP middle schools in the district (including non-KIPP charter schools) and remained in non-KIPP schools in their high school years. In the matched-school analysis, continuing students (those with an opportunity to attend a KIPP high school) are compared with those who did not have that opportunity, the majority of whom (at least 66 percent) attended private or boarding schools, magnet schools, and non-KIPP charter schools in the district. If those schools tend to be more effective than the high schools attended by comparison students in the matched-student analysis, then the KIPP schools would show a smaller impact even if they were equally effective relative to similar comparison schools.

[26] Another question is whether the new entrants to KIPP high schools might benefit from peer effects derived from attending high school alongside the higher-scoring students who arrived from a KIPP middle school. Available evidence from the literature suggests that peer effects are less pronounced during high school than in earlier grades, with a standard deviation increase in achievement of peers (measured against the student-level distribution of test scores) associated with an increase in a student's score of 0.01 to 0.06 standard deviations, depending on the sample and methods used (Lavy et al. 2012; Burke and Sass 2013). In our sample, new entrants are surrounded by peers with middle school scores 0.42 and 0.56 standard deviations higher than their own scores in reading and math, respectively. Based on peer effects reported in the literature, this could produce benefits for new entrants of about 0.00 to 0.03 standard deviations in reading and math. Our impact estimates for new entrants are much larger than this, suggesting that peer effects are unlikely to play a large role in explaining impacts for new entrants.

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

IV. KIPP'S IMPACTS ON STUDENT BEHAVIOR AND ATTITUDES

A student's long-term success depends on more than test scores, so we are interested in whether KIPP schools have effects on student outcomes other than academic achievement. We use data from surveys of students and their parents to investigate how attending a KIPP school influences student behavior, motivation and engagement, school experiences and satisfaction, educational goals and aspirations, and—for high school students—college preparation activities.

Findings on behavioral measures in this chapter should be interpreted with caution. Several outcomes may be affected by reference bias or by the fact that different groups of students or parents may have a consistently different frame of reference when answering survey questions about their attitudes, behavior, and experiences.[27] As a result, when comparing KIPP and non-KIPP students, differences in some outcomes may arise from differences between the groups in their frames of reference (related, for example, to differences in peer groups in KIPP schools and other schools) rather than differences in actual attitudes, behavior, and experiences. Reference bias should be less problematic for objective measures, such as courses high school students report having taken or other specific college preparation activities.

In many cases, the surveys included multiple items capturing the same underlying construct, so we created indices that summarize responses on related data items. We used principal components analysis to identify which group of related items to include in each index. For example, the index indicating the extent to which a student is well adjusted to school represents a parent's average response on seven items related to his or her child's adjustment to various aspects of school, such as getting along with others, liking school, working hard at school, and respecting adults. Details on all indices are presented in Appendix B.

We rely on the same designs for measuring KIPP schools' impacts on student behavior and attitudes as for measuring impacts on achievement outcomes, with one caveat. We did not have the opportunity to administer surveys to students in the matched-student analyses (or to their parents) at the middle school and high school levels. Thus, impacts on these outcomes are based on the lottery-based, randomized design at the elementary and middle school levels and on the matched-school design at the high school level. At the middle school level, we administered surveys to students and their parents in spring of the second follow-up year. Given the age of the students in the other samples, we administered surveys to parents only at the elementary school level (in spring of the second follow-up year) and to students only at the high school level (in spring of the fourth follow-up year).

As with the achievement impacts, the results presented here represent the effect of being offered admission to a KIPP school (that is, winning the lottery) at the elementary and middle school levels, and the effect of initial enrollment in a KIPP high school. Elementary and middle school students are included in the treatment group regardless of whether they actually enrolled in a KIPP school; high school students who initially enrolled at a KIPP school are included in the

[27] For further study on reference bias among educational surveys, see West et al. (2014). In addition, Duckworth and Yeager (2015) provide a review of the literature on limitations on measuring noncognitive outcomes for purposes of evaluating educational interventions.

treatment group even if they subsequently transferred out. These are therefore conservative estimates of KIPP's impact on students who remain enrolled.

Table IV.1 provides an overview of the results, based on separate analyses for KIPP elementary, middle, and high schools for a large number of student behavior and attitude outcomes. In each cell of the table, we report the number of outcomes examined for that level of KIPP school, as well as the number of those impact results that were statistically significant (at the 5 percent level). Cells shaded dark green represent domain-by-grade combinations where at least three outcomes (or all outcomes in cases with fewer than three outcomes examined) are statistically significant, suggesting a positive effect overall. Cells shaded light green include one or two statistically significant positive effects. More detailed tables showing these impacts are presented in Appendix B.

Table IV.1. Summary of significant effects, by domain

Domain	Elementary	Middle	High
Motivation and engagement Time spent on homework; indices of engagement, self-control, academic confidence, grit, and effort in school	1 of 5	2 of 10	0 of 8
Experiences and satisfaction Parent and student feelings about school, perceptions of classmates and teachers, school disciplinary environment, academic difficulty, parent outreach, parent engagement	3 of 9	6 of 14	0 of 8
Behavior Peer pressure, undesirable behaviors, positive behaviors, illegal activities, adjustment to school, parental concerns	0 of 1	0 of 11	--
Goals/aspirations Expectations for on-time high school graduation, college attendance, and college completion	1 of 3	0 of 7	0 of 3
College prep activities Discussions about college, assistance in planning for college, college testing, application to college	--	--	7 of 16
Coursetaking Years of coursework, number of AP/IB/honors courses	--	--	8 of 15

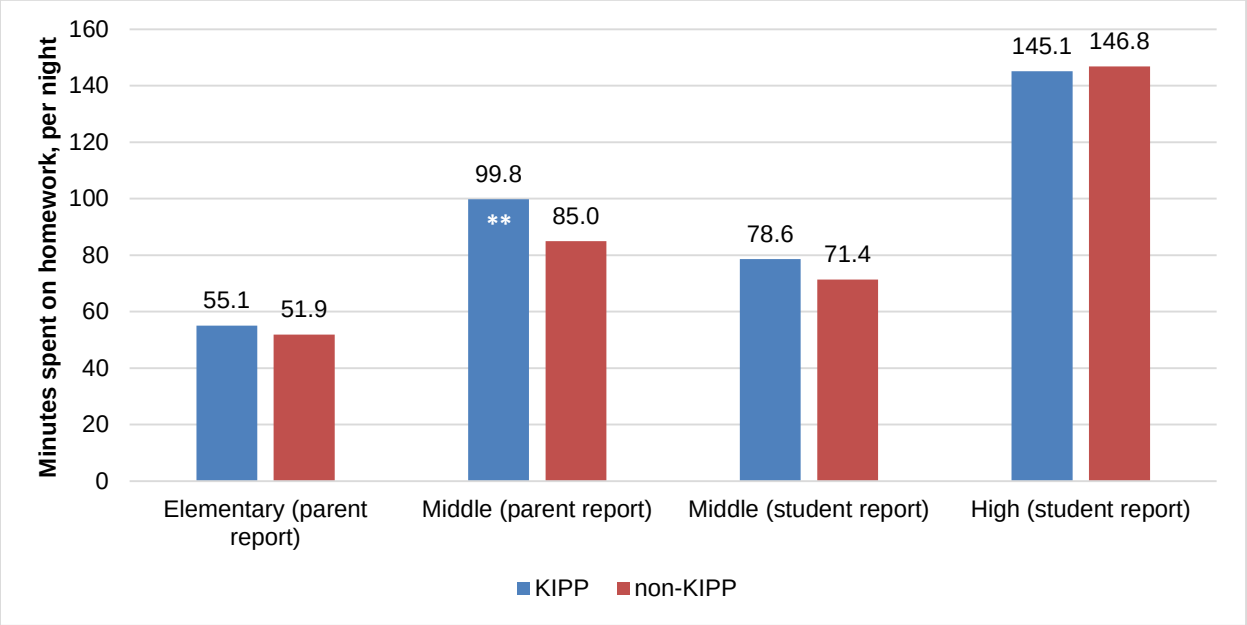
Notes: Each cell indicates the number of statistically significant outcomes from the total number of outcomes measured in a given domain and grade level. Cells shaded dark green represent domain-by-grade combinations where at least three outcomes (or all outcomes in cases with fewer than three outcomes examined) are statistically significant, suggesting a positive effect overall. Cells shaded light green include one or two statistically significant positive effects. Specific results, by outcome, are included in Appendix B.

Across grade levels, KIPP schools have no statistically significant impact on most measures of student motivation and engagement, behavior, or educational aspirations. By and large, we find no statistically significant impact of KIPP elementary, middle, or high schools on most measures included in three of the six categories of non-achievement outcomes we examined—student motivation and engagement, behavior, and goals and aspirations. In the motivation and engagement category, KIPP does not significantly affect measures of student

self-control, academic motivation, academic confidence, grit, school engagement (positive or negative), or effort in school, including student reports of the time spent on homework.

There are a few exceptions to the lack of impacts on outcomes in the motivation and engagement category. At the elementary school level, parents of KIPP students report a higher frequency of positive comments from the student about school and looking forward to going to school. At the middle school level, KIPP has a positive and significant impact on an index of a student's academic collaboration (how much a student enjoys working with, and learns while helping out, fellow students). The estimated impact of KIPP is statistically significant for only one other outcome measure in this category, the amount of time middle-school students spend on homework, according to their parents (Figure IV.2). Parents of treatment group students in middle school report that their children spend an average of almost 15 more minutes on homework each night compared with parents of control group students. This positive impact estimate is not supported by other evidence on homework completion. According to student reports, the impact of KIPP middle schools on homework time is about half as large (about seven minutes) and not statistically significant. Further, parents of KIPP middle school students do not report that their children are any more likely to complete their homework. Note that this result is different from estimates from our earlier study of KIPP middle schools (Tuttle et al. 2013), in which the impacts on time spent on homework were positive and significant according to both the student and parent reports. Further, in that study, the number of minutes spent was higher for both treatment and control groups (118 and 96 minutes, respectively, according to parent reports). At the elementary and high school levels, the estimated impact of KIPP on students' homework time is not statistically significant, regardless of the specific measure examined.

Figure IV.2. KIPP impacts on time spent on homework, by grade level



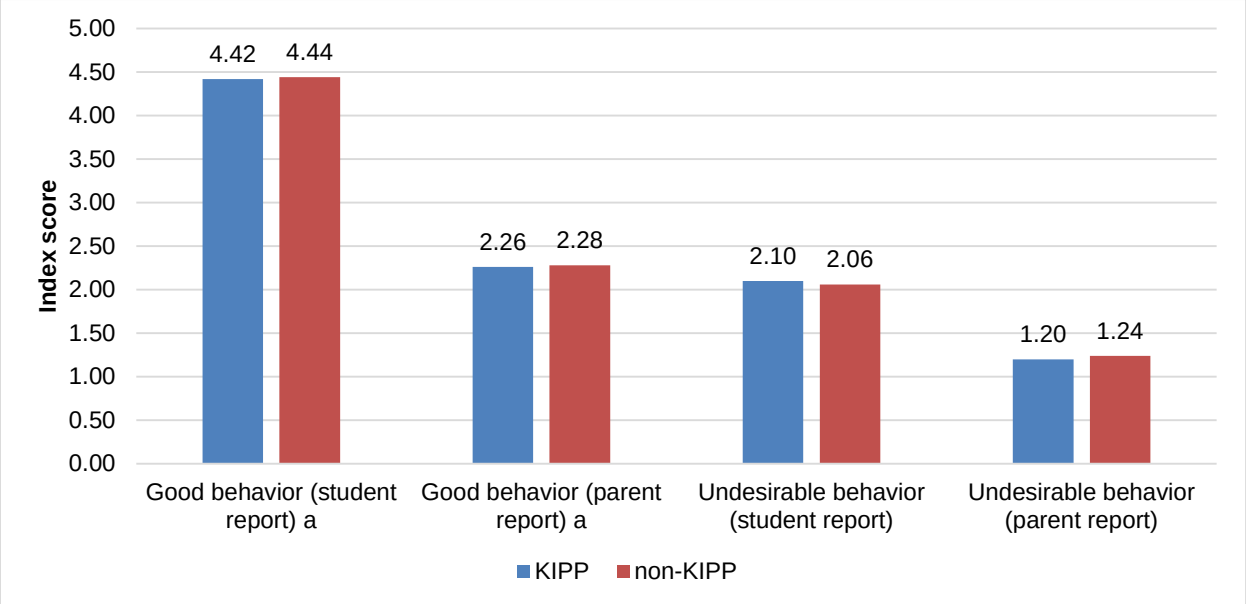
*Significantly different from zero at the .05 level, two-tailed test.

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

Similarly, we find no evidence that attending a KIPP school affects student behavior. At the elementary school level, we measured behavior using an index of items indicating the extent to which the student is well adjusted to school, and find that KIPP does not significantly affect this outcome. We used a range of measures of behavior at the middle school level, including indices of positive behaviors, undesirable behaviors, peer pressure, illegal activities, parental concerns about their child, frequency of school disciplinary actions (according to the parent), and the extent to which the child is well adjusted. In none of these cases is the impact of KIPP statistically significant. For example, Figure IV.3 shows estimates of the mean value of the indices of positive and undesirable behaviors for KIPP (treatment group) and non-KIPP (control group) middle school students, from both parent and student reports. The mean values of each of these student behavior outcomes for the KIPP and non-KIPP groups are nearly identical. This finding is in contrast to estimates from Tuttle et al. (2013), in which we found that KIPP schools led to increases in two indicators of student-reported undesirable behavior.

We measured educational goals and aspirations using responses from parents and students. In general, the educational goals and aspirations among these elementary, middle, and high school students are high in both the treatment (KIPP) and comparison (non-KIPP) groups. At the high school level, for example, 84 percent of students report that they think they will graduate from college. For 12 of 13 outcomes, the estimated impact of KIPP is not statistically significant. The single exception is among parents of students at KIPP elementary schools, who are 10 percentage points more likely than the comparison group to believe their child is very likely to complete college (81 versus 71 percent).

Figure IV.3. KIPP middle school impacts on behavior

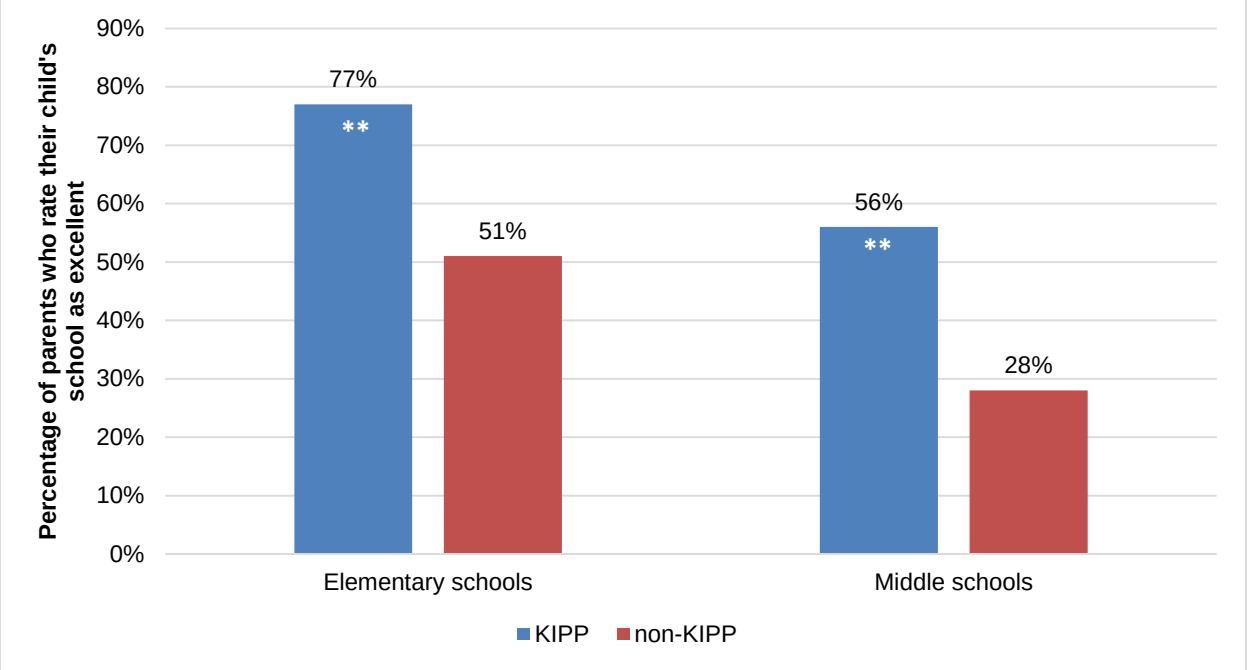


Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information. Indices based on student reports are measured on a scale of 1 (once or twice per year) to 5 (almost every day) and indices based on parent reports are measures on a scale of 1 (never) to 3 (often). None of the differences shown in the figure is statistically significant.

[a] Index has an alpha smaller than 0.7, indicating low reliability

KIPP elementary and middle schools have positive impacts on school satisfaction, particularly among parents. At both the elementary and middle school levels, being offered admission to KIPP leads to increases in parents' satisfaction with their child's school. At the elementary school level, for example, more than three-quarters of parents in the treatment group rate their child's school as excellent, compared to about half of parents in the control group (Figure IV.4). At the middle school level, 56 percent of treatment group parents and 28 percent of control group parents rate the school as excellent. These findings are consistent with previous research on KIPP in particular and oversubscribed charter middle schools in general (Tuttle et al. 2013; Gleason et al. 2010). Similarly, KIPP has significant positive impacts on a parent-based index capturing satisfaction with school facilities, academics, safety, and discipline. KIPP also has significant positive impacts on several other satisfaction measures, including indices of school efforts to engage parents at both the elementary and middle school level; an index of parental perceptions of problems in their child's middle school; and middle school students' perceptions of their schoolmates (see Appendix B). Evidence of KIPP impacts on satisfaction do not extend to the high school level, however, as none of eight measures of student-reported satisfaction at the high school level is statistically significant.

Figure IV.4. KIPP impacts on parental satisfaction



Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

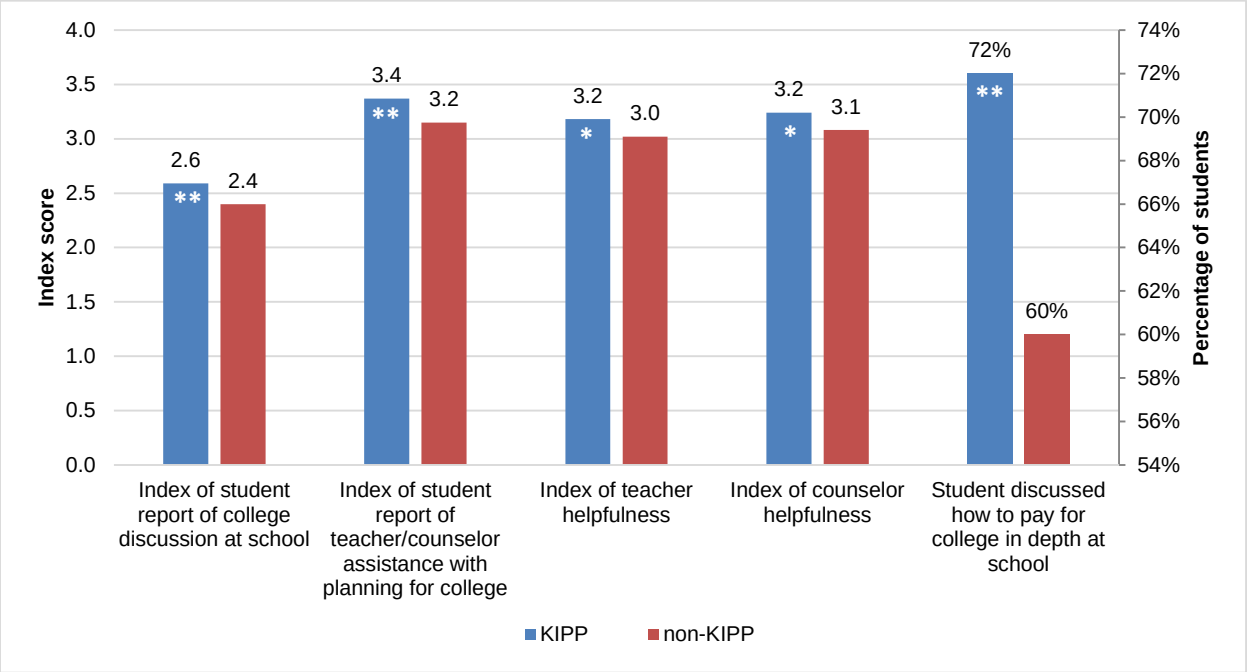
*Significantly different from zero at the .05 level, two-tailed test.

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

KIPP high schools have positive effects on several aspects of college preparation, including discussions about college, applying to college, and coursetaking. Based on surveys administered in spring of the fourth follow-up year—senior year for most students—KIPP high schools have positive and significant impacts on three measures related to school assistance in

planning for college. First, KIPP has a positive impact on the frequency of discussions about college at school, based on an index of student reports. This outcome includes indicators of whether school staff discussed with students nine different topics related to college admissions and readiness (Figure IV.5). Second, KIPP leads to students being more likely have in-depth discussions at school about how to pay for college, with 72 percent of treatment students reporting having these discussions compared to only 60 percent of comparison students. Third, KIPP high schools have a positive impact on an index summarizing student reports of teacher or counselor assistance with planning for college. This finding is supported by estimates of positive KIPP impacts on student reports of the helpfulness and encouragement of their school's teachers and counselors on various aspects related to college preparation (including help in selecting courses that meet graduation requirements and are needed for college admission, help in finding a suitable college and deciding what to do after graduation, encouragement to take AP/honors courses, and encouragement to continue their education through college and beyond).

Figure IV.5. KIPP high school impacts on assistance for college planning



Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

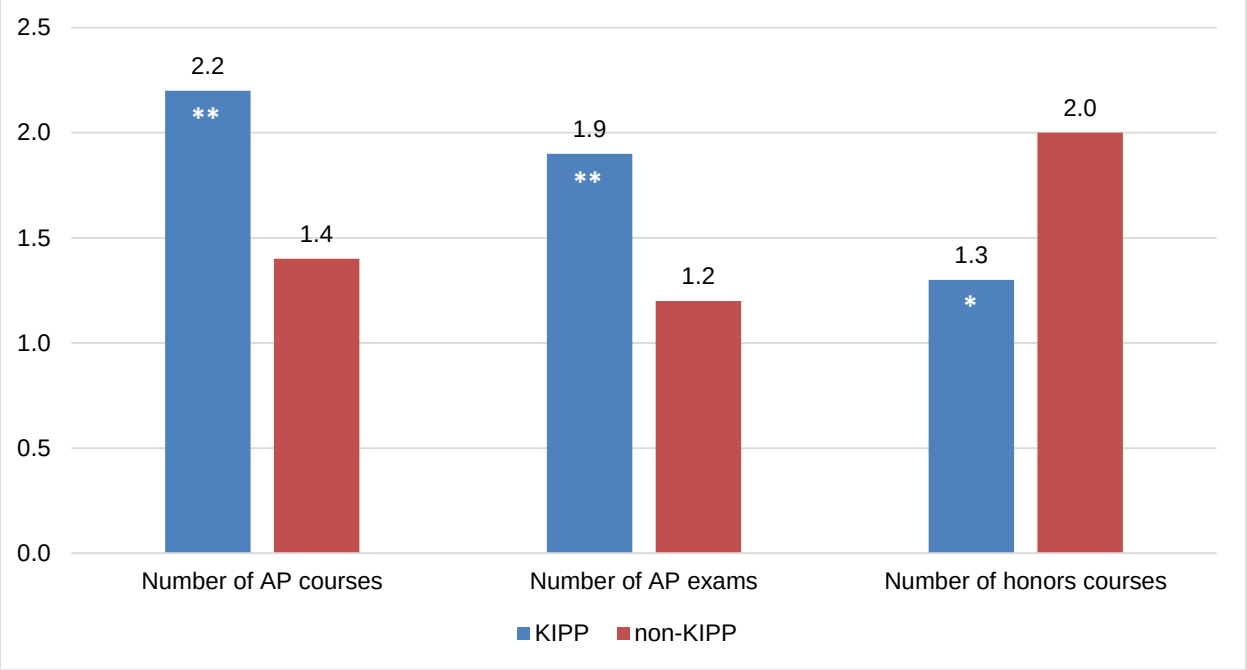
*Significantly different from zero at the .05 level, two-tailed test.

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

In addition to assistance provided by the school, we also measured specific activities students may have engaged in to better prepare themselves for college. KIPP high schools have a positive and significant effect on a summary index of seven different college preparation activities a student may have undertaken.[28] They also have a positive and significant effect on whether the student applied to at least one college or university by spring of senior year—93 percent of treatment students did so compared with 88 percent of comparison students.

A key aspect of students' college preparation involves which courses they take while in high school. Students with the opportunity to attend a KIPP high school enroll in more AP courses and, correspondingly, have taken or intend to take more AP exams (Figure IV.6). This difference is mostly offset by a negative effect of KIPP on the number of honors courses students have taken or plan to take, suggesting that students at KIPP high schools may be substituting AP courses for honors courses. KIPP high schools also have positive and statistically significant effects on the number of years of coursework students report having taken in several subjects, with treatment students averaging 1.4 more high school courses overall than comparison students, including more music, science, foreign language, and history.

Figure IV.6. KIPP high school impacts on advanced coursetaking



Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

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

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

[28] Impacts are positive and significant for three of the seven items included in the index: whether students visited in-state college campuses, whether they visited out-of-state college campuses, and whether they took practice ACT/SAT exams.

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

REFERENCES

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. "Inputs and Impacts in Charter Schools: KIPP Lynn." *American Economic Review: Papers & Proceedings*, vol. 100, May 2010, pp. 1-5.

Burke, Mary A. and Tim Sass. "Classroom Peer Effects and Student Achievement," *Journal of Labor Economics*, vol. 31, no. 1, 2013, pp. 51-82.

Duckworth, Angela L. and David Scott Yeager. "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes." *Educational Researcher*, vol. 44, no. 4, 2015.

Fortson, Kenneth, Philip Gleason, Emma Kopa, and Natalya Verbitsky-Savitz. "Horseshoes, Hand Grenades, and Treatment Effects? Reassessing Whether Nonexperimental Estimators are Biased." *Economics of Education Review*, vol. 44, February 2015, pp. 100-113.

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

Gill, Brian, Joshua Ferguson, Hanley Chiang, Bing-Ru Teh, Joshua Haimson, and Natalya Verbitsky-Savitz. "Replicating Experimental Impact Estimates in the Context of Control-Group Noncompliance." *Statistics and Public Policy.* Forthcoming.

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

Gleason, Philip M., Christina Clark Tuttle, Brian Gill, Ira Nichols-Barrer, and Bing-ru Teh. "Do KIPP Schools Boost Student Achievement?" *Education Finance and Policy*, vol. 9, no. 1, 2014, pp. 36-58.

Hill, Carolyn, Howard Bloom, Alison Rebeck Black, and Mark Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*, vol. 2, no. 3, December 2008, pp. 172–177.

KIPP Foundation. "KIPP: 2014 Report Card." 2015. Available at http://www.kipp.org/reportcard. Accessed August 1, 2015.

Lake, Robin, Melissa Bowen, Allison Demeritt, Moira McCullough, Joshua Haimson, and Brian Gill. "Learning from Charter School Management Organizations: Strategies for Student Behavior and Teacher Coaching." Seattle, WA: Center on Reinventing Public Education and Washington DC: Mathematica Policy Research, March 2012.

Lavy, Victor and M. Daniele Paserman. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *The Economic Journal*, vol. 122, 2012, pp. 208–237

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

Nichols-Barrer, Ira, Philip Gleason, Brian Gill, and Christina Clark Tuttle. "Student Selection, Attrition, and Replacement in KIPP Middle Schools." *Educational Evaluation and Policy Analysis*, published online before print, January 2015.

Tuttle Christina Clark, Brian Gill, Phillip Gleason, Virginia Knechtel, Ira Nichols-Barrer, and Alex Resch. "KIPP Middle Schools: Impacts on Achievement and Other Outcomes." Washington, DC: Mathematica Policy Research, February 27, 2013

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

West, Martin R., Matthew A. Kraft, Amy S. Finn, et al. "Promise and Paradox: Measuring Students' Non-cognitive Skills and the Impact of Schooling." Center for Education Policy Research, Harvard University. February 2014.

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

APPENDIX A

LIST OF KIPP SCHOOLS IN NETWORK

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

Table A.1. KIPP elementary schools (2014-2015)

School	Region	Year opened	Lottery-based Sample	Ineligible	Non-study
KIPP Academy Boston Elementary	KIPP Massachusetts	2014		X	
KIPP Academy Elementary School	KIPP NYC	2009	X		
KIPP AMP Elementary School	KIPP NYC	2013		X	
KIPP Ascend Primary School	KIPP Chicago	2010			X
KIPP Austin Comunidad	KIPP Austin	2010			X
KIPP Austin Connections Elementary	KIPP Austin	2011			X
KIPP Austin Leadership Elementary	KIPP Austin	2013		X	
KIPP Austin Obras	KIPP Austin	2013		X	
KIPP Believe Primary	KIPP New Orleans	2011			X
KIPP Central City Primary	KIPP New Orleans	2008			X
KIPP Columbus Elementary	KIPP Columbus	2014		X	
KIPP Comienza Community Prep	KIPP LA	2010			X
KIPP CONNECT Houston Primary School	KIPP Houston	2014		X	
KIPP DC: Arts & Technology Academy	KIPP DC	2014		X	
KIPP DC: Connect Academy	KIPP DC	2013		X	
KIPP DC: Discover Academy	KIPP DC	2009			X
KIPP DC: Grow Academy	KIPP DC	2010			X
KIPP DC: Heights Academy	KIPP DC	2011			X
KIPP DC: Lead Academy	KIPP DC	2012		X	
KIPP DC: LEAP Academy	KIPP DC	2007	X		
KIPP DC: Promise Academy	KIPP DC	2009			X
KIPP DC: Quest Academy	KIPP DC	2014		X	
KIPP DC: Spring Academy	KIPP DC	2013		X	
KIPP Delta Elem. Literacy Academy	KIPP Delta	2009			X
KIPP Destiny Elementary	KIPP Dallas-Fort Worth	2013		X	
KIPP Dream Prep	KIPP Houston	2006			X
KIPP East Community Primary	KIPP New Orleans	2014		X	
KIPP Empower Academy	KIPP LA	2010			X
KIPP Esperanza Dual Language Academy	KIPP San Antonio	2014		X	
KIPP Explore Academy	KIPP Houston	2009			X
KIPP GCP Primary	KIPP Eastern North Carolina	2012		X	
KIPP Harmony Academy	KIPP Baltimore	2009			X
KIPP Iluminar Academy	KIPP LA	2013		X	

Table A.1 *(continued)*

School	Region	Year opened	Lottery-based sample	Ineligible	Non-study
KIPP Indy Unite Elementary	KIPP Indianapolis	2014		X	
KIPP Infinity Elementary School	KIPP NYC	2010	X		
KIPP Legacy Preparatory School	KIPP Houston	2010			X
KIPP McDonogh 15 Primary	KIPP New Orleans	2006			X
KIPP Memphis Academy Elementary	KIPP Memphis	2013		X	
KIPP Memphis Collegiate Elementary	KIPP Memphis	2012		X	
KIPP New Orleans Leadership Primary	KIPP New Orleans	2011			X
KIPP PEACE Elementary School	KIPP Houston	2011			X
KIPP Philadelphia Elementary Academy	KIPP Philadelphia	2010	X		
KIPP Raíces Academy	KIPP LA	2008	X		
KIPP SHARP College Prep Lower School	KIPP Houston	2008	X		
KIPP SHINE Prep	KIPP Houston	2004	X		
KIPP STAR Harlem College Prep Elementary	KIPP NYC	2014		X	
KIPP STAR Washington Heights Elementary	KIPP NYC	2011			X
KIPP STRIVE Primary	KIPP Metro Atlanta	2012		X	
KIPP Un Mundo Dual Language Academy	KIPP San Antonio	2012		X	
KIPP Victory Academy	KIPP St. Louis	2014		X	
KIPP Vida Preparatory Academy	KIPP LA	2014		X	
KIPP Vision Primary	KIPP Metro Atlanta	2013		X	
KIPP VOICE Elementary	KIPP Jacksonville	2012		X	
KIPP WAYS Primary	KIPP Metro Atlanta	2014		X	
KIPP ZENITH Academy	KIPP Houston	2009			X
Life Academy at Bragaw, a KIPP School	KIPP New Jersey	2014		X	
Revolution Primary, a KIPP School	KIPP New Jersey	2014		X	
Seek Academy, a KIPP School	KIPP New Jersey	2013		X	
SPARK Academy, a KIPP School	KIPP New Jersey	2009	X		
THRIVE Academy, a KIPP School	KIPP New Jersey	2012		X	
	TOTAL		**8**	**31**	**21**

Note: Because the lotteries for inclusion in the elementary school study sample were conducted for entry in the 2011-2012 school year, schools were ineligible for the lottery-based sample if they opened in fall 2012 or later. Non-study schools are those elementary schools that were open as of 2011-2012, but were insufficiently oversubscribed to support the study design at an entry grade for that school year.

Table A.2. KIPP middle schools (2014-2015)

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP 3D Academy	KIPP Houston	2001		X			
KIPP Academy Boston Middle School	KIPP Massachusetts	2012			X		
KIPP Academy Lynn Middle School	KIPP Massachusetts	2004	X	X			
KIPP Academy (Houston)	KIPP Houston	1995		X			
KIPP Academy Middle School (New York)	KIPP NYC	1995	X	X			
KIPP Academy Nashville	KIPP Nashville	2005					X
KIPP Academy of Innovation	KIPP LA	2014				X	
KIPP Academy of Opportunity	KIPP LA	2003					X
KIPP Adelante Preparatory Academy	KIPP San Diego	2003					X
KIPP AMP Middle School	KIPP NYC	2005		X			
KIPP Ascend Middle School	KIPP Chicago	2003					X
KIPP Aspire Academy	KIPP San Antonio	2003		X			
KIPP Austin Academy of Arts & Letters	KIPP Austin	2009	X	X			
KIPP Austin Beacon Prep	KIPP Austin	2012			X		
KIPP Austin College Prep	KIPP Austin	2002	X	X			
KIPP Austin Vista Middle School	KIPP Austin	2012			X		
KIPP Bayview Academy	KIPP Bay Area	2003					X
KIPP Believe College Prep	KIPP New Orleans	2006					X
KIPP Bloom College Prep	KIPP Chicago	2013				X	
KIPP Blytheville College Prep School	KIPP Delta	2010		X			
KIPP Bridge Charter School	KIPP Bay Area	2002					X
KIPP Camino Academy	KIPP San Antonio	2010		X			
KIPP Central City Academy	KIPP New Orleans	2007					X
KIPP Charlotte	KIPP Charlotte	2007		X			
KIPP Columbus Middle	KIPP Columbus	2008					X

Table A.2 *(continued)*

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP CONNECT Houston Middle School	KIPP Houston	2014				X	
KIPP Courage College Prep	KIPP Houston	2012			X		
KIPP Create College Prep	KIPP Chicago	2012					X
KIPP DC: AIM Academy	KIPP DC	2005		X			
KIPP DC: KEY Academy	KIPP DC	2001		X			
KIPP DC: Northeast Academy	KIPP DC	2014				X	
KIPP DC: WILL Academy	KIPP DC	2006		X			
KIPP Delta College Preparatory School	KIPP Delta	2002		X			
KIPP Endeavor Academy	KIPP Kansas City	2007					X
KIPP Gaston College Preparatory	KIPP Eastern North Carolina	2001	X	X			
KIPP Halifax College Preparatory	KIPP Eastern North Carolina	2014				X	
KIPP Heartwood Academy	KIPP Bay Area	2004					X
KIPP Heritage Academy	KIPP Bay Area	2014				X	
KIPP Impact Middle School	KIPP Jacksonville	2010					X
KIPP Indy College Prep Middle	KIPP Indianapolis	2004					X
KIPP Infinity Middle School	KIPP NYC	2005	X	X			
KIPP Inspire Academy	KIPP St. Louis	2009					X
KIPP Intrepid Preparatory School	KIPP Houston	2008		X			
KIPP LA College Preparatory School	KIPP LA	2003	X				
KIPP Liberation College Prep	KIPP Houston	2006		X			
KIPP McDonogh 15 Middle	KIPP New Orleans	2006					X
KIPP Memphis Academy Middle	KIPP Memphis	2012			X		
KIPP Memphis Collegiate Middle	KIPP Memphis	2002		X			
KIPP Memphis Preparatory Middle	KIPP Memphis	2013				X	
KIPP Memphis University Middle	KIPP Memphis	2014				X	

Table A.2 *(continued)*

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP Montbello College Prep	KIPP Colorado	2011			X		
KIPP Nashville College Prep	KIPP Nashville	2013				X	
KIPP New Orleans Leadership Academy	KIPP New Orleans	2010					X
KIPP North Star Academy	KIPP Minnesota	2008					X
KIPP Philadelphia Charter School	KIPP Philadelphia	2003	X				
KIPP Philosophers Academy	KIPP LA	2012					X
KIPP Polaris Academy for Boys	KIPP Houston	2007		X			
KIPP Prize Preparatory Academy	KIPP Bay Area	2014				X	
KIPP Reach College Preparatory	— none —	2002					X
KIPP San Francisco Bay Academy	KIPP Bay Area	2003					X
KIPP Scholar Academy	KIPP LA	2012					X
KIPP Sharpstown College Prep	KIPP Houston	2007	X	X			
KIPP Sol Academy	KIPP LA	2013				X	
KIPP South Fulton Academy	KIPP Metro Atlanta	2003	X				
KIPP Spirit College Prep	KIPP Houston	2006		X			
KIPP STAR Harlem Middle School	KIPP NYC	2003	X	X			
KIPP STRIVE Academy	KIPP Metro Atlanta	2009		X			
KIPP Summit Academy	KIPP Bay Area	2003	X				
KIPP Sunshine Peak Academy	KIPP Colorado	2002		X			
KIPP TECH VALLEY	KIPP Albany	2005					X
KIPP TRUTH Academy	KIPP Dallas-Fort Worth	2003	X	X			
KIPP Tulsa College Preparatory	KIPP Tulsa	2005					X
KIPP Ujima Village Academy	KIPP Baltimore	2002	X				
KIPP Vision Academy	KIPP Metro Atlanta	2010		X			

Table A.2 *(continued)*

School	Region	Year opened	Lottery-based sample	Matched-student sample (est.)	Matched-student sample (new)	In-eligible	Non-study
KIPP Voyage Academy for Girls	KIPP Houston	2009		X			
KIPP Washington Heights Middle School	KIPP NYC	2012			X		
KIPP WAYS Academy	KIPP Metro Atlanta	2003	X	X			
KIPP West Philadelphia Preparatory	KIPP Philadelphia	2009	X				
Rise Academy, a KIPP School	KIPP New Jersey	2006					X
TEAM Academy, a KIPP School	KIPP New Jersey	2002					X
	TOTAL		**16**	**30**	**7**	**11**	**26**

Note: Schools in the matched-student sample that were open in 2010-2011 are considered "established;' those that opened in fall 2011 or later (during the scale-up period) are considered "new." Because a minimum of two years of available data were required for inclusion in any of the middle school study samples, schools that opened in fall 2013 or later are ineligible. Non-study schools are those located in jurisdictions (states or districts) that did not provide data for the matched-student analysis.

Table A.3. KIPP High Schools (2014-2015)

School	Region	Year opened	Matched-student sample	Matched-school sample	In-eligible	Non-study
KIPP Academy Lynn Collegiate High School	KIPP Massachusetts	2011	X			
KIPP Atlanta Collegiate	KIPP Metro Atlanta	2011	X			
KIPP Austin Collegiate	KIPP Austin	2008	X	X		
KIPP Blytheville Collegiate High School	KIPP Delta	2014			X	
KIPP DC: College Preparatory	KIPP DC	2009	X	X		
KIPP Delta Collegiate High School	KIPP Delta	2006	X			
KIPP Denver Collegiate High School	KIPP Colorado	2009	X			
KIPP DuBois Collegiate Academy	KIPP Philadelphia	2010		X		
KIPP Generations Collegiate	KIPP Houston	2011	X			
KIPP Houston High School	KIPP Houston	2004	X			
KIPP King Collegiate High School	KIPP Bay Area	2007		X		
KIPP Memphis Collegiate High	KIPP Memphis	2011	X			
KIPP Nashville Collegiate High School	KIPP Nashville	2014			X	
KIPP Northeast College Preparatory	KIPP Houston	2013	X			
KIPP NYC College Prep High School	KIPP NYC	2009	X	X		
KIPP Pride High School	KIPP Eastern North Carolina	2005	X			
KIPP Renaissance High School	KIPP New Orleans	2010				X
KIPP San Francisco College Preparatory	KIPP Bay Area	2013				X
KIPP San Jose Collegiate	KIPP Bay Area	2008		X		
KIPP Sunnyside High School	KIPP Houston	2010	X			
KIPP University Prep High School	KIPP San Antonio	2009	X	X		
Newark Collegiate Academy, a KIPP School	KIPP New Jersey	2007		X		
	TOTAL		**14**	**8**	**2**	**2**

Note: Because a minimum of one year of available data was required for inclusion in any of the high school study samples, schools that opened in fall 2014 or later are ineligible. Non-study schools are those located in jurisdictions (states or districts) that did not provide data for the matched-student analysis.

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

APPENDIX B

DETAIL ON SURVEY OUTCOMES

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

This appendix presents detailed information about the study's analysis of the impact of KIPP schools on student behavior and attitudes. First, we present information on response rates and the baseline equivalence of the analytic samples for the survey-based outcomes. Next, we define our methodology for constructing survey indices. We then present complete survey impact estimates for elementary, middle, and high schools. We conclude the appendix with tables listing the components of each index.

Detail on sample

Our survey outcome impact estimates come from four different data sources: an elementary school parent survey that includes eight KIPP elementary schools; a middle school parent survey and a middle school student survey, both of which include 16 KIPP middle schools; and a high school student survey that includes eight KIPP high schools. The elementary and middle school surveys were conducted using a randomized controlled trial design (see Appendix D for more information on the elementary school analytic methods and Appendix E for the middle school analytic methods). Table B.1 summarizes response rates by treatment and control group and overall for the RCT designs.

Table B.1. Rates of available outcome data by analytic sample (RCT)

Analytic Sample	Treatment	Control	Overall	# of schools
Elementary School Parent Survey	74%	59%	65%	8
Middle School Student Survey	61%	46%	54%	16
Middle School Parent Survey	75%	61%	68%	16

Note: Sample sizes for individual outcomes vary. The elementary and middle school surveys were administered during the spring of 2013.

The high school student survey was conducted using a quasi-experimental design detailed in Appendix H. The analytic sample for survey outcomes (which combines two models) included 865 students overall, 504 in the treatment group and 361 in the control group. Two schools are included in both models. As a result of this analytical approach, some outcome sample sizes are greater than 865. The survey was administered during the spring of the students' fourth year after completing 8th grade at a KIPP middle school (12th grade for most students) in either 2012, 2013, or 2014.

To check that our four outcome samples are equivalent on observable characteristics, we examined baseline equivalence separately for each. In particular, we compared mean values and proportions of baseline characteristics in the treatment and comparison groups among those with valid outcome data, separately for each survey instrument. For the elementary school parent survey, we examined 21 baseline characteristics including gender, age, race, language spoken, household income, mother's education, and Internet and computer access, and the number of books in the household (Table B.2). We found only one statistically significant difference, which is what we would have expected due to chance alone. For the middle school parent and student surveys, we examined 28 baseline characteristics including baseline and pre-baseline test scores in reading and math, gender, age, race, language spoken, special education status, household income, free and reduced price lunch status, mother's education, computer and Internet access,

parental help with homework, and parental discussions on college (Tables B.3—B.4). There were only two statistically significant differences for the middle school student survey and one statistically significant difference for the parent survey. Both samples were balanced on key test score and demographic variables.

Given these findings, we are confident that the admissions lotteries were conducted correctly and that the treatment and control groups in each analysis sample are similar in terms of their background characteristics, motivation, and prior educational experiences, aside from the outcome of the lottery itself. We address any potential differences by controlling for baseline characteristics in the impact models.

Table B.2. Elementary school parent survey baseline equivalence

Baseline characteristic	Mean lottery winner	Mean lottery loser	Difference	p-value	N_t	N_c	SD_t	SD_c
Female	0.451	0.489	-0.039	0.373	319	329	0.498	0.501
Age of respondent in years	4.12	4.14	-0.012	0.708	304	310	1.148	0.959
White, non-Hispanic	0.010	0.008	0.002	0.777	344	361	0.093	0.091
Hispanic (any race)	0.395	0.407	-0.013	0.649	344	361	0.501	0.492
Black, non-Hispanic	0.550	0.540	0.010	0.724	344	361	0.500	0.499
Other race	0.045	0.044	0.000	0.990	344	361	0.198	0.206
English: main language at home	0.612	0.589	0.023	0.460	306	314	0.500	0.493
Another language is main language at home	0.223	0.197	0.025	0.393	306	314	0.447	0.399
English and another language spoken equally at home	0.166	0.213	-0.048	0.129	306	314	0.395	0.410
One adult in household	0.255	0.307	-0.052	0.149	347	362	0.425	0.462
Family income less than 15K	0.221	0.272	-0.051	0.160	291	305	0.415	0.446
Family income between 15K and 25K	0.236	0.203	0.033	0.359	291	305	0.436	0.403
Family income between 25K and 35K	0.207	0.200	0.007	0.877	291	305	0.422	0.401
Family income between 35K and 50K	0.181	0.184	-0.003	0.931	291	305	0.378	0.388
Family income 50K or greater	0.155	0.141	0.014	0.613	291	305	0.330	0.349
Mother's education: less than HS	0.083	0.171	-0.088**	0.002	303	310	0.316	0.377
Mother's education: HS or GED	0.255	0.242	0.013	0.727	303	310	0.432	0.429
Mother's education: some college	0.327	0.277	0.050	0.193	303	310	0.461	0.448
Mother's education: college	0.334	0.310	0.025	0.569	303	310	0.473	0.463
Student has access to computer with internet at home	0.774	0.792	-0.017	0.610	303	312	0.309	0.331
Number of children's books at home	37.192	40.418	-3.226	0.353	292	292	0.484	0.490

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

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

Table B.3. Middle school student survey baseline equivalence

Baseline characteristic	Mean lottery winner	Mean lottery loser	Difference	p-value	N_t	N_c	SD_t	SD_c
Baseline reading exam (z-score)	-0.150	-0.165	0.015	0.879	169	126	0.888	0.811
Baseline math exam (z-score)	0.009	-0.139	0.147	0.163	169	126	0.931	0.845
Pre-baseline reading exam (z-score)	-0.156	-0.255	0.100	0.304	158	118	0.755	0.815
Pre-baseline math exam (z-score)	-0.130	-0.129	-0.001	0.992	159	118	0.902	0.882
Female	0.494	0.511	-0.017	0.740	270	188	0.501	0.501
Age of respondent in years	10.49	10.546	-0.061	0.271	244	172	0.779	0.668
White, non-Hispanic	0.012	0.020	-0.008	0.620	281	198	0.156	0.141
Hispanic or Latino	0.549	0.581	-0.032	0.340	281	198	0.501	0.495
Black, non-Hispanic	0.368	0.374	-0.006	0.848	281	198	0.486	0.485
Other race	0.071	0.025	0.045*	0.024	281	198	0.295	0.157
English: main language at home	0.476	0.460	0.017	0.633	281	198	0.501	0.500
Another language is main language at home	0.259	0.318	-0.059	0.139	281	198	0.420	0.467
English and another lang. spoken equally at home	0.265	0.222	0.042	0.271	281	198	0.441	0.417
One adult in household	0.292	0.253	0.039	0.355	279	198	0.453	0.436
Free/reduced price lunch status	0.818	0.864	-0.046	0.176	266	184	0.416	0.344
Special education status	0.059	0.110	-0.051	0.103	259	181	0.279	0.314
Family income less than 15K	0.213	0.201	0.012	0.773	240	164	0.416	0.402
Family income between 15K and 25K	0.277	0.250	0.027	0.543	240	164	0.418	0.434
Family income between 25K and 35K	0.157	0.250	-0.093*	0.034	240	164	0.377	0.434
Family income between 35K and 50K	0.185	0.165	0.020	0.643	240	164	0.391	0.372
Family income 50K or greater	0.168	0.134	0.034	0.358	240	164	0.398	0.342
Mother's education: less than HS	0.194	0.242	-0.049	0.215	280	198	0.378	0.430
Mother's education: HS or GED	0.249	0.263	-0.014	0.735	280	198	0.434	0.441
Mother's education: some college	0.248	0.232	0.016	0.703	280	198	0.434	0.423
Mother's education: college	0.309	0.263	0.047	0.257	280	198	0.471	0.441
Student has access to computer with internet at home	0.808	0.801	0.007	0.862	250	171	0.371	0.400
Parent helps student with homework at least 5 days per week	0.663	0.726	-0.064	0.177	248	168	0.455	0.447
Parent discussed college with student at least twice during pre-baseline school year	0.860	0.851	0.009	0.811	248	168	0.361	0.357

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

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

Table B.4. Middle school parent survey baseline equivalence

Baseline characteristic	Mean lottery winner	Mean lottery loser	Difference	p-value	N_t	N_c	SD_t	SD_c
Baseline reading exam (z-score)	-0.167	-0.182	0.015	0.864	201	173	0.929	0.821
Baseline math exam (z-score)	-0.050	-0.180	0.131	0.154	201	173	0.927	0.857
Pre-baseline reading exam (z-score)	-0.161	-0.260	0.099	0.257	192	163	0.780	0.798
Pre-baseline math exam (z-score)	-0.134	-0.172	0.038	0.695	193	164	0.899	0.876
Female	0.507	0.496	0.011	0.797	330	248	0.501	0.501
Age of respondent in years	10.53	10.565	-0.035	0.491	291	215	0.805	0.665
White, non-Hispanic	0.019	0.031	-0.012	0.434	343	262	0.160	0.172
Hispanic or Latino	0.520	0.553	-0.034	0.285	343	262	0.500	0.498
Black, non-Hispanic	0.387	0.366	0.021	0.450	343	262	0.492	0.483
Other race	0.074	0.050	0.025	0.222	343	262	0.299	0.218
English: main language at home	0.498	0.488	0.010	0.763	342	260	0.499	0.501
Another language is main language at home	0.249	0.296	-0.047	0.175	342	260	0.410	0.457
English and another lang. spoken equally at home	0.253	0.215	0.038	0.258	342	260	0.431	0.412
One adult in household	0.295	0.265	0.030	0.434	340	260	0.458	0.442
Free/reduced price lunch status	0.809	0.839	-0.030	0.342	322	242	0.425	0.368
Special education status	0.071	0.105	-0.034	0.220	311	239	0.277	0.307
Family income less than 15K	0.220	0.181	0.039	0.288	285	204	0.416	0.386
Family income between 15K and 25K	0.263	0.250	0.013	0.760	285	204	0.418	0.434
Family income between 25K and 35K	0.168	0.255	-0.087*	0.028	285	204	0.378	0.437
Family income between 35K and 50K	0.183	0.186	-0.003	0.936	285	204	0.393	0.390
Family income 50K or greater	0.166	0.127	0.038	0.237	285	204	0.395	0.334
Mother's education: less than HS	0.182	0.223	-0.041	0.231	341	260	0.368	0.417
Mother's education: HS or GED	0.244	0.258	-0.013	0.717	341	260	0.428	0.438
Mother's education: some college	0.272	0.265	0.007	0.857	341	260	0.453	0.442
Mother's education: college	0.301	0.254	0.047	0.192	341	260	0.464	0.436
Student has access to computer with internet at home	0.813	0.837	-0.025	0.493	301	215	0.370	0.370
Parent helps student with homework at least 5 days per week	0.709	0.698	0.011	0.800	299	212	0.449	0.460
Parent discussed college with student at least twice during pre-baseline school year	0.859	0.863	-0.004	0.905	299	212	0.365	0.344

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

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

For the high school student survey, there were no statistically significant differences between the treatment and comparison group on any baseline characteristics (Table B.5). Assessing baseline equivalence separately for the two different models, the treatment and comparison groups in the model using matched middle schools are equivalent, which is consistent with the model's design due to the matching process. In the model using adjacent cohorts from the same middle school, the groups are also generally equivalent—no differences between the groups are significant at the 5% level, and only one difference (whether a student is Hispanic) is marginally significant.

Table B.5. High school student survey baseline equivalence

Baseline characteristic	Treatment	Comp.	Difference	p-value	N_t	N_c	SD_t	SD_c
Matched middle school model								
Baseline reading exam (z-score)	0.100	0.101	-0.001	0.989	264	264	0.743	0.762
Baseline math exam (z-score)	0.547	0.593	-0.046	0.536	264	264	0.870	0.822
Male	0.409	0.405	0.004	0.930	264	264	0.493	0.492
Black, non-Hispanic	0.508	0.515	-0.008	0.862	264	264	0.501	0.501
Hispanic or Latino	0.356	0.390	-0.034	0.419	264	264	0.480	0.489
Old for grade	0.167	0.144	0.023	0.472	264	264	0.373	0.352
Adjacent cohort model								
Baseline reading exam (z-score)	0.444	0.367	0.077	0.347	206	209	0.825	0.834
Baseline math exam (z-score)	0.810	0.758	0.052	0.547	206	209	0.771	0.986
Male	0.353	0.393	-0.040	0.381	221	229	0.479	0.489
Black, non-Hispanic	0.520	0.555	-0.034	0.468	221	229	0.501	0.498
Hispanic or Latino	0.466	0.384	0.082	0.080	221	229	0.500	0.487
Old for grade	0.231	0.197	0.034	0.377	221	229	0.422	0.398

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

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

Defining indices

Several of the measures in Chapters II, III, and IV of this report are derived from survey items. Many of the outcomes are indices created by combining closely related survey items into a single measure, reducing measurement error, and capturing the breadth of a construct. The indices and their component items are listed at the end of this appendix in Tables B.9 through B.11. The tables also include survey outcomes or measures that are not indices, but are derived from one or more survey items.

The process for creating the indices included a number of steps to maximize reliability and reduce the number of separate outcome variables we examined (that is, to reduce dimensionality). We first identified all items from the surveys that were conceptually related to a specific construct. We used principal component analysis to confirm that the items were related to the underlying construct (and to one another) in the theoretically expected way, excluding

items not related to this underlying construct. We then created an index variable based on the included items to represent the underlying construct. Finally, we computed the standardized Cronbach's alpha, an estimate of the internal consistency or reliability of an index, and dropped indices with alpha values suggesting low reliability. Conventionally, indices with alpha values greater than 0.7 are considered reliable. Following Gleason et al. (2010), we retained indices with alpha values somewhat lower than this threshold but indicate that these indices may have low levels of reliability. Indices with values of alpha below 0.7 are noted in the tables.

Detailed survey outcomes tables

The following tables provide detailed results from our analysis of the impact of KIPP on survey-based (non-achievement) outcomes for elementary, middle, and high school students. Further details on the analytic methods used can be found in Appendix D for the elementary school analysis, Appendix E for the middle school analysis, and Appendix H for the high school analysis.

Table B.6. Estimated impact of offer of admission to KIPP elementary school

Outcome	Mean lottery winner	Mean lottery loser	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Student motivation and engagement								
Minutes spent on homework on typical night, parent report	55.11	51.86	3.26	0.25	341	327	38.31	38.01
Parent says student typically completes homework	0.95	0.92	0.03	0.05	340	327	0.18	0.28
Frequency of negative comments to parent about school (index)	1.26	1.26	0.00	0.99	342	346	0.42	0.42
Frequency of positive comments to parent about school (index)	2.79	2.69	0.10**	<0.01	341	347	0.42	0.48
Index of student development compared to peers	3.25	3.31	-0.06	0.09	341	346	0.46	0.44
Student behavior								
Index indicating student is well adjusted to school	3.66	3.60	0.06	0.08	338	343	0.42	0.42
School experiences and satisfaction								
Index of parental satisfaction with school	3.63	3.36	0.27**	<0.01	341	341	0.49	0.65
Parent rates school as excellent	0.77	0.51	0.26**	<0.01	344	351	0.42	0.50
Index of parental perceptions of problems in student's school	1.92	1.84	0.07	0.44	289	270	1.15	1.07
Index of parent outreach to school[a]	0.70	0.70	0.01	0.77	337	343	0.27	0.26
Index of school efforts to engage parents in school	2.53	2.39	0.14**	<0.01	327	326	0.44	0.53
Index of family engagement at home	2.83	2.81	0.03	0.59	340	348	0.62	0.62
Index of student engagement at home[a]	3.30	3.19	0.11	0.07	343	354	0.68	0.74
Index of parent indicating school is too easy[a]	0.10	0.12	-0.02	0.28	339	316	0.23	0.25
Index of parent indicating school is too hard[a]	0.02	0.04	-0.02	0.12	339	316	0.13	0.14
Education goals and aspirations								
Parent expects student to graduate high school on time	0.96	0.98	-0.02	0.17	338	345	0.21	0.14
Parent wishes student to complete college	0.99	0.98	0.00	0.66	341	351	0.11	0.13
Parent believes student very likely to complete college	0.81	0.71	0.10**	<0.01	340	349	0.40	0.45

Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information.

[a] Index has an alpha smaller than 0.7, indicating low reliability

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

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

Table B.7. Estimated impact of offer of admission to a KIPP middle school

Outcome	Mean lottery winner	Mean lottery loser	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Student motivation and engagement								
Minutes spent on homework on typical night, student report	78.64	71.42	7.22	0.18	272	176	53.38	50.65
Minutes spent on homework on typical night, parent report	99.79	84.97	14.82**	<0.01	326	244	57.41	52.56
Parent says student typically completes homework	0.91	0.92	-0.01	0.77	327	243	0.23	0.27
Index of positive school engagement	3.37	3.38	-0.01	0.82	279	197	0.46	0.45
Index of negative school engagement[a]	3.79	3.84	-0.05	0.24	280	198	0.47	0.43
Index of self-control	3.53	3.55	-0.02	0.66	280	194	0.50	0.49
Index of academic collaboration	3.30	3.18	0.12*	0.02	279	198	0.49	0.53
Index of effort in school	3.60	3.61	-0.01	0.89	281	198	0.43	0.44
Index of academic confidence	3.30	3.28	0.01	0.84	280	197	0.54	0.54
Index of grit	3.58	3.54	0.05	0.35	272	190	0.51	0.49
Student behavior								
Index of peer pressure for bad behavior	0.03	0.03	0.00	0.94	281	197	0.10	0.13
Index of student reported undesirable behavior	2.10	2.06	0.04	0.71	280	193	1.03	1.08
Index of parent report of undesirable behavior[a]	1.20	1.24	-0.04	0.16	337	255	0.32	0.35
Index of student reported illegal action[a]	0.03	0.02	0.00	0.59	281	196	0.09	0.08
Index of good behavior, student report[a]	4.42	4.44	-0.03	0.80	280	196	0.98	1.07
Index indicating well-adjusted student	3.44	3.49	-0.05	0.23	334	252	0.44	0.43
Parent reported any school disciplinary problems for student	0.35	0.32	0.02	0.53	338	253	0.47	0.47
Index of parent-reported frequency of school disciplinary actions for student[a]	0.22	0.18	0.04	0.17	338	253	0.38	0.31
Student never gets in trouble at school	0.44	0.51	-0.07	0.20	281	196	0.50	0.50
Index of good behavior, parent report[a]	2.26	2.28	-0.03	0.43	322	248	0.39	0.38
Index of parental concerns about student	1.41	1.41	-0.01	0.92	335	250	0.65	0.66
School experiences and satisfaction								
Index of student's feelings about school	3.35	3.27	0.08	0.06	274	196	0.41	0.45
Index of parental satisfaction with school	3.38	3.03	0.35**	<0.01	336	253	0.65	0.67
Parent rates school as excellent	0.56	0.28	0.29**	<0.01	339	256	0.50	0.45
Student likes school a lot	0.54	0.54	0.00	0.99	280	197	0.49	0.50
Index of parental perceptions of problems in student's school	2.00	2.24	-0.23*	0.04	277	184	1.12	1.14
Index of parent outreach to school[a]	0.64	0.64	0.01	0.81	324	251	0.24	0.27
Index of school efforts to engage parents in school	2.34	2.14	0.21**	<0.01	313	241	0.53	0.55
Parent talks to child about school experiences almost every day	0.80	0.79	0.01	0.82	340	258	0.39	0.41

Table B.7 *(continued)*

Outcome	Mean lottery winner	Mean lottery loser	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Parent helps child with homework almost every day	0.41	0.45	-0.04	0.32	339	257	0.49	0.50
Index of student perceptions of schoolmates	2.88	2.72	0.16**	<0.01	275	194	0.45	0.51
Index of student perceptions of teachers	3.48	3.41	0.07	0.08	281	195	0.39	0.42
Index of student perceptions of school disciplinary environment	3.37	3.24	0.13**	<0.01	281	196	0.44	0.41
Academic difficulty, parent report								
Index indicating school is too easy[a]	0.11	0.12	-0.01	0.73	323	238	0.23	0.24
Index indicating school is too difficult[a]	0.09	0.08	0.01	0.72	323	238	0.19	0.19
Education goals and aspirations								
Student expects to graduate high school on time	0.97	0.98	-0.01	0.41	281	198	0.14	0.14
Parent expects student to graduate high school on time	0.93	0.95	-0.03	0.12	337	254	0.21	0.21
Parent wishes student to complete college	0.98	0.99	-0.01	0.52	339	256	0.14	0.11
Parent believes student very likely to complete college	0.69	0.69	-0.01	0.90	331	253	0.45	0.46
Student reports having more than 2 discussions about college at school	0.46	0.47	-0.01	0.82	281	196	0.50	0.50
Student reports having more than 2 discussions about college at home	0.70	0.65	0.05	0.28	281	196	0.46	0.48
Parent report of having more than 2 discussions about college	0.84	0.82	0.02	0.60	338	256	0.37	0.38

Note: Estimates are from a model that pools impacts across grades and schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information. Some of the items presented in this table will be presented in an Appendix of the final report.

[a] Index has an alpha smaller than 0.7, indicating low reliability

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

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

Table B.8. Estimated impact of opportunity to attend KIPP high school

Outcome	Mean (T)	Mean (C)	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Student motivation and engagement								
Minutes spent on homework on a typical night	145.14	146.80	-1.66	0.85	410	398	103.37	94.43
Index of positive school engagement	3.19	3.18	0.01	0.89	491	492	0.61	0.59
Index of negative school engagement	1.75	1.67	0.07	0.32	493	500	0.89	0.81
Index of self control	4.43	4.38	0.05	0.45	401	394	0.71	0.68
Index of academic collaboration	3.05	3.07	-0.02	0.70	492	493	0.66	0.61
Index of academic motivation	3.21	3.20	0.01	0.84	486	491	0.54	0.55
Index of academic confidence	3.31	3.36	-0.05	0.26	493	501	0.56	0.55
Index of effort in school	3.28	3.22	0.06	0.32	492	496	0.65	0.68
Student dropped out of school	0.01	0.03	-0.02*	0.04	504	513	0.06	0.17
Education goals and aspirations								
Student thinks they will graduate college or go further in his/her education	0.84	0.85	-0.01	0.71	496	503	0.37	0.36
College preparation activities								
Index of student reports of discussions about college at school	2.59	2.40	0.19**	<0.01	471	481	0.52	0.58
Student discussed how to pay for college in depth at school	0.72	0.60	0.11**	<0.01	489	491	0.44	0.49
Index of student reports of teachers/counselor assistance with planning for college	3.37	3.15	0.22**	<0.01	477	481	0.77	0.87
Index of student reports of parent assistance with planning for college	3.04	2.94	0.10	0.22	475	473	0.93	1.03
Index of teacher helpfulness and encouragement	3.18	3.02	0.15*	0.02	477	478	0.71	0.77
Index of counselor helpfulness and encouragement	3.24	3.08	0.16*	0.03	466	471	0.79	0.85
Student reports having completed an education plan	0.86	0.86	0.00	1.00	459	466	0.34	0.33
Student reports having completed a career plan	0.75	0.71	0.04	0.36	459	466	0.44	0.45
College testing								
Student has taken PSAT or PLAN	0.82	0.80	0.02	0.50	485	500	0.37	0.40
Student has taken or plans to take SAT or ACT	0.89	0.85	0.03	0.26	487	500	0.31	0.36
Student has taken SAT or ACT	0.83	0.78	0.05	0.14	487	500	0.37	0.42
Student plans to take SAT or ACT	0.07	0.11	-0.04	0.13	487	500	0.25	0.31
Student reports participating in college prep program	0.25	0.21	0.04	0.21	484	477	0.44	0.40
Index of college preparation activities	0.88	0.80	0.08**	<0.01	460	459	0.19	0.23
Number of colleges or universities to which the student applied	6.99	6.90	0.09	0.84	490	496	4.95	5.83
College Application Process								
Student applied to at least 1 college/university	0.93	0.88	0.05*	0.04	490	496	0.24	0.32

Table B.8 *(continued)*

Outcome	Mean (T)	Mean (C)	Impact estimate	p-value	N_t	N_c	SD_t	SD_c
Accepted to at least 1 of the respondent's top 3 choices for college/university	0.81	0.76	0.05	0.12	483	487	0.38	0.43
Applied early admission/action/decision to one of top 3 college choices	0.41	0.42	-0.01	0.80	484	483	0.49	0.50
Coursetaking								
School offers honors courses	0.80	0.84	-0.04	0.24	426	426	0.40	0.37
School offers AP or IB courses	0.98	0.88	0.09**	<0.01	461	434	0.15	0.32
Number of honors courses student took or planned to take	1.25	1.95	-0.70*	0.01	411	440	1.97	2.99
Number of AP courses student took or planned to take	2.23	1.33	0.90**	<0.01	498	453	2.10	1.92
Number of AP exams student took or planned to take	1.96	1.14	0.82**	<0.01	480	435	2.11	1.77
How many years of the following subjects will the student complete:								
English or language arts	3.97	3.91	0.06	0.12	469	463	0.46	0.48
Math	3.93	3.87	0.06	0.20	470	466	0.51	0.59
Science or engineering	3.78	3.54	0.23**	<0.01	466	457	0.78	0.89
Social studies or history	3.79	3.42	0.37**	<0.01	469	460	0.59	0.83
Foreign language	2.92	2.73	0.19*	0.02	467	451	0.90	0.97
Physical education	2.36	2.30	0.06	0.56	467	451	1.33	1.29
Art	1.50	1.53	-0.02	0.82	462	460	1.09	1.22
Music	1.33	0.90	0.44**	<0.01	451	435	1.29	1.28
Technical or vocational education	0.53	0.61	-0.08	0.44	413	416	1.02	1.14
Total number of high school courses student took or plans to take	24.31	22.91	1.40**	<0.01	470	466	4.67	5.31
School experiences and satisfaction								
Index of student's feelings about school	3.39	3.40	-0.01	0.77	490	494	0.51	0.52
Student likes school a lot	0.37	0.45	-0.08	0.06	500	507	0.49	0.50
Index of student perceptions of schoolmates	3.01	3.04	-0.03	0.55	493	502	0.55	0.59
Index of student perceptions of teachers	3.33	3.34	-0.01	0.89	483	501	0.56	0.57
Index of school disciplinary environment	3.40	3.38	0.02	0.62	490	501	0.51	0.50
Student was suspended or expelled within the last school year	0.12	0.11	0.00	0.91	491	499	0.32	0.32
Student feels is getting a good education at the high school (strongly agree)	0.54	0.54	0.00	0.98	497	495	0.49	0.50
Index of helpfulness of high school in career planning	2.64	2.50	0.14	0.08	487	485	0.94	0.93

Note: Estimates are from a model that pools impacts across schools, weights schools based on sample size, and regression imputes missing values of covariates based on available baseline information. All indices have alpha values larger than 0.7. Standard errors are adjusted for the fact that some students are included in the sample multiple times.

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

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

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

DESCRIPTION OF INDICES

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

Table B.9. Description of indices: elementary school parent survey

Variable	Principal survey items included	Scale/Definition
Frequency of negative comments to parent about school (index)	F1. On average, during the first two months of this school year how often... **a. did the student complain about school?** **b. was the student upset or reluctant to go to school?** **c. did the student pretend to be sick or stay home from school?**	Mean across items (a), (b), and (c) using the following scale: Not at all (1) Once a week or less (2) More than once a week (3)
Frequency of positive comments to parent about school (index)	F1. On average, during the first two months of this school year how often... **d. did the student say good things about school?** **e. did the student say he/she like his/her teacher?** **f. did the student look forward to going to school?**	Mean across items (d), (e), and (f) using the following scale: Not at all (1) Once a week or less (2) More than once a week (3)
Index of student development compared to peers	F2. How does the student compare to other children of the same age? **a. Student is independent and takes care of himself/herself** **b. Student pays attention** **c. Student learns, thinks, and solves problems**	Mean across items (a), (b), and (c) using the following scale: Much less well than other children (1) Slightly less well than other children (2) As well as other children (3) Slightly better than other children (4) Better than other children his/her age (5)
Index indicating a student is well-adjusted to school	F3. For each of the following statements, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree." **a. The student gets along with others** **b. The student likes schools** **c. The student works hard at school** **d. The student is self-confident** **e. The student is creative** **f. The student is happy** **g. The student respects adults**	Mean across items (a) through (g) using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of parental satisfaction with school	D6. Please rate each of the following features of the school the student attends/attended for the 2012-2013 school year, as "excellent," "good," "fair," or "poor." **a. Facilities, like the classrooms, library, cafeteria, or the gym** **b. Academics, the teachers, and classes** **c. Safety** **d. Discipline**	Mean across items (a) through (d) using the following scale: Poor (1) Fair (2) Good (3) Excellent (4)

Table B.9 *(continued)*

Variable	Principal Survey Items Included	Scale/Definition
Index of parental perceptions of problems in student's school	D1. For each of the following issues, please tell me if you feel it [is/was] "not a problem," "a small problem," "a medium problem," or "a big problem" at this school. **a. Student destroying property** **b. Student being late for school** **c. Student missing school** **d. Fighting** **e. Bullying** **f. Stealing**	Mean across items (a) through (f) using the following scale: Not a problem (1) A small problem (2) A medium problem(3) A big problem (4)
Index of parent outreach to school	C3. Since the beginning of this school year, have you or any other adults in your household… **a. Attended an open house or a back-to-school night?** **b. Attended a meeting of a PTA, PTO, or Parent-Teacher Student Organization?** c. Gone to a parent-teacher conference or meeting with his/her teacher? **d. Attended a school or class event, such as a play, sports event, or science fair?** **e. Acted as a volunteer at the school or served on a committee?** **f. Participated in fundraising for the student's school?** g. Visited the school's website for information?	Mean across items (a), (b), (d), (e), and (f) using the following scale: No (0) Yes (1)
Index of school efforts to engage parents in school	C4. How well would you say the student's school… **a. Lets you know between report cards how the student is doing in school?** **b. Helps you understand what children at the student's age are like?** **c. Makes you aware of chances to volunteer at the school?** **d. Provides workshops, materials, or advice about how to help he student learn at home?** **e. Provides information on community services to help the student or your family?** **f. Communicates with you through email or online forums?** **g. Provides online resources?**	Mean across items (a) through (g) using the following scale: Not at all (1) Just O.K (2) Very well (3)

Table B.9 *(continued)*

Variable	Principal survey items included	Scale/Definition
Index of family engagement at home	C8. In a typical week, how often do you or any other family or household member do the following things with him/her? a. Read books to him/her? b. Tell stories to him/her? **c. Sing songs with him/her?** **d. Help him/her to do arts and crafts?** **e. Involve him/her in household chores, like cooking, cleaning, setting the table, or caring for pets?** **f. Play games or do puzzles with him/her?** **g. Talk about nature or do science projects with him/her?** **h. Build something or play with construction toys with him/her?** **i. Play a sport or exercise together?**	Mean across items (c) though (i) using the following scale: Not at all (1) Once or twice (2) 3 to 6 times (3) Every day (4)
Index of student engagement at home	C10. How often did he/she look at picture books outside of school in the past week? C11. In the past week, how often did he/she read to or pretend to read to himself/herself or to others outside of school?	Mean across C10 and C11 using the following scales: Never (1) Once or twice (2) 3 to 6 times (3) Every day (4)
Index of parent indicating school is too easy	B4. Do you think the homework (is/was) too difficult, about right, or too easy for the student? B5a. Do you think the material covered in (his/her) math class (is/was) too difficult, about right, or too easy for the student? B6a. Do you think the material covered in (his/her) English/language arts class (is/was) too difficult, about right, or too easy for the student?	Mean across all items using the following scale: Too easy (1) About right/Too difficult (0)
Index of parent indicating school is too difficult	See items B4, B5a, and B6a	Mean across all items using the following scale: Too difficult (1) Other responses (0)

Table B.10. Description of indices: middle school parent and student surveys

Variable	Survey questions	Scale/definition and included items
Index of positive school engagement	Student survey C3. For the following statements about school, do you do each one "almost always," "often," "sometimes," or "almost never?" **a. stick with a class assignment or task until it is done** **b. put in your best effort on class assignments, projects, and homework** **c. ask a teacher for help when you don't understand an assignment** **d. ask another student for help when you don't understand an assignment** **e. take part in class discussions or activities** **f. feel challenged in class** **g. receive recognition or praise for doing good school work** **h. learn from your mistakes at school** **j. complete class assignments, projects, and homework on time**	Mean across items (a) though (i) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost Always (4)
Index of negative school engagement	Student survey C3. For the following statements about school, do you do each one "almost always," "often," "sometimes," or "almost never?" **i. think of dropping out of school** **j. try to stay home from school**	Mean across items (j) and (k) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost Always (4)
Index of self-control	Student survey C4. Thinking about a typical week during the 2012-2013 school year, how often did you do each of the following things? **a. Went to all of your classes prepared** **b. Remained calm even when things happened that could upset you** **c. Paid attention in all of your classes** **d. Listened to other students speak without interrupting them** **e. Were polite to adults and other students** **f. Remembered and followed directions** **g. Controlled your temper** i. Got to work right away rather than procrastinating	Mean across items (a) though (g) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost Always (4)

Table B.10 *(continued)*

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

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of academic confidence	Student survey B1. I'm going to read you some statements about your schoolwork. For each please tell me if these things apply to you "almost always," "often," "sometimes," or "almost never"? **a. You're sure you can understand even the most difficult material presented in textbooks or other written material** **b. You can learn something really difficult when you want to** c. In school you work as hard as possible **d. You're certain you can understand even the most difficult material presented by the teacher** **e. If you decide to not get any bad grades, you can really do it** f. In school, you keep working even if the material is difficult **g. You're certain you can do an excellent job on assignments and tests** h. You try to do your best to learn the knowledge and skills taught i. You work hard in school so you can get a good job j. If you want to learn something well, you can **k. You're certain you can master the material you are taught** l. If you don't understand something in your schoolwork, you try to find additional information to help you learn **m. You put forth your best effort in school**	Mean across items (a), (b), (d), (e), (g) and (k) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of Grit	Student survey C5. For each item, tell us whether it is very much like you, mostly like you, somewhat like you, not much like you or not at all like you. **a. New ideas and projects sometimes distract you from previous ones.** **b. Delays and obstacles don't discourage you** **c. You have been obsessed with a certain idea or project for a short time but later lost interest.** **d. You are a hard worker** **e. You often set a goal but later choose to follow a different one** **f. You have difficulty keeping your focus on projects that take more than a few months to complete** **g. You finish whatever you begin** **h. You are hard working and careful**	Mean across all items using the following two scales: For items (a), (c), (e), and (f): Very much like you (1) Mostly like you (2) Somewhat like you (3) Not much like you (4) Not like you at all (5) For items: (b), (d), (g), and (h): Very much like you (5) Mostly like you (4) Somewhat like you (3) Not much like you (2) Not like you at all (1)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of Peer Pressure for Bad Behaviors	Student survey E2a. During the 2012-2013 school year, did your friends pressure you to do any of the following things? **a. Skip class or school?** **b. Drink alcohol?** **c. Smoke cigarettes?** **d. Use marijuana or other drugs?** e. Commit a crime or do something violent?	Mean across items (a) through (d) using the following scale: Yes (1) No (0)
Index of student reported undesirable behavior	Student survey E1. During the 2012-2013 school year, how often have you: **a. Argued with your parents or guardians?** b. Smoked cigarettes? **c. Lied to your parents or guardians?** d. Stolen something from someone else? e. Taken something from a store without paying for it? **f. Given a teacher a hard time?** g. Drunk alcohol? h. Skipped, or cut, classes during school? i. Skipped, or cut, the entire school day? j. Used marijuana or other drugs? **k. Gotten in trouble at school?** **l. Lost your temper at home or at school?** m. Gotten arrested or held by police?	Mean across items (a), (c), (f), (k), and (l) using the following scale: Just once or twice during the past year (1) Every couple of months (2) About once a month (3) About once a week (4) Almost every day (5)
Index of student reported illegal action	Student survey E1. During the 2012-2013 school year, have you ever done any of the following things? a. Argued with your parents or guardians? b. Smoked cigarettes? c. Lied to your parents or guardians? **d. Stolen something from someone else?** **e. Taken something from a store without paying for it?** f. Given a teacher a hard time? **g. Drunk alcohol?** h. Skipped, or cut, classes during school? i. Skipped, or cut, the entire school day? j. Used marijuana or other drugs? k. Gotten in trouble at school? l. Lost your temper at home or at school? m. Gotten arrested or held by police?	Mean across items (d), (e), (g), and (m) using the following scale: Yes (1) No (0)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of parent report of undesirable behavior	Parent survey F6. During the 2012-2013 school year, how often did [STUDENT] do the following things? **a. Break something on purpose** **b. Punch or hit someone in anger** **c. Argue with you**	Mean across all items using the following scale: Never (1) Sometimes (2) Often (3)
Parent reported any school disciplinary problems for student	Parent survey F5. During the 2012-2013 school year, how many times was [STUDENT]… a. Sent out of class for disciplinary reasons? b. Suspended from school? c. Expelled from school?	Respondent receives a value of 1 if they provided a nonzero answer to (a), (b), or (c), and 0 otherwise.
Index of Parent-Reported Frequency of School Disciplinary Actions for Student	See item F5 above.	Mean across all items using the following scale: Not at all (0) 1 to 3 times (1) 4 to 6 times (2) 7 to 10 times (3) More than 10 times (4)
Index of good behavior, student report	Student survey E3b.During the 2012-2013 school year, how often have you: **a. Helped another student with school work** **b. Helped people in your local community, for example, help a neighbor, or do volunteer work** c. Read for fun d. Gone to the library outside of school **e. Helped your parents or guardians with chores**	Mean across items (a), (b), and (e) using the following scale: Just once or twice during the past year (1) Every couple of months (2) About once a month (3) About once a week (4) Almost every day (5)
Index of good behavior, parent report	Parent survey F7. During the 2012-2013 school year, how often did [STUDENT] do the following things? **a. Help you with chores or other tasks** **b. Stay and help teachers in (his/her) classrooms** **c. Help people in your local community, for example, help a neighbor, or do volunteer work** **d. Read for fun** **e. Go to the library outside of school**	Mean across all items using the following scale: Never (1) Sometimes (2) Often (3)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index indicating well-adjusted student	Parent survey F3. For each of the following statements, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree." **a. The student gets along with others** **b. The student likes school** **c. The student works hard at school** **d. The student is self-confident** **e. The student is creative** **f. The student is happy** **g. The student respects adults**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of Parental Concerns About Student	Parent survey F4. For each of the following statements, please tell me if it is "not a problem," "a small problem," "a medium problem," or "a big problem" with [STUDENT] in or out of school. **a. Getting into trouble** **b. Smoking, drinking alcohol or using drugs** **c. The friends (he/she) has chosen** **d. (His/Her) academic achievement**	Mean across all items using the following scale: Not a problem (1) A small problem (2) A medium problem (3) A big problem (4)
Index of student's feelings about school	Student survey A1. Now, I'm going to read you some statements on how you (feel/felt) about school. For each, statement, please tell me if you "strongly agree," "agree," "disagree," or "strongly disagree". a. You have good friends at your school **b. You are treated fairly at your school** **c. You are happy to be at your school** **d. You feel like you are part of the community in your school** **e. You feel safe in your school** **f. You are treated with respect at your school** g. You know how you are doing in school h. You have the materials and equipment you need to do your school work right **i. You get the chance to be independent at school** j. You have opportunities to choose how you learn	Mean across items (b), (c), (d), (e), (f), (i), and (j) using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of parental satisfaction with school	Parent survey D6. Please rate each of the following features of the school the student attends/attended for the 2012-2013 school year, as "excellent," "good," "fair," or "poor." **a. Facilities, like the classrooms, library, cafeteria, or the gym** **b. Academics, the teachers, and classes** **c. Safety** **d. Discipline**	Mean across all items using the following scale: Poor (1) Fair (2) Good (3) Excellent (4)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of student perceptions of schoolmates	Student survey A2. Please tell me how much you agree or disagree with these statements about the students in your classes (this/last) year at school: **a. Students usually complete their homework** **b. Students get along well with the teachers** **c. Students are interested in learning** **d. Students help one another** e. Students are well behaved	Mean across items (a), (b), (c), and (d) using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of student perceptions of teachers	Student survey A3. These next statements are about your teachers (this/last) year at school. Again please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree" with each statement. **a. They are available for help** **b. They listen to what you have to say** **c. They give corrections and suggestions for improvement** **d. They care about students** **e. They encourage you to think about your future** **f. Their classes are challenging** **g. They make you feel like your school work is important** **h. You like your teachers**	Mean across all items using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of school disciplinary environment	Student survey A4. For each of the following statements about the rules (this/last) year at your school, please tell me whether you "strongly agree," "agree," "disagree," or "strongly disagree." **a. Everyone knows what the school rules are** **b. The school rules are fair** **c. The punishment for breaking school rules is the same no matter who you are** **d. If a school rule is broken, students know what the punishment will be** **e. Students receive specific positive rewards for good behavior** **f. You follow the rules at school**	Mean across all items using the following scale: Strongly disagree (1) Disagree (2) Agree (3) Strongly agree (4)
Index of parental perceptions of problems in student's school	Parent survey D1. For each of the following issues, please tell me if you feel it [is/was] "not a problem," "a small problem," "a medium problem," or "a big problem" at this school. **a. Students destroying property** **b. Students being late for school** **c. Students missing school** **d. Fighting** **e. Bullying** **f. Stealing**	Mean across all items using the following scale: Not a problem (1) A small problem (2) A medium problem(3) A big problem (4)

Table B.10 *(continued)*

Variable	Survey questions	Scale/definition and included items
Index of parent outreach to school	Parent survey C3. Since the beginning of this school year, have you or any other adults in your household… **a. Attended an open house or a back-to-school night?** **b. Attended a meeting of a PTA, PTO, or Parent-Teacher Student Organization?** **c. Gone to a parent-teacher conference or meeting with his/her teacher?** **d. Attended a school or class event, such as a play, sports event, or science fair?** **e. Acted as a volunteer at the school or served on a committee?** **f. Participated in fundraising for the student's school?** **g. Visited the school's website for information?**	Mean across all items using the following scale: No (0) Yes (1)
Index of school efforts to engage parents in school	Parent survey C4. How well would you say the student's school… **a. Lets you know between report cards how the student is doing in school?** **b. Helps you understand what children at the student's age are like?** **c. Makes you aware of chances to volunteer at the school?** **d. Provides workshops, materials, or advice about how to help the student learn at home?** **e. Provides information on community services to help the student or your family?** **f. Communicates with you through email or online forums?** **g. Provides online resources?**	Mean across all items using the following scale: Not at all (1) Just O.K (2) Very well (3)
Index indicating school is too easy	Parent survey B4. Do you think the homework (is/was) too difficult, about right, or too easy for the student? Parent survey B5a. Do you think the material covered in (his/her) math class (is/was) too difficult, about right, or too easy for the student? Parent survey B6a. Do you think the material covered in (his/her) English/language arts class (is/was) too difficult, about right, or too easy for the student?	Mean across all items using the following scale: Too easy (1) About right/Too difficult (0)
Index indicating school is too difficult	See items B4, B5a, and B6a above	Mean across all items using the following scale: This responses (0)

Table B.11. Description of indices: high school student survey

Variable	Principal survey items included	Scale/definition
Index of positive school engagement	C3. Please select how often you (do/did) each of the following things: **a. I (stick/stuck) with a class assignment or task until it (is/was) done** **b. I (ask/ed) a teacher for help when I (don't/didn't) understand an assignment** **c. I (take/took) part in class discussions or activities** **d. I (receive/received) recognition or praise for doing good school work** **e. I (learn/learned) from my mistakes at school** **f. I (complete/completed) class assignments, projects, and homework on time**	Mean across items (a) through (f) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of negative school engagement	C3. Please select how often you (do/did) each of the following things: **g. I (think/thought) of dropping out of school** **h. I (try/tried) to stay home from school** **i. I (am/was) absent from school**	Mean across items (g), (h), and (i) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of self control	C4. Last week in school, how many days did you do each of the following things? a. I went to all of my classes prepared b. I remained calm even when things happened that could upset me c. I paid attention in all of my classes d. I listened to other students speak without interrupting them e. I was polite to adults and other students f. I remembered and followed directions g. I controlled my temper h. I got to work right away rather than procrastinating	Mean number of days across all items
Index of academic collaboration	C2. For the following statements, select how much you agree or disagree with each. **a. I like to work with other students** b. I learn things quickly in most school subjects c. Because reading is fun, I wouldn't want to give it up d. I am good at most school subjects **e. I learn most when I work with other students** f. English/Language Arts is one of my best subjects **g. I do my best work when I work with other students** h. Math is one of my best subjects **i. I like to help other people do well in group assignments** j. I do well in tests in most school subjects **k. It is helpful to put together everyone's ideas when I work on a project** l. If I don't do well on a test, I study harder next time m. I set aside time to do my homework and study n. I try to do well on my schoolwork even when it isn't interesting to me o. Grades in high school matter for success in college p. What I learn in class is necessary for success in the future	Mean across items (a), (e), (g), (i), and (k) using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of academic motivation	C2. For the following statements, select how much you agree or disagree with each. **l. If I don't do well on a test, I study harder next time** **m. I set aside time to do my homework and study** **n. I try to do well on my schoolwork even when it isn't interesting to me** **o. Grades in high school matter for success in college** **p. What I learn in class is necessary for success in the future**	Mean across items (l) through (p) using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of academic confidence	C1. For each please select if these things apply to you "almost always," "often," "sometimes," or "almost never." **a. I (am/was) certain I (can/could) understand even the most difficult material presented in textbooks or other written material** **b. I (can/could) learn something really difficult when I (want/ed) to** c. In school I (work/ed) as hard as possible **d. I (am/was) certain I (can/could) understand even the most difficult material presented by the teacher** **e. If I (decide/d) to not get any bad grades, I (can/could) really do it** **f. I (am/was) certain I (can/could) do an excellent job on assignments and tests** g. I (work/worked) hard in school so I (can/could) get a good job h. I (work/ed) hard in school so I (can/could) go to college **i. If I (want/ed) to learn something well, I (can/could)** **j. (am/was) certain I (can/could) master the material I (am/was) taught** k. If I (don't/didn't) understand something in my schoolwork, I (try/tried) to find other resources to help me learn	Mean across items (a), (b), (d), (e), (f), (i), and (j) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of effort in school	C1. For each please select if these things apply to you "almost always," "often," "sometimes," or "almost never." **a. I (am/was) certain I (can/could) understand even the most difficult material presented in textbooks or other written material** **b. I (can/could) learn something really difficult when I (want/ed) to** c. In school I (work/ed) as hard as possible **d. I (am/was) certain I (can/could) understand even the most difficult material presented by the teacher** **e. If I (decide/d) to not get any bad grades, I (can/could) really do it** **f. I (am/was) certain I (can/could) do an excellent job on assignments and tests** g. I (work/worked) hard in school so I (can/could) get a good job h. I (work/ed) hard in school so I (can/could) go to college **i. If I (want/ed) to learn something well, I (can/could)** **j. I (am/was) certain I (can/could) master the material I (am/was) taught** k. If I (don't/didn't) understand something in my schoolwork, I (try/tried) to find other resources to help me learn	Mean across items (c), (g), (h), and (k) using the following scale: Almost never (1) Sometimes (2) Often (3) Almost always (4)
Index of student reports of discussions about college at school	F7. (Has/Did) anyone at your school (discussed/discuss) the following with you? **a. Different admissions requirements for two-year vs. four-year colleges** **b. Different admissions requirements among four-year colleges** **c. How to decide which college to attend** **d. Your likelihood of being accepted at different types of schools** **e. What ACT/SAT scores you need to get into the colleges you want to attend** **f. Opportunities to attend out-of-state schools** **g. Your readiness for college-level coursework** **h. What kinds of study skills you will need in college or vocational/technical school** **i. How to pay for college**	Mean across all items using the following scale: Did not discuss (1) Discussed briefly (2) Discussed in depth (3)
Index of student reports of teachers/counselor assistance with planning for college	F8. How much (do/did) your high school teachers or guidance counselors… **a. Encourage you to apply to several different schools** **b. Talk to you about what college would be like** **c. Help you fill out applications for colleges or vocational/technical schools** **d. Help you find scholarships to apply for** **e. Help you decide which school to attend** **f. Help you plan how to pay for tuition and other expenses** **g. Help you with your college application essays or personal statements**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of student reports of parent assistance with planning for college	F9. How much (do/did) your parent(s)/guardian(s)… **a. Encourage you to apply to several different schools** **b. Talk to you about what college would be like** **c. Help you fill out applications for colleges or vocational/technical schools** **d. Help you find scholarships to apply for** **e. Help you decide which school to attend** **f. Help you plan how to pay for tuition and other expenses** **g. Help you with your college application essays or personal statements**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)
Index of college preparation activities	F5. At anytime during high school did you. . . **a. Attend college fairs** **b. Speak with college representatives** **c. Visit in-state college campuses** **d. Visit out-of-state college campuses** e. Sit in on a college-level course f. Participate in a college summer program or early college access program **g. Take practice ACT/SAT exams** h. Use college guidebooks (on-line or print) **i. Obtain information from college websites** j. Apply for financial aid k. Apply for a scholarship **l. Research career possibilities**	Mean across items (a), (b), (c), (d), (g), (i), and (l) using the following scale: No (0) Yes (1)
Index of student's feelings about school	B1. Here are some statements related to how you feel about your school. For each statement, please select whether you “strongly agree,” “agree,” “disagree,” or “strongly disagree.” **a. I (have/had) good friends at my high school** **b. I (am/was) treated fairly at my high school** **c. I (am/was) happy to be at my high school** **d. I (feel/felt) like I (am/was) part of the community in my high school** **e. I (feel/felt) safe in my high school** **f. I (am/was) treated with respect at my high school** **g. I (know/knew) how I (am/was) doing in high school** **h. I (have/had) the materials and equipment I need(ed) to do my school work right** **i. I (get/got) the chance to be independent at high school** **j. High school (is/was) seen as preparation for the future** **k. All students (are/were) encouraged to go to college**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of student perceptions of schoolmates	B2. Please select how much you agree or disagree with these statements about the students in your classes at your school. **a. Students usually (complete/d) their homework** **b. Students (get/got) along well with the teachers** **c. Students(are/were) interested in learning** **d. Students(help/ed) one another** **e. Students (are/were) well behaved**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of student perceptions of teachers	B3. Please select how much you agree or disagree with these statements about your teachers at your school. **a. Teachers are/were available for help** **b. Teachers listen/ed to what I have to say** **c. Teachers give/gave corrections and suggestions for improvement** **d. Teachers care/d about students** **e. Teachers make/made me feel like my school work is important** **f. Teachers work/ed hard to make sure that all students are learning** **g. Teachers work/ed hard to make sure that students stay in school** **h. I like/d my teachers**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of school disciplinary environment	B5. Please select to what extent you agree or disagree with the following statements about your school. **a. Behavioral standards and discipline policies (are/were) established and enforced consistently across the entire school** **b. The school (has/had) a zero-tolerance policy for potentially dangerous behaviors** **c. The school (has/had) a school-wide behavior code that includes specific positive rewards for students who consistently (behave/d) well** **d. The school (has/had) a school-wide behavior code that (includes/ed) specific consequences for students who break the rules** **e. I (follow/ed) the rules at school**	Mean across all items using the following scale: Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4)
Index of helpfulness of high school in career planning	D2. How helpful has/was your school in the following: **a. Assessing your career interests and abilities** **b. Developing a career plan** **c. Providing information about occupations (e.g., salaries, working conditions, and future outlook of various occupations)** **d. Teaching job search techniques (e.g., where/how to look for jobs)** **e. Teaching resume writing** **f. Helping you find a job** **g. Teaching job interviewing skills**	Mean across all items using the following scale: Not helpful (1) Somewhat helpful (2) Helpful (3) Very helpful (4)

Table B.11 *(continued)*

Variable	Principal survey items included	Scale/definition
Index of teacher helpfulness and encouragement	D3. Think about the teachers that you (have) met with throughout high school. Please indicate how much your teachers (have done/did) the following: **a. Helped me select courses that (meet/met) my high school's graduation requirements** **b. Helped me select courses that I (need/ed) for work or admission to college** **c. Helped me decide what I (want/ed) to do after I graduate** **d. Encouraged me to take Advanced Placement (AP)/honors courses** **e. Encouraged me to continue my education after high school** **f. Talked to me about how to get a job** **g. Talked to me about colleges/schools that (are/were) suited to my interests and abilities**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)
Index of counselor helpfulness and encouragement	D4. Thinking about the counselors that you have met with throughout high school. Please indicate how much your counselors have done the following: **a. Helped me select courses that (meet/met) my high school's graduation requirements** **b. Helped me select courses that I (need/ed) for work or admission to college** **c. Helped me decide what I (want/ed) to do after I graduate** **d. Encouraged me to take Advanced Placement (AP)/honors courses** **e. Encouraged me to continue my education after high school** **f. Talked to me about how to get a job** **g. Talked to me about colleges/schools that (are/were) suited to my interests and abilities**	Mean across all items using the following scale: Not at all (1) A little (2) Some (3) A lot (4)

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

APPENDIX C

CUMULATIVE MIDDLE AND HIGH SCHOOL RESULTS

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

Our primary analyses of KIPP high schools employ two different approaches. The first analysis estimates the *marginal* impact of KIPP high schools among students who previously attended a KIPP middle school. This involves comparing one group of KIPP middle school students (those who had the option to attend a KIPP high school) to a different group of KIPP middle school students (those who did not have the option to attend a KIPP high school, because there was no KIPP high school available in their area at the time they completed middle school). A more detailed description of this analysis can be found in Appendix H. The second analysis focuses on students who entered KIPP for the first time in grade 9 and matches to these new entrants a set of comparison students who never attended KIPP at any level (elementary, middle, or high school). These matches between KIPP and comparison students were based on the students' demographic characteristics and "baseline" test scores in grade 7 and grade 8 (see Appendix G for more information about the matching model and other details of this second analysis).

As an additional exploratory analysis, this appendix presents the results of a third analysis of high school outcomes that examines the *combined* impact of KIPP middle schools and KIPP high schools. These estimates compare a set of KIPP high school students who came through KIPP middle schools with a matched set of comparison students who never attended KIPP at any point in their schooling. The KIPP and non-KIPP groups both attended non-KIPP elementary schools, but at that point their educational paths diverged. Since the groups attended different middle schools and high schools, the resulting impact estimates capture the cumulative effect of KIPP at both levels, and cannot distinguish between the separate effects of KIPP middle versus high schools.

The cumulative impact analysis of KIPP middle and high schools on high school outcomes uses a matched-student design. In particular, students who attended KIPP middle and high schools after attending non-KIPP elementary schools were identified, and comparison students who attended non-KIPP schools at all levels were matched to them. This matching was based primarily on their demographic characteristics and test scores in elementary school (typically grades 3 and 4). In addition, each of these KIPP students was matched to a comparison student who did not attend a KIPP middle or high school, but did remain in our data from elementary school through grade 9. In other words, students who dropped out or transferred to a different school district during middle school were not eligible for the comparison group. This requirement was added because the group of KIPP students, by definition, also remained in our data (and in KIPP schools) through grade 9, at minimum.

The sample of KIPP middle schools and linked high schools in this analysis (as well as the number of KIPP and matched comparison students in the sample for each school) is summarized in Table C.1. Overall, the analysis is based on a sample of 7,404 students and includes KIPP students who went through 25 KIPP middle schools and 14 KIPP high schools. To match the KIPP students and comparison group, we used the same propensity score model and imputation procedures as we used for the matched-student analysis of KIPP middle school impacts (described in Appendix F). The impact models also closely resemble the models used for our main analysis of high school outcomes described in Appendix G; the only difference is that for this supplemental analysis baseline test scores were defined in elementary school rather than middle school grades.

Table C.1. Cumulative middle and high school analysis sample

Middle school	Linked high school	Analytic baseline sample			Number of KIPP cohorts in data (School Years)
		Treatment (N)	Comparison (N)	Total sample size	
KIPP Academy Lynn Middle School	KIPP Academy Lynn Collegiate High School	182	182	364	4 (2006-07 to 2009-10)
KIPP WAYS Academy	KIPP Atlanta Collegiate	120	120	240	5 (2006-07 to 2010-11)
KIPP STRIVE Academy	KIPP Atlanta Collegiate	35	35	70	2 (2009-10 to 2010-11)
KIPP Austin College Prep	KIPP Austin Collegiate	308	308	616	7 (2003-04 to 2009-10)
KIPP Austin Acad. of Arts & Letters	KIPP Austin Collegiate	49	49	98	1 (2009-10)
KIPP DC: KEY Academy	KIPP DC College Preparatory	130	130	260	6 (2005-06 to 2010-11)
KIPP DC: AIM Academy	KIPP DC College Preparatory	168	168	336	6 (2005-06 to 2010-11)
KIPP DC: WILL Academy	KIPP DC College Preparatory	84	84	168	5 (2006-07 to 2010-11)
KIPP Delta College Preparatory School	KIPP Delta Collegiate High School	115	115	230	5 (2005-06 to 2009-10)
KIPP Sunshine Peak Academy	KIPP Denver Collegiate High School	254	254	508	7 (2004-05 to 2010-11)
KIPP 3D Academy	KIPP Generations Collegiate	163	163	326	5 (2006-07 to 2010-11)
KIPP Intrepid Preparatory	KIPP Generations Collegiate	69	69	138	3 (2008-09 to 2010-11)
KIPP Academy (Houston)	KIPP Houston High School	360	360	720	7 (2003-04 to 2009-10)
KIPP Sharpstown College Prep	KIPP Houston High School	130	130	260	3 (2007-08 to 2009-10)
KIPP Memphis Collegiate Middle	KIPP Memphis Collegiate High	157	157	314	4 (2008-09 to 2010-11)
KIPP Polaris Academy for Boys	KIPP Northeast College Prep	22	22	44	2 (2009-10 to 2010-11)
KIPP Voyage Academy for Girls	KIPP Northeast College Prep	35	35	70	2 (2009-10 to 2010-11)
KIPP Academy Middle School (New York)	KIPP NYC College Prep High School	162	162	324	6 (2004-05 to 2009-10)
KIPP STAR Harlem Middle School	KIPP NYC College Prep High School	147	147	294	6 (2004-05 to 2009-10)
KIPP Infinity Middle School	KIPP NYC College Prep High School	177	177	354	5 (2005-06 to 2009-10)
KIPP AMP Middle School	KIPP NYC College Prep High School	82	82	164	5 (2005-06 to 2009-10)
KIPP Gaston College Preparatory	KIPP Pride High School	315	315	630	9 (2001-02 to 2009-10)
KIPP Liberation College Prep	KIPP Sunnyside	64	64	128	5 (2006-07 to 2010-11)
KIPP Spirit College Prep	KIPP Sunnyside	139	139	278	5 (2006-07 to 2010-11)
KIPP Aspire Academy	KIPP University Prep High School	235	235	470	6 (2004-05 to 2009-10)
Total	**25 middle schools 14 high schools**	**3,702**	**3,702**	**7,404**	

Notes: Test outcomes are drawn from administrative records collected following enrollment in high school. Treatment students attended KIPP in both middle and high school, and comparison students are matched on baseline (elementary school) characteristics.

In comparison to the other matched-student analyses of KIPP schools presented in the main text of this report, the matched student design for estimating the cumulative impact of KIPP middle and high schools relies on a stronger set of assumptions. In particular, the matched-student analysis of middle school impacts uses matching on elementary school characteristics to account for students' decision to enter or not enter a KIPP middle school (that is, to account for selection into middle school). By contrast, the matched-student analysis of cumulative KIPP middle and high school impacts uses matching on elementary school characteristics to account for both selection into a KIPP middle school *and persistence in KIPP into the high school grades.* Persistence in a KIPP middle school may be correlated with attributes that we cannot observe in our data, to the extent that factors such as parental involvement, discipline, and students' "grit" and determination may not be fully reflected in their baseline test scores or demographic characteristics.[1] If these unobserved factors are also associated with improved outcomes in high school, the impact estimates we discuss below would be upwardly biased. We believe this potential for upward bias may be particularly likely to occur for the graduation outcome, because choosing to persist in the KIPP network throughout middle school may also predict persistence to graduation in high school. As a result, we believe the findings from this supplemental analysis should be interpreted with caution.[2]

Table C.2 presents the results. We find a pattern of positive and statistically significant impacts in all the academic subjects we examined (math, ELA, science, and social studies).[3] For linked KIPP middle and high schools, the estimated impacts in math, ELA, science, and social studies are 0.34, 0.29, 0.40, and 0.27 standard deviations, respectively. Relative to the high school test scores in the matched comparison group, these impacts represent an increase from the 49th percentile (equivalent to the average percentile of the comparison group) to the 63rd percentile in math, an increase from the 53rd to the 64th percentile in ELA, an increase from the 50th to the 65th percentile in science, and an increase from the 46th to the 56th percentile in social studies relative to the distribution of students in the same district. We also estimate that this group of linked KIPP middle and high schools increased four-year graduation rates by 13 percentage points, with a graduation effect that is positive and statistically significant.

[1] In our data, on average less than half of students entering a KIPP middle school and had the option to attend a KIPP high school persisted at KIPP into grade 9.

[2] On observable characteristics, however, the treatment and comparison groups are very similar for each of the outcome samples we examined. Full baseline equivalence results for this analysis are available from the authors upon request.

[3] For more information on the data and measures used for these high school outcomes, see Appendix G.

Table C.2. Cumulative impact of KIPP middle and high school on high school achievement and graduation

	Mean, KIPP students	Mean, non-KIPP students	Impact estimate	Number of schools	Number of students
Impacts on High School Standardized Subject Tests (z-scores)					
Mathematics achievement	0.32	-0.03	0.34** (0.04)	11	3,122[a]
ELA achievement	0.36	0.07	0.29** (0.02)	14	4,208
Science achievement	0.39	-0.01	0.40** (0.03)	14	3,704
Social studies achievement	0.16	-0.10	0.27** (0.05)	9	1,573[a]
Impacts on high school graduation rates (percentage points)					
Graduation within 4 years after entering grade 9	78.6	65.3	13.3** (2.2)	9	2,216

Source: State and district administrative records data.

Notes: Impacts for matched student analyses were calculated by comparing the outcomes of KIPP students to a set of matched comparison students with similar baseline (grade 4) achievement profiles and demographic characteristics. Impacts were calculated separately for each KIPP high school; the average impact estimates reported here assign an equal weight to each of the school-level impact estimates. In a given high school, the included test may be a either an end-of-course exam (i.e. algebra), or an end-of-grade exam (i.e. grade 10 mathematics). Means for the comparison group are unadjusted; means for the treatment group are equal to the comparison group mean plus the estimated impact. Standard errors are shown in parentheses.

[a] The high school exams used for this analysis varied by jurisdiction. For each subject and site, we selected the exam that was observed for the largest number of students during high school, provided that the percentage of students taking the exam was similar in the treatment and matched comparison group. Each site (and each student cohort within sites) was only included in the analysis of a given test if both the treatment and matched comparison groups took the relevant exam during high school at a similar rate. In this sample, data coverage was less consistent for mathematics and social studies exams, compared to ELA and science exams.

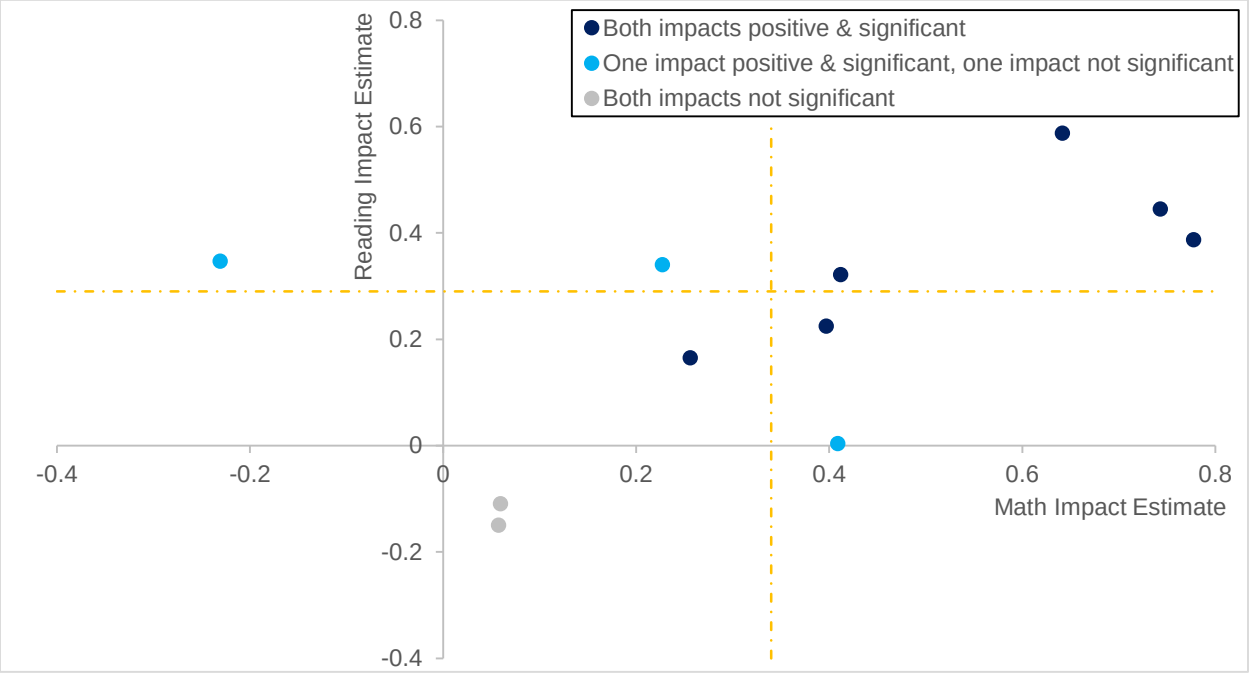
* Impact estimate is statistically significant at the 0.05 level, two-tailed test.

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

To summarize the variation in these impact estimates, Figure C.1 plots both the ELA and math impacts for the 11 sites with high school data on both outcomes. Of these, six sites have a significant and positive impact in both ELA and math, three sites have a significant and positive impact in one subject but not the other, and two sites do not have significant impacts in either subject. None of the sites has a negative and statistically significant impact on test scores in ELA or math.

While the magnitudes of these impacts are substantial, it is important to remember that the estimates represent the combined effect of KIPP middle schools and KIPP high schools. The analysis does not distinguish between the relative contributions from these two types of KIPP schools. This differs from both the analysis of new grade 9 entrants to KIPP and the matched-school analysis of the marginal impacts of KIPP high schools on students continuing from a KIPP middle school—both of those analyses examined the impacts of KIPP high schools alone. This makes it more difficult to interpret differences between the magnitudes of the impact estimates from this supplemental set of results and the study's primary two analyses of high school outcomes.

Figure C.1. Distribution of cumulative middle and high school impacts in ELA and math



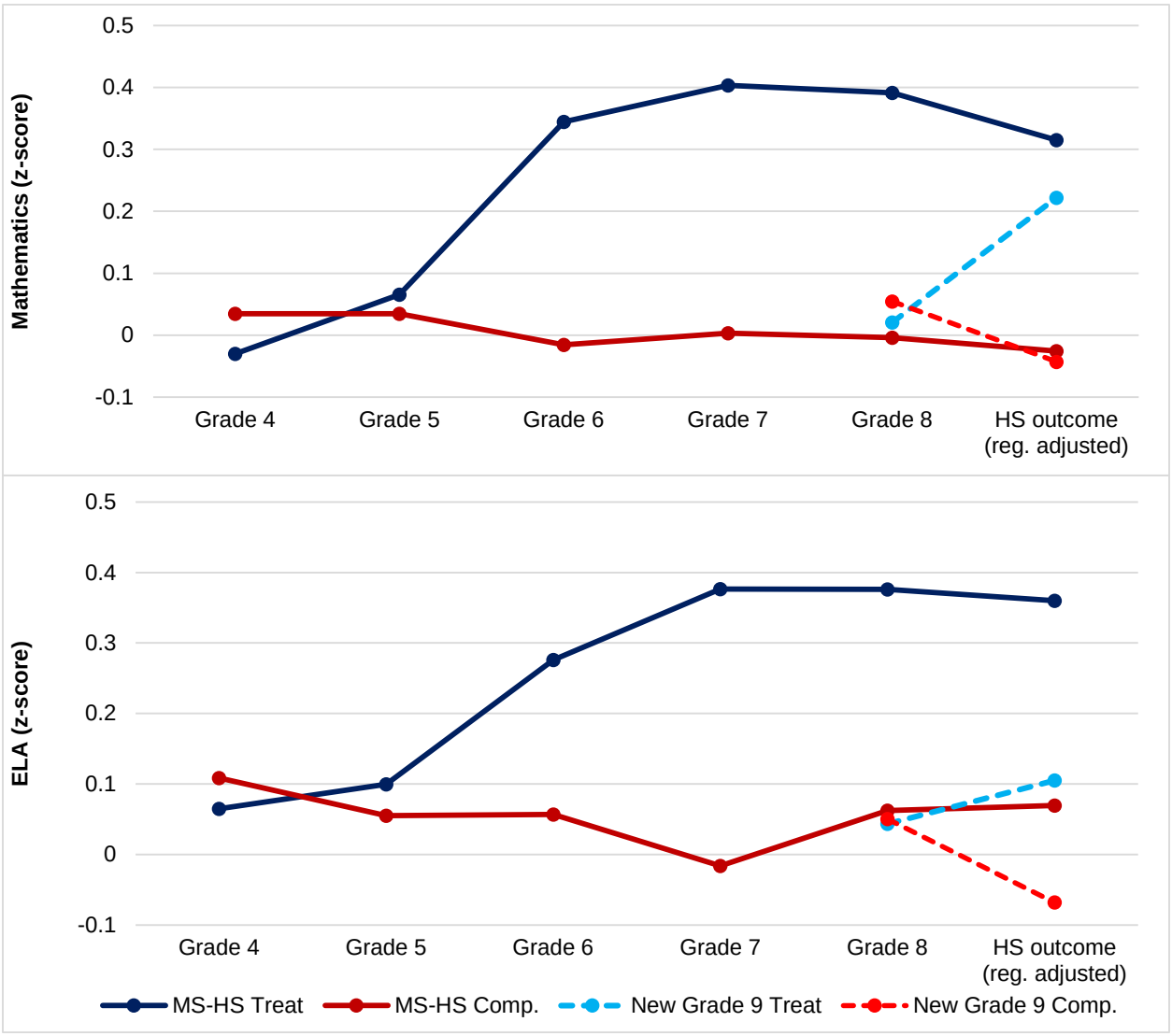
Note: Each circle represents the math and ELA impact estimate for one KIPP site (a linked set of one or more middle schools and one high school). Dark-blue circles indicate that impacts in both subjects are statistically significant and positive at the 0.05 level, two-tailed test. Light-blue circles indicate that the impact in only one of the two test subjects is statistically significant and positive. Grey indicates that both impacts are statistically indistinguishable from zero. No site has a statistically significant and negative impact in either subject. The dashed orange lines represent the average impact in math (vertical line) and ELA (horizontal line).

Compared to the primary high school impact estimates discussed in our main report, we find a pattern of larger impacts on high school test scores when we consider the effect of KIPP middle and high schools combined, relative to students who never attend a KIPP school at any level. The impacts of KIPP high schools on students who are new to the network in grade 9 are slightly smaller than these cumulative impacts, but the effects are still large and statistically significant. In the matched-school design that produced estimates of the marginal impacts of KIPP high schools on students who are continuing from a KIPP middle school, we found that those impacts were not statistically significant. These estimates from the matched-school analysis suggest that attending a KIPP high school neither helps nor harms students who also attended a KIPP middle school in terms of their high school achievement, on average.

To investigate the possible contribution of KIPP middle schools and high schools further, we compared the average achievement levels throughout middle school and high school of middle school entrants and grade 9 entrants to KIPP, alongside achievement levels for the non-KIPP comparison groups in each matched-student impact analysis (Figure C.2). The solid lines in the figure represent KIPP and non-KIPP (matched comparison group) test score outcomes, by grade, for students in the matched-student analysis of cumulative KIPP middle and high school impacts. The dotted lines represent outcomes by grade for students in the matched-student analysis of KIPP high school impacts for new entrants. In each case, matching produced similar

test scores for KIPP and comparison group students in the baseline period before entry into KIPP (grade 4 for the cumulative analysis and grade 8 for the new entrant analysis). The divergence in these estimates after the baseline period reflect possible impacts of KIPP.

Figure C.2. Average test scores by grade level, comparing the MS-HS analytic sample to the analytic sample of new grade 9 entrants to KIPP



Notes: This figure reports the average math and reading (in middle school) or ELA (in high school) z-score in each grade for four different groups of students: (1) the MS-HS analysis treatment group (solid blue); (2) the matched comparison group for the MS-HS analysis (solid red); (3) the treatment group for the analysis of students entering KIPP for the first time in grade 9 (dashed blue); and (4) the matched comparison group for the students entering KIPP in grade 9 (dashed red). The sample is restricted to students with an observed high school test score for each outcome. The HS outcome scores are regression adjusted, such that the difference between a treatment group and its associated matched comparison group equals the study's impact estimate for that analysis. The average scores shown in grade 4 through grade 8 are not regression adjusted.

The patterns in this figure suggest that middle school KIPP entrants and grade 9 new entrants to KIPP both experience large achievement impacts in their first few years after entry into the KIPP network, but for the middle school entrants the gains appear to be concentrated in middle school rather than high school years. In other words, it is possible that the new grade 9 entrants to KIPP could have attributes (such as low middle school test scores) that help make them more receptive to the "achievement benefits" of KIPP high schools than students who arrived from a KIPP middle school and have already been exposed to KIPP for several years. This pattern of results is also consistent with the matched-school design presented in the main text, in which the estimated marginal impact of KIPP high schools among those who also attended KIPP middle schools was not statistically significant.

To test whether our benchmark results are sensitive to our baseline test score imputation strategy described in Appendix F, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores. The results for this smaller sample are nearly identical to our benchmark impact estimates for the cumulative impact of KIPP middle and high schools (Table C.3). There are no statistically significant differences on any baseline measure and the KIPP impact remains positive and statistically significant for all test outcomes and high school graduation. The magnitude of each impact estimate is nearly identical to the benchmark estimate as well.

Table C.3. Baseline equivalence and impact estimates on sample with non-imputed baseline data (cumulative impact of KIPP middle and high School)

	Treatment group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	0.065	2072	0.100	1929	-0.035	0.261
Math scores (math)	-0.032	1516	0.037	1414	-0.069	0.083
Reading scores (social studies)	0.117	788	0.109	707	0.007	0.871
Math scores (science)	0.065	1946	0.133	1636	-0.067	0.078
Reading scores (4-year graduation)	0.037	1025	0.055	1008	-0.018	0.658
Math scores (4-year graduation)	0.019	1025	0.080	1008	-0.061	0.138
Free and reduced-price lunch status (4-year graduation)	0.883	1025	0.892	1008	-0.009	0.562
	Treatment group		**Comparison group**			
Outcome measure	**Adjusted mean**	**Sample size**	**Mean**	**Sample size**	**Impact estimate**	**p-value**
ELA achievement	0.381	2072	0.085	1929	0.295**	0.000
Mathematics achievement	0.341	1516	-0.001	1414	0.341**	0.000
Science achievement	0.417	788	0.001	707	0.417**	0.000
Social studies achievement	0.180	1946	-0.087	1636	0.267**	0.000
Four-year high school graduation	0.785	1025	0.651	1008	0.134**	0.000

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

APPENDIX D

DETAILED ANALYTIC METHODS: ELEMENTARY SCHOOL (LOTTERY-BASED ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP elementary schools. First we present information on the sample and the baseline equivalence of students who won an admission lottery (the treatment group) and those who did not win (the control group). Next we discuss the data and analytic methods used for the analysis. We conclude the appendix by discussing results of sensitivity analyses.

Detail on sample

Of the 23 KIPP elementary schools open in Spring 2011, 8 were sufficiently oversubscribed to be in the analysis and held admissions lotteries in Spring 2011.[4] Elementary school admissions lotteries are held primarily at the pre-kindergarten and kindergarten level, so the analysis sample is drawn from these two entry grades. Among students in the study sample schools and grades, 1,250 were admitted based on a lottery result (Table D.1).[5] The original lottery sample, as defined above, is larger than required to meet the study's targeted level of statistical power. There was a substantial imbalance in the size of the treatment and control groups at the school level for several schools. In other words, a larger number of students participating in the lotteries at these schools were offered admission and included in the treatment group than the number who were not offered admission and included in the control group, or vice versa. To conserve resources while simultaneously maximizing our ability to detect impacts, a subsample of 1,097 students at these schools was randomly selected to comprise the baseline sample for inclusion in the study's data collection (the Woodcock-Johnson test).

Table D.1. Elementary school student sample sizes (lottery-based analysis)

	Treatment	Control	Overall
Original lottery sample	499	751	1,250
Baseline sample	473	624	1,097
Analytic sample (for achievement outcomes)	284	370	654

In total, 654 students who applied to eight KIPP elementary schools were included in the final analytic sample for which we have study-administered test outcomes (284 in the treatment group and 370 in the control group).[6] These students participated in lotteries to enroll in the entry grade at a KIPP elementary school, and this entry grade varied by school: prekindergarten—age 3 (PK3) at three schools and kindergarten at the remaining five schools. Among the students in the final analytic sample, 79 percent of lottery winners ever enrolled in the KIPP school to which they applied, while 6 percent of those who did not win an admissions lottery still ended up attending KIPP at some point during the follow-up period. The 73 percentage-point difference in

[4] A ninth school was sufficiently oversubscribed to be included in the study, but ultimately dropped from our baseline and analytic samples because more than half the sample at that site lacked outcome data.

[5] Students who apply to oversubscribed schools may be guaranteed admission and thus not be eligible for the study. For example, applicants may be admitted to the school outside of the typical lottery process if they have a sibling already attending the school or if a predetermined number of seats are reserved for district residents and these reserved seats are not oversubscribed.

[6] Overall attrition from the sample of students at baseline was 40 percent (443 out of 1,097). The rates of attrition in the treatment and control groups were very similar (within 1 percentage point).

enrollment rates provides a clear contrast between lottery winners and non-winners in exposure to KIPP schools.

Further, the treatment and control groups remained enrolled in different elementary schools as of the second year following the lottery (Table D.2). For applicants at both grade levels, control group students were likely to attend either a traditional public school or a non-KIPP charter school. The rate of private school attendance was 5 percent or less for both groups.

Table D.2. Type of schools attended by KIPP elementary school applicants

	PK-3 entrants		Kindergarten entrants	
Type of school attended	Percentage of lottery winners (treatment)	Percentage of lottery losers (control)	Percentage of lottery winners (treatment)	Percentage of lottery losers (control)
KIPP charter	77	19	69	3
Non-KIPP charter	12	32	17	39
Private	1	5	2	3
Traditional public	10	44	12	54

Notes: Proportions reflect the schools students attended during the 2012–2013 school year, the second year following admissions lotteries, as identified in the parent survey. Among students included in the achievement analysis, 66 percent had nonmissing data on the school they attended. The proportions reported here reflect school attendance among those students. Type of school was determined using the National Center for Education Statistics Common Core of Data.

Properly executed randomization should ensure that there are no differences (observed or unobserved) between the treatment and control groups. In principle, we can test whether or not this is true by examining the baseline characteristics of the treatment and control groups. The baseline characteristics were assessed using data from a baseline survey administered to the parents of students participating in KIPP lotteries close to the time of random assignment.[7] Information from the baseline survey was supplemented with information from the follow-up parent survey where there was missing data from the baseline survey and where we determined the characteristic should not vary systematically with lottery outcome over time (gender, race/ethnicity, age, and an indicator for single-parent household).

We compared the baseline characteristics of the full data collection sample of students in the treatment and comparison groups and found only one significant difference in twenty characteristics (Table D.3), which is what we would have expected due to chance alone.[8] Since

[7] Overall, about 76 percent of parents of students in the elementary school analysis sample completed the baseline survey. Among these students, about 60 percent completed the baseline survey prior to the time of the lottery and the remaining 40 percent completed it after the lottery. For the group that completed the baseline survey after the lottery, there is some risk that the lottery outcome could have influenced their responses.

[8] One complication of this approach is that we do not have full information on the baseline characteristics of the student sample since they are based on surveys that were not completed by all parents. As a result, treatment and control group differences on these characteristics may reflect both actual differences in the two groups' baseline characteristics and also subsequent patterns of non-response on the baseline and follow-up parent surveys. In other words, if losing the lottery made some parents less likely to complete the surveys and the parents who chose not to complete the survey differed systematically from those who did choose to complete it, this could result in

the estimates of KIPP elementary schools' impact are based on the analytic sample of students who completed a WJ-III test, it is also important to examine whether there are any differences in the baseline characteristics of the treatment and control students for whom we have outcome data on the WJ-III. The second panel of Table D.3 compares the baseline characteristics of students for whom we have at least one test score. For three of the four outcome tests, the sample sizes for students with outcomes are similar, differing by fewer than five students, so a single set of baseline treatment-control comparisons is presented. The third panel presents baseline equivalence results for the Calculation test sample, which is substantially smaller since that test was only administered to students in second grade.

Comparisons of baseline characteristics within the analytic sample indicate that there were some baseline differences between treatment and control group students on these measures. In particular, treatment students are less likely to live in a household with income less than $15,000 and more likely to have a mother who completed college, in addition to being less likely to have a mother with less than a high school education (as in the baseline sample). In the Calculation sample, there is no significant difference in family income but the differences in mother's education is of a larger magnitude. These additional baseline differences in the analytic sample may be driven by different processes in the treatment and control group influencing which students take the WJ-III. It is also possible that the observed differences in family income and mother's education are caused in part by missing data on baseline characteristics. However, in order to account for differences in baseline characteristics that might influence students' achievement, impact models statistically control for all of the baseline characteristics listed in Table D.3.

differences in the overall composition of the treatment and comparison groups on these measures. However, we found no evidence of treatment-control differences even accounting for the possibility of both sources of bias.

Table D.3. Baseline equivalence for elementary school samples

Baseline characteristic	Lottery winners	Lottery losers	Difference	Std. Error	Nt	Nc
Baseline Sample						
Female	0.44	0.48	-0.04	0.04	397	488
Average age in years	4.15	4.17	-0.01	0.03	381	465
White, non-hispanic	0.01	0.01	0.00	0.01	421	516
Hispanic (any race)	0.39	0.40	-0.01	0.02	421	516
Black, non-hispanic	0.57	0.54	0.02	0.03	421	516
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.04	0.04	-0.01	0.01	421	516
English is main language spoken at home	0.62	0.60	0.03	0.03	384	473
Another language is main language spoken at home	0.21	0.19	0.02	0.03	384	473
English and another lang. spoken equally at home	0.17	0.21	-0.04	0.03	384	473
One adult in household	0.26	0.32	-0.06	0.03	422	513
Family income less than 15K	0.24	0.26	-0.03	0.03	363	450
Family income between 15K and 25K	0.25	0.24	0.01	0.03	363	450
Family income between 25K and 35K	0.21	0.20	0.01	0.03	363	450
Family income between 35K and 50K	0.16	0.17	0.00	0.03	363	450
Family income 50K or greater	0.14	0.13	0.01	0.02	363	450
Mother's education: less than HS	0.09	0.17	-0.08**	0.03	378	463
Mother's education: HS or GED	0.25	0.27	-0.01	0.03	378	463
Mother's education: some college	0.34	0.32	0.02	0.03	378	463
Mother's education: college	0.32	0.25	0.07	0.04	378	463
Student has access to computer with internet at home	0.77	0.76	0.00	0.03	379	465
Average number of children's books at home	37.52	39.08	-1.56	3.00	359	433
Analytic Sample (with Test Scores)						
Female	0.43	0.46	-0.03	0.05	258	317
Average age in years	4.17	4.20	-0.03	0.03	249	303
White, non-hispanic	0.01	0.01	0.01	0.01	270	337
Hispanic (any race)	0.38	0.42	-0.04	0.03	270	337
Black, non-hispanic	0.56	0.54	0.02	0.03	270	337
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.04	0.03	0.01	0.02	270	337
English is main language spoken at home	0.61	0.59	0.02	0.03	251	307
Another language is main language spoken at home	0.22	0.21	0.00	0.03	251	307
English and another lang. spoken equally at home	0.17	0.20	-0.02	0.03	251	307
One adult in household	0.26	0.29	-0.03	0.04	272	335
Family income less than 15K	0.22	0.29	-0.08*	0.04	238	295
Family income between 15K and 25K	0.24	0.24	0.00	0.04	238	295
Family income between 25K and 35K	0.21	0.17	0.04	0.04	238	295
Family income between 35K and 50K	0.17	0.18	-0.01	0.04	238	295
Family income 50K or greater	0.17	0.12	0.05	0.03	238	295
Mother's education: less than HS	0.08	0.18	-0.10**	0.03	248	302
Mother's education: HS or GED	0.22	0.28	-0.06	0.04	248	302
Mother's education: some college	0.34	0.28	0.06	0.04	248	302
Mother's education: college	0.36	0.25	0.11*	0.05	248	302
Student has access to computer with internet at home	0.77	0.76	0.01	0.04	247	303
Average number of children's books at home	38.99	40.77	-1.78	3.77	231	280

Table D.3 *(continued)*

Baseline characteristic	Lottery winners	Lottery losers	Difference	Std. Error	Nt	Nc
Analytic subsample with calculation score						
Female	0.36	0.39	-0.03	0.06	166	175
Average age in years	5.01	5.03	-0.02	0.04	158	164
White, non-hispanic	0.01	0.01	0.00	0.01	168	179
Hispanic (any race)	0.49	0.46	0.03	0.04	168	179
Black, non-hispanic	0.46	0.51	-0.04	0.04	168	179
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.02	0.01	0.01	168	179
English is main language spoken at home	0.57	0.56	0.01	0.04	159	166
Another language is main language spoken at home	0.24	0.20	0.04	0.04	159	166
English and another lang. spoken equally at home	0.19	0.23	-0.04	0.05	159	166
One adult in household	0.25	0.29	-0.04	0.05	170	179
Family income less than 15K	0.22	0.29	-0.07	0.05	152	160
Family income between 15K and 25K	0.23	0.22	0.01	0.05	152	160
Family income between 25K and 35K	0.21	0.17	0.03	0.06	152	160
Family income between 35K and 50K	0.21	0.20	0.01	0.05	152	160
Family income 50K or greater	0.12	0.11	0.02	0.03	152	160
Mother's education: less than HS	0.09	0.23	-0.14**	0.04	156	163
Mother's education: HS or GED	0.22	0.25	-0.03	0.05	156	163
Mother's education: some college	0.40	0.26	0.14**	0.05	156	163
Mother's education: college	0.30	0.27	0.03	0.06	156	163
Student has access to computer with internet at home	0.81	0.74	0.06	0.05	155	164
Average number of children's books at home	44.56	45.95	-1.39	5.33	145	154

Source: Baseline characteristics are drawn from a baseline survey of parents. Missing values for age, gender, race/ethnicity, and a single parent household indicator from the baseline survey are filled in using information from the follow-up survey, where possible.

Note: All values in this table are based on non-imputed data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in the “Lottery Winner” and “Non-Winner” columns.

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

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

Detail on achievement measures

For the elementary school sample, academic achievement was measured using scores on the WJ-III Letter-Word Identification (Test 1) and Reading Comprehension (Test 9) tests in reading and on the Calculation (Test 5) and Applied Problems (Test 10) tests in math. We administered the WJ-III tests to all students in spring 2014, the third follow-up year, regardless of entry grade or age. Typically students in the PK3 sample would be in kindergarten at this time and students in the kindergarten sample would be in second grade. The Calculation test was administered only to the sample of students who participated in lotteries to enroll in a KIPP kindergarten, since it was not age appropriate for the PK3 sample.

Students' scores on the WJ-III tests were standardized (into z-scores) using information on the performance of a nationally representative norming population. Thus, values reflect students' performance relative to the national population: positive values indicate that sampled students outperformed the average student nationally (in the norming population) and negative values indicate that sample students performed below the national average. The standardized score has been scaled so that a 1 unit change represents 1 standard deviation of the national population. For example, a z-score of 1 for a given student would indicate that the student's score was 1 standard deviation above that of the average student nationally, which would put them at about the 84th percentile.

Detail on analytic methods

Model specification

To obtain estimates of the impact of KIPP admissions for the subset of KIPP elementary schools included in the lottery-based analysis sample, we use the following model:

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

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

Our analysis includes the following student covariates (in X):

- gender,
- student age in years,
- race/ethnicity,

- language spoken at home,
- whether there is only one adult in the household,
- family income,
- mother's education,
- whether the student has access to a computer with internet at home, and
- the number of children's books in the home.

Weighting

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

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

In sites with sibling preference rules, the basic approach to calculating sample weights is the same as in the simple case above.[9] The difference, however, is in the calculation of the probability of admission. No longer can we simply use the number of students offered admission divided by the number of lottery participants. The exact probabilities of admission depend on the

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

number of sets of siblings who participate in the lottery at the school as well as the number of students within each sibling set. With sibling preference rules, each sibling in the lottery has a higher probability of admissions than non-siblings, so the probabilities are adjusted to account for the number of siblings in each affected lottery.

Imputation of baseline characteristics

If there were missing values for the model's covariates, we imputed these values based on other baseline information we collected from the student so that he or she could be included in the sample and contribute to our impact estimates. Our imputation procedure, known as multiple imputation by chained equations, uses non-missing values of baseline covariates to estimate plausible values of baseline characteristics for observations with missing baseline data. In particular, this method first generates multiple datasets with estimated ("imputed") values for missing baseline characteristics. A separate impact estimate is then calculated with each of the imputed datasets. Finally, these impact estimates are combined using procedures described in Rubin (1987) that account for the variability of estimates calculated using the different imputed datasets. The standard error of each combined impact estimate is adjusted to reflect this variability. The imputation procedure and impact estimation using imputed data are conducted using standard commands in Stata and 20 imputations are used. Imputation is conducted separately by treatment group, and all baseline characteristics included as covariates in the impact model are included in the imputation model. Finally, no outcome measures are imputed, only baseline characteristics.

While we use these imputed baseline covariates in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variable were simply treated as being missing from the sample.

Additional analyses

Complier Average Causal Effect (CACE) estimate of the impact of KIPP attendance

For the subset of KIPP elementary schools in which randomized lotteries created viable treatment and control groups, we present two sets of impact estimates: (1) intent-to-treat (ITT) estimates that rely on treatment status as defined by the random lotteries to estimate the impact of being offered admission to a KIPP elementary school and (2) Complier Average Causal Effect (CACE) estimates that represent the impact of attending a KIPP elementary school.[10]

Because families and students choose whether or not to attend KIPP after winning an admissions lottery, and not all lottery winners ultimately attend KIPP, we cannot simply compare outcomes of KIPP attendees and non-attendees to get an unbiased estimate of attending a KIPP elementary school. To generate CACE estimates of the impact of attending a KIPP elementary school, we use the outcome of the lottery for each student as an instrumental variable for KIPP attendance, where attendance is defined as ever having attended a KIPP elementary school. In other words, to obtain CACE estimates we calculate the difference between the outcomes of treatment and control students, adjusting it to reflect the difference between the proportion of

[10] CACE estimates are sometimes referred to as treatment-on-the-treated, or TOT, estimates.

treatment and control students who enroll at KIPP.[11] The CACE model uses the same covariates and weights as the ITT model, and a similar imputation strategy is used, but with a single imputation rather than multiple imputation.

Specifically, we used two-stage least squares to first estimate the effect of winning an admissions lottery on KIPP attendance (Equation D2a), and in the second stage estimated the impact of KIPP attendance on outcomes (Equation D2b). In effect, the CACE approach adjusts the ITT results to account for whether students actually attended a KIPP school.

(D2a) $$attendKIPP_i = \eta + \lambda * T_i + \sum_{k=1}^{K} \theta_k * SCHOOL_{i,k} * T_i + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \gamma * X_i + \upsilon_i$$

(D2b) $$y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * \widehat{attendKIPP_i} + \gamma * X_i + \varepsilon_i$$

As shown in Table D.4 below, the CACE estimates are generally larger than the ITT estimates, as expected. These estimates indicate that, among students whom the admissions lottery caused to attend a KIPP elementary school, KIPP had impacts of 0.41 standard deviations in Calculation, 0.23 standard deviations in Letter and Word Identification, and 0.29 standard deviations in Passage Comprehension, all of which are statistically significant. The estimated impact on Applied Problems is small and not statistically significant, as is the ITT estimate for that outcome. The CACE estimate in Calculation is approximately equivalent to a student moving from the 58th percentile to the 73rd percentile; the Letter and Word Identification estimate is equivalent to moving from the 78th percentile to the 86th percentile; and the Passage Comprehension estimate is equivalent to moving from the 48th percentile to the 60th percentile. (In each of these examples, the starting percentile corresponds with the control group mean score for that outcome.)

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

Outcome (z-score)	ITT	Std. Error	CACE	Std. Error	$F_{instruments}$	Number of instruments	N_t	N_c
WJ Calculation	0.28**	0.11	0.41**	0.15	66.62	5	176	195
WJ Applied Problems	0.07	0.05	0.06	0.07	62.56	8	282	370
WJ Letter-Word Identification	0.25**	0.07	0.34**	0.10	61.38	8	281	370
WJ Passage Comprehension	0.22**	0.07	0.29**	0.09	64.70	8	280	368

Notes: Outcomes are measured on Woodcock-Johnson III, administered in the spring of the third follow-up year. Impacts in the first column of this table are intent-to-treat (ITT), based on regression models that pool all lottery elementary schools and that control for baseline covariates. Impacts in the third column are complier average causal effect (CACE, sometimes referred to as treatment-on-treated or TOT estimate) for each outcome. The same covariate set is used for ITT and CACE models. Standard errors are to the right of each impact estimate. The F-statistic and number of instruments are provided for each CACE estimate to document the instruments' predictive strength. Instruments used for the calculation outcome reflect that only five schools were in that sample. CACE estimates are calculated using a dataset in which missing values of baseline characteristics are imputed using a single imputation rather than multiple imputation.

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

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

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

Sensitivity to weighting and imputation method

The weighting and imputation approaches used in our main elementary school impact estimates are described earlier in this appendix. This section presents evidence on the sensitivity of impact estimates to alternative weighting and imputation approaches. As described above, the normalization of our main sample weights causes each school to contribute to the overall impact estimate in proportion to its sample size. We also calculated impacts using a different normalization approach in which each school contributes equally to the overall impact estimates.[12] As shown in the third column of Table D.5, impacts on Calculation calculated using this approach is substantially smaller than the benchmark estimate and is not statistically significant, in contrast to the benchmark estimate. This difference is driven in large part by a single school with a very small sample size; when this school is omitted from the sample the main impact estimate and alternative estimate differ by less than 0.01 standard deviation and are each statistically significant. Impacts on other test scores estimated using the alternative weighting approach are similar to the main estimates in magnitude and statistical significance.

In order to test the sensitivity of impact estimates to our approach for imputing missing baseline information, we also estimated impacts using mean-imputed baseline characteristics. In this alternative approach, missing baseline characteristics were imputed as the mean among students not missing that characteristic within the same school and treatment group. As shown in the fifth column of Table D.5, there are no substantive differences in the magnitude or statistical significance of these estimates, relative to the main impacts.

Table D.5. ITT Estimates from alternative analysis approaches

Outcome (z-score)	Main ITT Estimate	Std. Error	Schools weighted equally	Std. Error	Mean imputed covariates	Std. Error	N_t	N_c
WJ Calculation	0.28**	0.11	0.18	0.10	0.29**	0.10	176	195
WJ Applied Problems	0.07	0.05	0.04	0.05	0.04	0.05	282	370
WJ Letter-Word Identification	0.25**	0.07	0.22*	0.08	0.22**	0.07	281	370
WJ Passage Comprehension	0.22**	0.07	0.21**	0.07	0.18**	0.07	280	368

Notes: Outcomes are measured on Woodcock-Johnson III, administered in the spring of the third follow-up year. Impacts in the first column of this table are intent-to-treat (ITT), based on regression models that pool all lottery elementary schools and that control for baseline covariates. Impacts in the third column are calculated using a model that weights schools equally, regardless of enrollment size. Impacts in the fifth column use mean imputed data rather than the multiple imputation approach used in the main impact estimates. The same set of baseline characteristics is controlled for in all models. Estimates calculated with schools weighted equally use a single imputed dataset rather than multiple imputed datasets. Standard errors are to the right of each impact estimate.

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

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

[12] This alternative weighting approach does not change the method by which we accounted for differential probabilities of admission within a school for siblings versus non-siblings. Rather, it only affects the weight given to each school when combining each school's impact into an overall cross-site impact.

APPENDIX E

DETAILED ANALYTIC METHODS: MIDDLE SCHOOL (LOTTERY-BASED ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP middle schools. First we present information on the sample and the baseline equivalence of students who won an admission lottery (the treatment group) and those who did not win (the control group). Next we discuss the data and analytic methods used for the analysis. We conclude the appendix by discussing results of sensitivity analyses.

Detail on sample

Of the 59 KIPP middle schools open in Spring 2011, 16 (27 percent) were sufficiently oversubscribed to include in the lottery-based analysis.[13] Middle school admissions lotteries are held primarily at the 5th and 6th grade level, so the analysis sample is drawn from these two entry grades. Among students in the study sample schools and grades, 996 were admitted based on a lottery result (Table E.1).[14] The original lottery sample, as defined above, is larger than required to meet the study's targeted level of statistical power. There was a substantial imbalance in the size of the treatment and control groups at several schools in the baseline sample. In other words, a larger number of students participating in the lotteries at these schools were offered admission and included in the treatment group than the number who were not offered admission and included in the control group, or vice versa. To conserve resources while simultaneously maximizing our ability to detect impacts, a subsample of 857 students at these schools was randomly selected to comprise the baseline sample for inclusion in the study's data collection (follow-up surveys and state test records).[15]

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

	Treatment	Control	Overall
Original lottery sample	530	466	996
Baseline sample	436	421	857
Analytic sample (Year 1)	314	295	609
Analytic sample (Year 2)	291	274	565
Analytic sample (Year 3)	234	225	459

In each of the three outcome years for which state test data were collected, students with test scores available were included in the analysis sample. In total, 609 students were included in the

[13] One oversubscribed school was located in a jurisdiction that did not provide any state test data and is therefore not include in the achievement analysis sample. Of the remaining 43 schools, 19 also conducted lotteries for admission but either exhausted their waitlists or did not provide a sufficiently large treatment or control group for the analysis.

[14] Students who apply to oversubscribed schools may be guaranteed admission and thus not be eligible for the study. For example, applicants may be admitted to the school outside of the typical lottery process if they have a sibling already attending the school or if a predetermined number of seats are reserved for district residents and these reserved seats are not oversubscribed.

[15] The sample of students used in lottery-based estimates of KIPP middle schools' impact on student achievement differs slightly from the sample used when estimating impacts on survey measures because one jurisdiction categorically declined to provide test score data on students whose parent had signed a consent form electronically rather than manually. These students were considered to have been randomly removed from the sample after verifying that there was not a substantial difference in the rate at which lottery winners and losers were excluded on this basis.

year one analysis sample, 565 students were included in the second year's sample and 459 students were included in year three. The main lottery-based impact analysis measures the impact of admission to—rather than attendance at—a KIPP school (in other words, it is an intent-to-treat [ITT] estimate), although admission and attendance are closely related. Among lottery participants, 72 percent of treatment group students and 5 percent of control group students attended a KIPP middle school.

Properly executed randomization should ensure that there are no differences (observed or unobserved) between the treatment and control groups. In principle, we can test whether or not this is true by examining the baseline characteristics of the treatment and control groups. The baseline characteristics were assessed using data from administrative records and from a baseline survey administered to the parents of students participating in KIPP lotteries close to the time of random assignment.[16] Information from the baseline survey was supplemented with information from the follow-up parent survey where there was missing data from the baseline survey on characteristics not expected to vary systematically with lottery outcome over time (for example, gender, race/ethnicity, and date of birth).

We compared the baseline characteristics of the baseline sample of students in the treatment and control groups and found only one difference that was statistically significant at the 5 percent level among the 28 characteristics we examined, which is what we would have expected due to chance alone (Table E.2). Since the lottery-based impact analysis is based on the sample of students with test outcomes in each of three years following the lottery, it is also important to examine whether there are any differences in the baseline characteristics of the treatment and control students for whom we have outcome data in each sample (that is, in the analytic samples).

We examined baseline equivalence separately for each of these analytic samples, and the results provide evidence that the treatment and control groups were similar at baseline (Tables E.3-E.5). In each of the analytic samples—academic achievement in 2012, 2013, and 2014—only one out of 28 measured baseline characteristics shows a statistically significant difference between groups, which is again what would be expected due to chance alone. Given these findings, we are confident that the admissions lotteries were conducted correctly and that the treatment and control groups in each analysis sample are similar in terms of their background characteristics, motivation, and prior educational experiences, aside from the outcome of the lottery itself.

[16] Overall, about 72 percent of parents of students in the middle school analysis sample completed the baseline survey. Among those who completed the baseline survey, about 35 percent completed it prior to the time of the lottery and the remaining 65 percent completed it after the lottery. For the group that completed the baseline survey after the lottery, there is some risk that the lottery outcome could have influenced their responses. However, the test scores are based on administrative data and cover the year (or two years) prior to middle school entry.

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

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std. Error	N_t	N_c
Baseline Reading (z-score)	-0.21	-0.28	0.07	0.07	282	292
Baseline Math (z-score)	-0.12	-0.23	0.10	0.08	280	291
Pre-Baseline Reading (z-score)	-0.21	-0.26	0.05	0.08	270	271
Pre-Baseline Math (z-score)	-0.18	-0.21	0.03	0.08	270	273
Female	0.52	0.52	0.00	0.04	372	363
Age in years	10.63	10.64	-0.01	0.05	300	289
White, non-hispanic	0.02	0.04	-0.02	0.01	382	378
Hispanic or latino	0.53	0.51	0.02	0.03	382	378
Black, non-hispanic	0.43	0.42	0.01	0.03	382	378
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.03	0.00	0.01	382	378
English is main language spoken at home	0.51	0.55	-0.04	0.03	347	331
Another language is main language spoken at home	0.25	0.24	0.01	0.03	347	331
English and another lang. spoken equally at home	0.24	0.21	0.03	0.03	347	331
One adult in household	0.32	0.30	0.01	0.04	340	326
Free/Reduced Price Lunch Eligible	0.84	0.86	-0.02	0.03	358	353
Has an Individualized Education Program (IEP)	0.10	0.10	0.00	0.02	355	355
Family income less than 15K	0.25	0.20	0.05	0.04	293	270
Family income between 15K and 25K	0.25	0.27	-0.02	0.04	293	270
Family income between 25K and 35K	0.17	0.27	-0.10**	0.04	293	270
Family income between 35K and 50K	0.20	0.17	0.03	0.04	293	270
Family income 50K or greater	0.14	0.10	0.04	0.03	293	270
Mother's education: less than HS	0.21	0.21	0.00	0.03	342	325
Mother's education: HS or GED	0.24	0.26	-0.02	0.04	342	325
Mother's education: some college	0.28	0.29	-0.01	0.04	342	325
Mother's education: college	0.28	0.25	0.03	0.03	342	325
Student has access to computer with internet at home	0.78	0.84	-0.06	0.03	307	286
Parent helps student with homework 5 days per week or more	0.73	0.72	0.01	0.04	304	282
Parent discussed college with student over 2x during pre-baseline school year	0.87	0.85	0.02	0.03	304	281

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample randomly assigned by the admissions lottery, then randomly subsampled for data collection. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

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

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std Error	Nt	Nc
Baseline Reading (z-score)	-0.16	-0.24	0.08	0.07	268	264
Baseline Math (z-score)	-0.09	-0.18	0.09	0.08	266	264
Pre-Baseline Reading (z-score)	-0.17	-0.21	0.03	0.08	256	243
Pre-Baseline Math (z-score)	-0.14	-0.16	0.02	0.08	256	244
Female	0.52	0.52	0.00	0.04	311	293
Age in years	10.63	10.64	-0.01	0.05	248	224
White, non-hispanic	0.03	0.04	-0.01	0.01	314	295
Hispanic or latino	0.56	0.54	0.02	0.03	314	295
Black, non-hispanic	0.38	0.39	-0.01	0.03	314	295
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.03	0.03	0.00	0.01	314	295
English is main language spoken at home	0.48	0.52	-0.05	0.03	283	255
Another language is main language spoken at home	0.27	0.25	0.02	0.04	283	255
English and another lang. spoken equally at home	0.25	0.22	0.03	0.04	283	255
One adult in household	0.31	0.27	0.04	0.04	277	250
Free/Reduced Price Lunch Eligible	0.87	0.87	-0.01	0.03	302	284
Has an Individualized Education Program (IEP)	0.09	0.08	0.00	0.03	305	286
Family income less than 15K	0.26	0.21	0.04	0.04	243	210
Family income between 15K and 25K	0.24	0.28	-0.04	0.05	243	210
Family income between 25K and 35K	0.19	0.27	-0.09*	0.04	243	210
Family income between 35K and 50K	0.21	0.14	0.07	0.04	243	210
Family income 50K or greater	0.11	0.10	0.01	0.03	243	210
Mother's education: less than HS	0.22	0.24	-0.02	0.04	279	250
Mother's education: HS or GED	0.24	0.26	-0.02	0.04	279	250
Mother's education: some college	0.29	0.27	0.02	0.04	279	250
Mother's education: college	0.25	0.23	0.02	0.04	279	250
Student has access to computer with internet at home	0.77	0.83	-0.07	0.04	253	222
Parent helps student with homework 5 days per week or more	0.71	0.71	0.01	0.04	250	218
Parent discussed college with student over 2x during pre-baseline school year	0.86	0.83	0.03	0.04	250	218

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data and non-missing baseline data,. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

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

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std Error	Nt	Nc
Baseline Reading (z-score)	-0.13	-0.29	0.16*	0.07	246	243
Baseline Math (z-score)	-0.03	-0.17	0.14	0.08	245	242
Pre-Baseline Reading (z-score)	-0.14	-0.24	0.10	0.08	237	224
Pre-Baseline Math (z-score)	-0.08	-0.15	0.07	0.08	237	226
Female	0.53	0.51	0.02	0.04	288	274
Age in years	10.66	10.65	0.02	0.05	231	211
White, non-hispanic	0.03	0.04	-0.01	0.01	291	274
Hispanic or latino	0.56	0.54	0.02	0.03	291	274
Black, non-hispanic	0.39	0.39	0.00	0.03	291	274
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.02	0.03	-0.01	0.01	291	274
English is main language spoken at home	0.47	0.51	-0.04	0.04	262	238
Another language is main language spoken at home	0.27	0.26	0.01	0.04	262	238
English and another lang. spoken equally at home	0.25	0.22	0.03	0.04	262	238
One adult in household	0.32	0.26	0.06	0.04	257	233
Free/Reduced Price Lunch Eligible	0.87	0.88	-0.01	0.03	279	266
Has an Individualized Education Program (IEP)	0.09	0.10	-0.02	0.03	281	268
Family income less than 15K	0.27	0.21	0.06	0.04	225	198
Family income between 15K and 25K	0.23	0.29	-0.07	0.05	225	198
Family income between 25K and 35K	0.20	0.27	-0.07	0.04	225	198
Family income between 35K and 50K	0.21	0.14	0.07	0.04	225	198
Family income 50K or greater	0.10	0.09	0.01	0.03	225	198
Mother's education: less than HS	0.21	0.24	-0.04	0.04	259	233
Mother's education: HS or GED	0.26	0.25	0.01	0.04	259	233
Mother's education: some college	0.28	0.26	0.02	0.04	259	233
Mother's education: college	0.25	0.24	0.01	0.04	259	233
Student has access to computer with internet at home	0.76	0.82	-0.06	0.04	234	209
Parent helps student with homework 5 days per week or more	0.72	0.70	0.01	0.04	231	206
Parent discussed college with student over 2x during pre-baseline school year	0.85	0.84	0.01	0.04	231	206

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data and non-missing baseline data. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

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

Characteristic	Mean lottery winner	Mean lottery loser	Difference	Std Error	N_t	N_c
Baseline Reading (z-score)	-0.11	-0.26	0.16	0.08	196	195
Baseline Math (z-score)	-0.04	-0.14	0.10	0.09	195	194
Pre-Baseline Reading (z-score)	-0.18	-0.20	0.02	0.09	188	180
Pre-Baseline Math (z-score)	-0.12	-0.09	-0.03	0.10	188	182
Female	0.53	0.51	0.02	0.05	231	224
Age in years	10.63	10.64	-0.01	0.06	182	173
White, non-hispanic	0.03	0.05	-0.02	0.02	234	225
Hispanic or latino	0.51	0.47	0.04	0.04	234	225
Black, non-hispanic	0.44	0.44	-0.01	0.04	234	225
Asian, Pac. Isl., AK Native, Native Amer., or Multi-Race	0.02	0.04	-0.02	0.02	234	225
English is main language spoken at home	0.50	0.56	-0.06	0.04	209	196
Another language is main language spoken at home	0.25	0.26	-0.02	0.04	209	196
English and another lang. spoken equally at home	0.26	0.18	0.07	0.04	209	196
One adult in household	0.35	0.27	0.08	0.05	206	191
Free/Reduced Price Lunch Eligible	0.86	0.86	-0.01	0.03	222	218
Has an Individualized Education Program (IEP)	0.09	0.11	-0.02	0.03	224	219
Family income less than 15K	0.28	0.20	0.08	0.05	177	164
Family income between 15K and 25K	0.24	0.27	-0.03	0.05	177	164
Family income between 25K and 35K	0.18	0.30	-0.12*	0.05	177	164
Family income between 35K and 50K	0.20	0.14	0.06	0.04	177	164
Family income 50K or greater	0.11	0.10	0.01	0.03	177	164
Mother's education: less than HS	0.20	0.23	-0.02	0.04	208	191
Mother's education: HS or GED	0.24	0.27	-0.03	0.05	208	191
Mother's education: some college	0.31	0.28	0.02	0.05	208	191
Mother's education: college	0.25	0.23	0.03	0.04	208	191
Student has access to computer with internet at home	0.78	0.82	-0.05	0.04	186	171
Parent helps student with homework 5 days per week or more	0.72	0.68	0.04	0.05	183	168
Parent discussed college with student over 2x during pre-baseline school year	0.84	0.84	0.01	0.04	183	168

Note: Standard errors reported in parentheses. All values in this table are based on non-imputed data and the sample for which we have state tests outcome data and non-missing baseline data. The difference between lottery winners and losers is based on a regression the baseline characteristic on treatment status and site indicators. The difference is the coefficient on treatment status from that regression. The lottery loser mean is the unadjusted mean for lottery losers. The lottery winner mean is the sum of the lottery loser mean and the regression-adjusted difference between groups. Values are proportions unless otherwise indicated. Totals may not equal difference due to rounding.

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

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

Detail on achievement measures

Academic achievement was measured using scores on state-wide assessments drawn from administrative records collected from states and districts. Students' scores on the state tests were standardized (that is, converted into z-scores) using state-wide means and standard deviations. Thus, values reflect students' performance relative to all tested students in a given state (within grade and subject): positive values indicate that sample students outperformed the average student state-wide and negative values indicate that sample students performed below the state average. This standardization to a common scale allows us to combine outcomes for students in different states. For most students and jurisdictions we were able to collect test score outcomes in the spring of 2012, 2013, and 2014, corresponding to the first, second, and third year after random assignment.[17]

Detail on analytic methods

Model specification

To obtain estimates of the impact of KIPP admissions for the subset of KIPP middle schools included in the lottery-based analysis sample, we use the following model:

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

where *i* and *k* index students and schools, respectively, and *y* is the student-level outcome of interest. *SCHOOL* is a set of binary variables indicating the school that the student applied to and thus the lottery in which the student participated, *T* is a binary treatment status variable indicating whether the student was offered admission to the school via the lottery, and *X* is a set of demographic and other controls, which are listed in Table E.6 below[18]. The βs represent site/lottery fixed effects, which capture differences in outcomes across sites that are not related to KIPP school attendance itself. These effects may capture variation across schools in the characteristics of KIPP applicants and/or the characteristics and performance of non-KIPP schools attended by control students. Including fixed effects in the model (as opposed to random effects) implies that that KIPP schools were selected purposefully for the lottery-based analysis and that the results cannot be generalized beyond the study schools. The parameter δ represents the average impact of winning a KIPP middle school lottery; this is an intent-to-treat (ITT) estimate.

[17] One jurisdiction did not provide test score data for the 2013-14 school year, so the school from that jurisdiction is not included in the last year of test score analysis. Two jurisdictions did not provide their statewide means and standard deviations for the 2014 test administration. The impact estimates presented in this memo for schools in these states from 2014 use z-scores based on 2013 means and standard deviations.

[18] In addition to demographics and baseline test scores, these covariates include school-specific indicators for a handful of students who are a grade ahead of their admissions-lottery peers *before and after* the admissions lottery. These indicators control for systematic differences between the test scores in different grades. Such students are only included in the analysis sample if they have a counterpart in the opposite treatment group at their school. An additional indicator variable for students who took an end-of-course math exam in year three similarly controls for systematic differences between end-of-course and end-of-grade subject tests. As with students ahead of grade, end-of-course exam takers are only included if their school includes such students in the treatment and control group.

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

Math baseline test score from 1 year prior to lottery
Math baseline test score from 2 years prior to lottery
Reading baseline test score from 1 year prior to lottery
Reading baseline test score from 2 years prior to lottery
Gender indicator variable
Age in years at baseline
Set of race/ethnicity indicator variables
Set of indicators for home language: English, Non-English, or bilingual
Indicator for single-parent household
Free or reduced price lunch status indicator variable
Special education status indicator variable
Set of family income indicator variables
Set of mother's education indicator variables
Indicator variable for access to a computer with internet at home
Indicator variable for parent helping with homework five days per week or more
Indicator variable for parent discussing college with student more than two times in the baseline school year
Set of math and reading imputation dummies indicating whether math and reading baseline test scores are imputed
Set of dummy variables indicating whether demographic variables were imputed
School-specific dummy variables indicating students who are ahead of grade before and after the admission lottery, based on the lottery entry grade
Dummy variable for lottery entry grade
Dummy variable indicating that a student took an end of course exam (math, year three only)

Note: Baseline test scores and other covariates were imputed when missing.

Weighting

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

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

normalization factor such that the weights of each experimental group sum to one-half of the total sample size within the site. Thus, the sum of all students' weights within a site will be equal to the overall sample size in that site (that is, the number of consenting lottery participants), with the sum of weights among treatments equal to that among controls.

In sites with sibling preference rules, the basic approach to calculating sample weights is the same as in the simple case above.[19] The difference, however, is in the calculation of the probability of admission. No longer can we simply use the number of students offered admission divided by the number of lottery participants. The exact probabilities of admission depend on the number of sets of siblings who participate in the lottery at the school as well as the number of students within each sibling set. With sibling preference rules, each sibling in the lottery has a higher probability of admissions than non-siblings, so the probabilities are adjusted to account for the number of siblings in each affected lottery.

Imputation of Baseline Characteristics

If there were missing values for the model's covariates, we imputed these values based on other baseline information we collected from the student so that he or she could be included in the sample and contribute to our impact estimates. Our imputation procedure, known as multiple imputation by chained equations, uses non-missing values of baseline covariates to estimate plausible values of baseline characteristics for observations with missing baseline data. In particular, this method first generates multiple datasets with estimated ("imputed") values for missing baseline characteristics. A separate impact estimate is then calculated with each of the imputed datasets. Finally, these impact estimates are combined using procedures described in Rubin (1987) that account for the variability of estimates calculated using the different imputed datasets. The standard error of each combined impact estimate is adjusted to reflect this variability. The imputation procedure and impact estimation using imputed data are conducted using standard commands in Stata and 20 imputations are used. Imputation is conducted separately by treatment group, and all baseline characteristics included as covariates in the impact model are included in the imputation model. Finally, no outcome measures are imputed, only baseline characteristics.

While we use these imputed baseline covariates in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variable were simply treated as being missing from the sample.

Grade Repeaters and Skippers

In our main lottery-based analysis of state test scores, we excluded students who repeated or skipped a grade after the admission lottery because they do not have the same grade progression as their peers and therefore do not have the same pretest-posttest relationship.[20] This strategy is

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

[20] A very small number of students were ahead of the grade progression indicated by their lottery entry grade *both before and after* the lottery. We retained these students in the sample and included their actual (not imputed) test

in contrast to the matching approach that "freezes" grade repeaters in the test score distribution. As discussed later in this appendix, we tested the sensitivity of our impacts on state test scores to this approach by calculating impacts using the alternate approach. As shown below, when we retain grade repeaters in the sample but impute their outcome score, the impact estimates do not change substantively.

Additional analyses

Complier Average Causal Effect (CACE) estimate of the impact of KIPP attendance

For the subset of KIPP middle schools in which randomized lotteries created viable treatment and control groups, we present two sets of impact estimates: (1) intent-to-treat (ITT) estimates that rely on treatment status as defined by the random lotteries to estimate the impact of being offered admission to a KIPP middle school and (2) Complier Average Causal Effect (CACE) estimates that represent the impact of attending a KIPP middle school.[21]

Because families and students choose whether or not to attend KIPP after winning an admissions lottery, and not all lottery winners ultimately attend KIPP, we cannot simply compare outcomes of KIPP attendees and non-attendees to get an unbiased estimate of attending a KIPP middle school. To generate CACE estimates of the impact of attending a KIPP middle school, we use the outcome of the lottery for each student as an instrumental variable for KIPP attendance, where attendance is defined as ever having attended a KIPP middle school. In other words, to obtain CACE estimates we calculate the difference between the outcomes of treatment and control students, adjusting it to reflect the difference between the proportion of treatment and control students who enroll at KIPP.[22] The CACE model uses the same covariates and weights as the ITT model, and a similar imputation strategy is used, but with a single imputation rather than multiple imputation. Specifically, we used two-stage least squares to first estimate the effect of winning an admissions lottery on KIPP attendance (Equation E2a), and in the second stage estimated the impact of KIPP attendance on outcomes (Equation E2b). In effect, the CACE approach adjusts the ITT results to account for whether students actually attended a KIPP school.

(E2a) $$attendKIPP_i = \eta + \lambda * T_i + \sum_{k=1}^{K} \theta_k * SCHOOL_{i,k} * T_i + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \gamma * X_i + \upsilon_i$$

(E2b) $$y_i = \alpha + \sum_{k=1}^{K} \beta_k * SCHOOL_{i,k} + \delta * \widehat{attendKIPP_i} + \gamma * X_i + \varepsilon_i$$

As shown in Table E.7 below, the CACE estimates are larger than the ITT estimates, as expected. These indicate that in the second year, for example, KIPP had impacts of 0.32 standard

scores in the impact model, including an indicator variable to control for any systematic differences in their test scores, relative to peers in other grades. We only retained such students in the sample if they had a counterpart in the opposite treatment group at the same school. That is, if at a given school all students in this situation were in the treatment group and none were in the control group, then they were not included in the sample.

[21] CACE estimates are sometimes referred to as treatment-on-the-treated, or TOT, estimates.

[22] Control students may end up enrolling at KIPP if they are offered admission after the October cut-off date for assignment to the treatment group (for example, during the second semester), if they apply and are offered admission for the following school year, or in rare cases when they are offered admission out of order off the waitlist. For each outcome year, attendance is defined for each student as having ever attended a KIPP school between the lottery and that school year.

deviations in Math and 0.27 standard deviations in reading among students whom the admissions lottery caused to attend a KIPP middle school. The CACE estimate in math in year two is approximately equivalent to a student moving from the 40th percentile to the 53rd percentile; the CACE estimate in reading in the same year is equivalent to a student moving from the 37th percentile to the 47th percentile. (In each of these examples, the starting percentile corresponds with the control group mean score for that outcome.)

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

Outcome (z-score)	ITT	Std. Error	CACE	Std. Error	$F_{instruments}$	Number of instruments	N_t	N_c
Math								
Year 1	0.10*	0.05	0.12	0.06	73.3	16	313	294
Year 2	0.24**	0.06	0.32**	0.08	64.7	16	287	268
Year 3	0.18**	0.07	0.25**	0.09	56.6	14	233	222
Reading								
Year 1	0.03	0.05	0.05	0.06	75.1	16	314	294
Year 2	0.18**	0.05	0.27**	0.07	66.5	16	291	272
Year 3	0.14*	0.06	0.16	0.08	56.9	14	234	224

Source: State and district administrative records data, study-administered survey data (baseline characteristics)

Note: Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Impact estimates in the third column measure the impact of an offer of admission to a KIPP middle school (the ITT estimate) and are based on regression models that pool all lottery schools and control for baseline covariates. Impact estimates in the fifth column are complier average causal effect (CACE, sometimes referred to as treatment-on-treated or TOT estimate) for each outcome. The same covariate set is used for ITT and CACE models. Standard errors are to the right of each impact estimate. The F-statistic and number of instruments are provided for each CACE estimate to document the instruments' predictive strength. The number of instruments for year 3 estimates reflects that two grades of one school were excluded from this outcome sample because they did not provide test scores for that year. CACE estimates are calculated using a dataset in which missing values of baseline characteristics are imputed using a single imputation rather than multiple imputation.

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

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

Sensitivity to weighting

The weighting and imputation approaches used in our main middle school impact estimates are described earlier in this appendix. This section presents evidence on the sensitivity of impact estimates to alternative weighting and imputation approaches. As described above, the normalization of our main sample weights causes each school to contribute to the overall impact estimate in proportion to its sample size. We also calculated impacts using a different normalization approach in which each school contributes equally to the overall impact estimates.[23] As shown in the third column of Table E.8, impacts calculated using the alternate approach generally have a magnitude similar to the benchmark model, with the exception of the reading impact in year three. In the third year, sites with smaller sample sizes tended to have more lower (or negative impacts); because the alternative approach gives small sites more weight

[23] This alternative weighting approach does not change the method by which we accounted for differential probabilities of admission within a school for siblings versus non-siblings. Rather, it only affects the weight given to each school when combining each school's impact into an overall cross-site impact.

than the main impact estimates do, these small sites contributed to a lower overall impact estimate.

Table E.8. ITT estimates using alternative methods

Outcome (z-score)	Main impact	Std. error	Impact weighting schools equally	Std. Error	N_t	N_c	Impact including repeaters, skippers	Std. Error	N_t	N_c
Math										
Year 1	0.10	0.05	0.06	0.05	313	294	0.10*	0.04	315	300
Year 2	0.24**	0.06	0.19**	0.06	287	268	0.25**	0.06	303	278
Year 3	0.18*	0.07	0.15*	0.06	233	222	0.20**	0.06	251	236
Reading										
Year 1	0.03	0.05	0.01	0.05	314	294	0.02	0.04	316	300
Year 2	0.18**	0.05	0.17**	0.06	291	272	0.18**	0.05	306	278
Year 3	0.14*	0.06	0.03	0.06	234	224	0.13*	0.05	252	237

Source: State and district administrative records data, study-administered survey data (baseline characteristics)

Note: Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Estimates measure the impact of an offer of admission to a KIPP middle school (the ITT estimate) and are based on regression models that pool all lottery schools and control for baseline covariates. Impact estimates in the first column are those presented in the main report body, which combine school-level estimates weighting each school's estimate by its sample size. Impacts in the third column weight schools equally, regardless of their sample size. Impacts in the seventh column include students who were excluded from other impact models based on having repeated or skipped grades only following the lottery. The alternative models use a single imputed dataset to account for missing baseline data, while the main impact model uses multiple imputation, as described earlier in this appendix. All models control for the same set of baseline characteristics. Standard errors are to the right of each impact estimate.

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

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

Sensitivity to including grade repeaters and skippers

As described above, in our main lottery-based analyses of KIPP middle school impacts, we exclude students who repeated or skipped a grade following the admissions lottery. To test whether impact estimates are sensitive to the approach used to address students these students, we also calculate impacts including these students. In this sensitivity analysis, students' test scores are imputed in repeated or advanced grades: the last score before the student skipped a grade or was held back is used as the outcome measure for that student in every year after starting with the first skipped or repeated grade. As shown in Table E.8 above, the impacts calculated using this approach are generally slightly larger, and the year one math impact is statistically significant in this analysis but not in the main analysis. However, overall, the magnitudes and patterns of impacts are similar.

APPENDIX F

DETAILED ANALYTIC METHODS: MIDDLE SCHOOL (MATCHED-STUDENT ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP middle schools. First we present information on the sample and the baseline equivalence of the treatment and matched comparison group for each outcome. Next we discuss the data used in the analysis, and we conclude the appendix with a discussion of the analytic methods used for the analysis.

Detail on sample

Overall, the middle school analysis included 37 KIPP middle schools. Seven of these middle schools opened in fall 2011, during the period of KIPP network expansion facilitated by i3 scale-up grant issued to KIPP by the federal Department of Education. The total sample of treatment and matched comparison students for each included school is shown in Table F.1, along with the number of cohorts and years of outcome data included.

A key goal for the matching analysis was to include as many schools as possible so that our estimated impacts would be based on as large a portion of the KIPP network as possible. Two criteria were used to select KIPP middle schools for the analysis. First, all included schools had to be established in the 2012–13 school year or earlier to ensure that a minimum of two cohorts of students per school would be observed by spring of 2014.[24] Second, the schools had to be located in jurisdictions (states or school districts) that provided at least three consecutive years of complete, longitudinally linked student-level data for both traditional public and charter schools. These data were needed to track individual KIPP and non-KIPP students in the years prior to middle school enrollment, as well as during the middle school. Throughout this report, we use the term "baseline year" to refer to the school year that began one year prior to when a cohort of students first entered middle school at KIPP; the term "pre-baseline year" refers to the point two years before middle school entry.

The middle school sample includes a combination of schools that were included in our previous report on KIPP middle school impacts (Tuttle et al. 2013) and newer schools that could not be included in that report. In addition, some middle schools that were included in the previous report could not be included here because we could not obtain data for them. In the current study, 25 of the 37 schools were also included in the analyses for the 2013 report.[25] The 37 schools in our sample represent more than half of the KIPP middle school network as of the 2012-2013 school year. Our sample also includes 7 of the 10 KIPP middle schools that opened in fall 2011 or fall 2012.

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

[25] For our current study we did not receive district-wide data from Philadelphia (which only provided data on the RCT sample for this study), New Orleans, Los Angeles, Indiana, Oklahoma, and the San Francisco Bay Area (all of these jurisdictions were present in the 2013 study).

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

State	School	Region	Analytic baseline sample			Total Number of KIPP cohorts through 2013-14	Number of KIPP cohorts in data (school years)
			Treatment (N)	Comp. (N)	Total sample size		
AR	KIPP Blytheville College Prep. School	KIPP Delta	169	169	338	4	4 (2010-11 to 2013-14)
	KIPP Delta College Preparatory School	KIPP Delta	601	601	1,202	12	9 (2005-06 to 2013-14)
CO	KIPP Montbello College Prep*	KIPP Denver	322	322	644	3	3 (2011-12 to 2013-14)
	KIPP Sunshine Peak Academy	KIPP Denver	1,134	1,134	2,268	12	11 (2004-05 to 2013-14)
DC	KIPP DC: AIM Academy	KIPP DC	707	707	1,414	9	9 (2005-06 to 2013-14)
	KIPP DC: KEY Academy	KIPP DC	884	884	1,768	13	11 (2003-04 to 2013-14)
	KIPP DC: WILL Academy	KIPP DC	632	632	1,264	8	8 (2006-07 to 2013-14)
GA	KIPP STRIVE Academy	KIPP Atlanta	396	396	792	5	5 (2009-10 to 2013-14)
	KIPP Vision Academy	KIPP Atlanta	377	377	754	4	4 (2010-11 to 2013-14)
	KIPP WAYS Academy	KIPP Atlanta	794	794	1,588	11	10 (2004-05 to 2013-14)
MA	KIPP Academy Boston Middle School*	KIPP Massachusetts	143	143	286	2	2 (2012-13 to 2013-14)
	KIPP Academy Lynn Middle School	KIPP Massachusetts	892	892	1,784	10	10 (2004-05 to 2013-14)
NY	KIPP Academy Middle School (New York)	KIPP NYC	480	480	960	19	10 (2004-05 to 2013-14)
	KIPP AMP Middle School	KIPP NYC	595	595	1,190	9	10 (2005-06 to 2013-14)
	KIPP STAR Harlem Middle School	KIPP NYC	606	606	1,212	11	10 (2004-05 to 2013-14)
	KIPP Infinity Middle School	KIPP NYC	568	568	1,136	9	10 (2005-06 to 2013-14)
	KIPP Washington Heights Middle School*	KIPP NYC	180	180	360	2	2 (2012-13 to 2013-14)
NC	KIPP Charlotte	KIPP Charlotte	335	335	670	7	7 (2007-08 to 2013-14)
	KIPP Gaston College Preparatory	KIPP Eastern NC	931	931	1,862	13	13 (2001-02 to 2013-14)
TN	KIPP Memphis Academy Middle*	KIPP Memphis	210	210	420	2	2 (2012-13 to 2013-14)
	KIPP Memphis Collegiate Middle	KIPP Memphis	1,113	1113	2,226	12	12 (2002-03 to 2013-14)

Table F.1 *(continued)*

State	School	Region	Analytic baseline sample			Total Number of KIPP cohorts through 2013-14	Number of KIPP cohorts in data (school years)
			Treatment (N)	Comp. (N)	Total sample size		
TX	KIPP Austin Academy of Arts & Letters	KIPP Austin	371	371	742	5	4 (2009-10 to 2013-14)[a]
	KIPP Austin Beacon Prep*	KIPP Austin	180	180	360	2	2 (2012-13 to 2013-14)
	KIPP Austin College Prep	KIPP Austin	816	816	1,632	12	10 (2003-04 to 2013-14)[a]
	KIPP Austin Vista Middle School*	KIPP Austin	170	170	340	2	2 (2012-13 to 2013-14)
	KIPP TRUTH Academy	KIPP Dallas-Fort Worth	616	616	1,232	11	10 (2003-04 to 2013-14)[a]
	KIPP Aspire Academy	KIPP San Antonio	834	834	1,668	11	10 (2003-04 to 2013-14)[a]
	KIPP Camino Academy	KIPP San Antonio	362	362	724	4	4 (2010-11 to 2013-14)
	KIPP 3D Academy	KIPP Houston	948	948	1,896	13	11 (2004-05 to 2013-14)
	KIPP Academy (Houston)	KIPP Houston	692	692	1,384	19	10 (2003-04 to 2013-14)[a]
	KIPP Courage College Prep*	KIPP Houston	160	160	320	2	2 (2012-13 to 2013-14)
	KIPP Intrepid Preparatory School	KIPP Houston	516	516	1,032	6	6 (2008-09 to 2013-14)
	KIPP Liberation College Prep	KIPP Houston	531	531	1,062	8	8 (2006-07 to 2013-14)
	KIPP Polaris Academy for Boys	KIPP Houston	518	518	1,036	7	7 (2007-08 to 2013-14)
	KIPP Sharpstown College Prep	KIPP Houston	454	454	908	7	6 (2007-08 to 2013-14)[a]
	KIPP Spirit College Prep	KIPP Houston	682	682	1,364	8	8 (2006-07 to 2013-14)
	KIPP Voyage Academy for Girls	KIPP Houston	393	393	786	5	5 (2009-10 to 2013-14)
	Total (schools opened prior to 2011)	**30 schools**	**18,947**	**18,947**	**37,894**		
	Total (schools opened 2011 or later)	**7 schools**	**1,365**	**1,365**	**2,730**		
	Total (all schools)	**37 schools**	**20,312**	**20,312**	**40,624**		

Notes: Test outcomes are drawn from administrative records for each of the first four years following enrollment in middle school. Treatment students are KIPP middle school students who never attended a KIPP elementary school, and are matched with comparison students based on baseline (grade 4) characteristics. Starred (*) schools opened in fall 2011 or later. Data was either provided by states or individual school districts. In each school year, data files included the following student-level variables: school of enrollment; indicators for gender, race/ethnicity, special education status, free or reduced price lunch status, and limited English proficiency status (except in TX); and test scores in reading, math, and science (except for New York City). History test scores were provided in Atlanta, Texas, and Memphis.

[a] Data does not include cohort from the 2010-2011 school year

While the study's matching procedures (described below) ensured that the treatment group (KIPP students) and comparison group (non-KIPP students) were equivalent at baseline, we did not observe outcome measures for every matched student in every outcome sample. To check that treatment and comparison students included in each of our analysis samples (that is, with valid data on a particular outcome) are equivalent on observable characteristics, we examined baseline equivalence separately for the study's middle school outcomes: math and reading exams one through four years after enrollment at KIPP, and middle school social studies and science exams.[26] We looked at nine baseline characteristics including baseline reading and math test scores; gender, race, special education, limited English proficiency, and free- or reduced price lunch status; and whether the student repeated a grade in the baseline year. The following tables (Tables F.2 through F.6) show the baseline equivalence of each outcome's analytic sample.

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

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 1 outcome						
Reading scores (z-score)	-0.104	-0.087	-0.017	0.011	17,668	17,562
Math scores (z-score)	-0.097	-0.083	-0.013	0.011	17,634	17,502
Student is male	0.489	0.495	-0.006	0.006	18,460	18,314
Student is black	0.512	0.515	-0.003	0.004	18,460	18,314
Student is Hispanic	0.465	0.459	0.006	0.004	18,460	18,314
Special Education	0.066	0.069	-0.003	0.003	18,460	18,314
Limited English Proficiency	0.102	0.108	-0.007	0.005	11,615	11,348
Free- or reduced-price lunch	0.889	0.882	0.007*	0.004	18,460	18,314
Grade Repeaters	0.011	0.013	-0.002	0.001	18,460	18,314
Sample for math year 1 outcome						
Reading scores (z-score)	-0.104	-0.088	-0.016	0.011	17,668	17,564
Math scores (z-score)	-0.097	-0.085	-0.012	0.011	17,646	17,538
Student is male	0.489	0.497	-0.007	0.006	18,461	18,337
Student is black	0.512	0.515	-0.003	0.004	18,461	18,337
Student is Hispanic	0.465	0.459	0.005	0.004	18,461	18,337
Special Education	0.066	0.070	-0.004	0.003	18,461	18,337
Limited English Proficiency	0.102	0.109	-0.008	0.005	11,627	11,370
Free- or reduced-price lunch	0.889	0.882	0.008*	0.004	18,461	18,337
Grade Repeaters	0.011	0.013	-0.002	0.001	18,461	18,337

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 1 year after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator.

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

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

[26] While math and reading exams are typically given annually in middle school, science and social studies are administered once, twice, or not at all, and in different grade levels depending on the jurisdiction. For science and social studies, we used the outcome test from the highest middle school grade in each jurisdiction where the test could be observed for more than one cohort of KIPP students in our sample.

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

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 2 outcome						
Reading scores (z-score)	-0.089	-0.089	0.000	0.014	14,210	13,816
Math scores (z-score)	-0.072	-0.082	0.010	0.015	14,156	13,782
Student is male	0.492	0.498	-0.007	0.008	14,855	14,531
Student is black	0.516	0.518	-0.002	0.005	14,855	14,531
Student is Hispanic	0.461	0.457	0.004	0.005	14,855	14,531
Special Education	0.065	0.070	-0.005	0.004	14,855	14,531
Limited English Proficiency	0.106	0.117	-0.011	0.006	9,479	8,752
Free- or reduced-price lunch	0.886	0.882	0.003	0.004	14,855	14,531
Grade Repeaters	0.011	0.012	-0.002	0.002	14,855	14,531
Sample for math year 2 outcome						
Reading scores (z-score)	-0.089	-0.092	0.003	0.014	14,193	13,798
Math scores (z-score)	-0.072	-0.085	0.012	0.015	14,149	13,782
Student is male	0.492	0.500	-0.008	0.008	14,848	14,523
Student is black	0.516	0.519	-0.003	0.005	14,848	14,523
Student is Hispanic	0.461	0.456	0.005	0.005	14,848	14,523
Special Education	0.065	0.071	-0.006	0.004	14,848	14,523
Limited English Proficiency	0.105	0.117	-0.011	0.006	9,477	8,747
Free- or reduced-price lunch	0.885	0.882	0.003	0.004	14,848	14,523
Grade Repeaters	0.011	0.012	-0.002	0.002	14,848	14,523

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 2 years after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator. *Significantly different from zero at the .05 level, two-tailed test.

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

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

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 3 outcome						
Reading scores (z-score)	-0.050	-0.048	-0.002	0.014	11,424	10,959
Math scores (z-score)	-0.042	-0.055	0.013	0.015	11,373	10,916
Student is male	0.485	0.499	-0.014	0.008	11,928	11,505
Student is black	0.563	0.556	0.007	0.005	11,928	11,505
Student is Hispanic	0.411	0.413	-0.002	0.005	11,928	11,505
Special Education	0.064	0.063	0.001	0.004	11,928	11,505
Limited English Proficiency	0.097	0.102	-0.005	0.006	7,585	6,934
Free- or reduced-price lunch	0.872	0.871	0.001	0.005	11,928	11,505
Grade Repeaters	0.013	0.013	0.000	0.002	11,928	11,505
Sample for math year 3 outcome						
Reading scores (z-score)	-0.052	-0.056	0.003	0.014	11,291	10,855
Math scores (z-score)	-0.045	-0.065	0.020	0.015	11,245	10,821
Student is male	0.484	0.499	-0.015	0.008	11,796	11,388
Student is black	0.564	0.556	0.008*	0.005	11,796	11,388
Student is Hispanic	0.410	0.413	-0.002	0.005	11,796	11,388
Special Education	0.064	0.063	0.001	0.004	11,796	11,388
Limited English Proficiency	0.097	0.102	-0.005	0.006	7,575	6,912
Free- or reduced-price lunch	0.871	0.870	0.001	0.005	11,796	11,388
Grade Repeaters	0.013	0.013	0.001	0.002	11,796	11,388

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 3 years after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator.

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

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

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

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for reading year 4 outcome						
Reading scores (z-score)	-0.009	-0.039	0.029	0.019	7,523	7,210
Math scores (z-score)	-0.022	-0.029	0.006	0.020	7,458	7,159
Student is male	0.482	0.496	-0.015	0.010	7,863	7,528
Student is black	0.587	0.571	0.017**	0.006	7,863	7,528
Student is Hispanic	0.392	0.399	-0.007	0.005	7,863	7,528
Special Education	0.062	0.061	0.001	0.005	7,863	7,528
Limited English Proficiency	0.072	0.079	-0.007	0.005	5,058	4,604
Free- or reduced-price lunch	0.853	0.858	-0.005	0.007	7,863	7,528
Grade Repeaters	0.013	0.014	-0.002	0.002	7,863	7,528
Sample for math year 4 outcome						
Reading scores (z-score)	-0.059	-0.084	0.025	0.020	6,977	6,822
Math scores (z-score)	-0.067	-0.086	0.018	0.021	6,912	6,776
Student is male	0.490	0.497	-0.007	0.011	7,290	7,135
Student is black	0.595	0.575	0.020**	0.006	7,290	7,135
Student is Hispanic	0.385	0.395	-0.011	0.006	7,290	7,135
Special Education	0.062	0.061	0.001	0.005	7,290	7,135
Limited English Proficiency	0.072	0.079	-0.006	0.005	5,010	4,552
Free- or reduced-price lunch	0.856	0.860	-0.004	0.007	7,290	7,135
Grade Repeaters	0.013	0.015	-0.002	0.002	7,290	7,135

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have reading or math test outcome data 4 years after enrollment at KIPP. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator. *Significantly different from zero at the .05 level, two-tailed test.

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

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

Baseline characteristic	KIPP	Comparison	Difference	Std. Error	# KIPP	# Comparison
Sample for social studies outcome						
Reading scores (z-score)	0.020	-0.019	0.040	0.022	4,892	4,938
Math scores (z-score)	0.016	-0.025	0.041	0.022	4,895	4,937
Student is male	0.477	0.491	-0.014	0.011	5,191	5,249
Student is black	0.513	0.510	0.003	0.006	5,191	5,249
Student is Hispanic	0.484	0.483	0.002	0.006	5,191	5,249
Special Education	0.038	0.043	-0.005	0.005	5,191	5,249
Limited English Proficiency	0.003	0.006	-0.003	0.003	1,670	1,460
Free- or reduced-price lunch	0.892	0.888	0.004	0.007	5,191	5,249
Grade Repeaters	0.014	0.013	0.001	0.003	5,191	5,249
Sample for science outcome						
Reading scores (z-score)	-0.068	-0.075	0.007	0.015	8,928	8,700
Math scores (z-score)	-0.055	-0.063	0.009	0.016	8,886	8,652
Student is male	0.479	0.493	-0.014	0.008	9,329	9,104
Student is black	0.529	0.528	0.002	0.005	9,329	9,104
Student is Hispanic	0.447	0.445	0.002	0.005	9,329	9,104
Special Education	0.058	0.059	-0.002	0.004	9,329	9,104
Limited English Proficiency	0.089	0.105	-0.015**	0.006	5,136	4,658
Free- or reduced-price lunch	0.875	0.878	-0.003	0.005	9,329	9,104
Grade Repeaters	0.011	0.011	0.000	0.002	9,329	9,104

Source: Test outcomes are drawn from state or district administrative records.

Note: All values in this table are based on non-imputed data and the sample for which we have social studies or science test outcome data. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each middle school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator. *Significantly different from zero at the .05 level, two-tailed test.

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

Out of the 10 outcome samples in our middle school analysis, none of them had a statistically significant difference on baseline math or reading test scores. With respect to the baseline demographic characteristics we examined, four show zero statistically significant differences between the treatment and control group and six show only one significant difference on a demographic characteristic. Where we did find a statistically significant difference on a demographic indicator, the magnitude was small: for all of the baseline demographic attributes we examined, the treatment group is within two percentage points of the control group in each outcome sample.

Detail on data

For the matched comparison group analysis, we used de-identified, longitudinally-linked student- data from jurisdictions (states or districts) hosting at least one KIPP school and able to provide student-level records at the time of data collection. The variables from jurisdictions' administrative data systems included: test scores in reading, mathematics, social studies, and science (where middle school scores represent the primary outcome and elementary school scores a key matching variable and baseline covariate); demographic characteristics, used for

matching and as baseline covariates; and schools attended and dates of enrollment, identifying students' exposure to KIPP. Within each jurisdiction, we requested data for all school years beginning with the year prior to the KIPP middle school's first year (to capture baseline data) through the 2013–14 school year. We obtained data from districts for 13 of the 37 schools in the analysis; for the other 24 schools, we obtained records from the state in which the school was located but limited our data to the district (or districts) from which the KIPP school drew students.

To make the analysis of state test scores comparable across states and districts, all raw test scores were converted to z-scores defined relative to the distribution of scores in each grade, year, subject, and jurisdiction. That is, for each jurisdiction associated with a given KIPP school, we calculated the difference between each student's raw score and the mean score in that grade, year, and subject, and then divided the difference by the standard deviation of raw scores in the jurisdiction in that grade, year, and subject. Thus, each z-score reflects a student's achievement level relative to the average student in the relevant cohort and jurisdiction (in terms of the number of standard deviations above or below the mean).[27]

For a variety of reasons, some students may not have valid data in the year when a given outcome was measured. For example, some students may transfer to a jurisdiction outside of our data catchment area, while others may transfer to local private schools or drop out of school altogether. In a small number of cases, students may simply have missing variable values in a given year or subject. We categorize these cases when students disappear from the analytic sample as out-of-district transfers. If KIPP students transfer out-of-district at a different rate than matched comparison students, it could undermine the validity of impact estimates. As noted above, we checked this by examining the baseline equivalence of the sample for each of our 10 outcome measures. There were no differences on baseline reading or math scores on any of the outcome samples and no outcome sample had more than one significant difference on a demographic characteristic. All variables assessed for baseline equivalence are also controlled for in our impact regression model.

Different analytic sample attrition might occur when students are missing one or more baseline or pre-baseline test scores. To address this we imputed missing baseline data, ensuring that all students with at least one recorded baseline test score remain in the sample. For a detailed discussion of our imputation methods, see the following section of this appendix.

Detail on analytic methods

This study relied on a matched comparison group design that used "nearest neighbor" matching to identify a similar comparison student for each treatment student entering a KIPP middle school in grade 5 or grade 6. The validity of our matched comparison group design depends on the ability to eliminate or minimize differences in key characteristics between students who enter KIPP and students in the comparison group who remain in non-KIPP public

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

schools.[28] Our approach achieved this in two ways. First, we used student-level data that included a rich set of student characteristics and multiple years of baseline (prior to KIPP entry) test scores. We used this information to identify a matched comparison group of students who are similar to KIPP students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in elementary school. By matching on more than one year of baseline test score data, we accounted for achievement levels at the time when students applied to KIPP schools as well as pre-KIPP trends in student achievement. After we identified the matched comparison group, the second feature of our approach estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between KIPP students and comparison students. Specifically, the impact estimates adjust for any differences between KIPP students and the matched comparison group pertaining to demographic characteristics or students' prior two years of math and reading test scores.

The combination of propensity-score matching and OLS accounts for differences in observed baseline characteristics and achievement scores between KIPP students and comparison students (in other words, the differences associated with initial selection into KIPP schools). But it remains possible that KIPP students and comparison students differ in unobserved ways that may affect later test scores. However, previous studies have suggested that applying a combination of propensity-score matching and OLS, as we did here, can succeed in replicating experimental impact estimates in certain contexts (Cook et al. 2008; Bifulco 2012; Furgeson et al. 2012; Tuttle et al. 2013; Fortson et al. 2015). We used the same analytic approach for the propensity score matching model as we implemented in our previous report on KIPP middle schools (Tuttle et al. 2013). As part of that report, we also ran a variety of sensitivity tests to check the robustness of our model to alternatives to our main specifications.[29]

There are several other threats to the validity of these impact estimates that we addressed, including students moving from KIPP middle schools to other district schools (attrition from KIPP schools), students who are retained in grade, and attrition from the sample.

Attrition from KIPP Schools. The fact that some students depart KIPP schools and return to non-KIPP schools in the surrounding district before the end of 8th grade could potentially introduce selection bias if not appropriately handled. At both KIPP and district schools, students who transfer before the end of middle school tend to be those who are not doing as well academically as those who remain (Nichols-Barrer et al. 2012). In this way, an analysis that only includes persistently enrolled KIPP students in the treatment group would positively bias the estimated impact of KIPP schools (that is, make KIPP impacts look more positive than they actually are). We addressed this problem by permanently assigning to the treatment group any student who can be found in the records as ever enrolling at KIPP in grades 5 or 6, regardless of whether the student remained in a KIPP school or transferred elsewhere before the end of middle

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

[29] Sensitivity tests included running models that: excluded pre-baseline test scores, included dummies for 4th grade schools, weighted schools by sample size, matched with replacement, used caliper matching, only included students currently enrolled in KIPP in the treatment group, only included students with non-imputed baseline data, and a variety of alternative specifications for grade repeaters. The overall average impact estimates are not sensitive to any of these changes in specification.

school.[30] In other words, a student who enrolled at KIPP in 5th grade for the 2009–10 school year but left KIPP after completing 6th grade in the 2010–11 school year is included in the treatment group for all four years he or she appears in the data (from 2009–10 to 2012–13, inclusive). By including all students observed attending a KIPP school, regardless of whether they stay through eighth grade, we avoid the problem of overstating the effect of KIPP. Instead, this approach is likely to produce a conservative estimate of KIPP's full impact on students during the years they actually attended KIPP schools.

Grade Repetition. KIPP schools retain students in grades 5 and 6 at a substantially higher rate than do conventional public schools in their local districts (Tuttle et al. 2013). This produces a missing data problem for the analysis of state test scores during middle school, as students who repeat a grade do not take the same tests as others in their original cohort. Because KIPP students and comparison students are retained at different rates, our impact estimates could also be biased if we simply excluded all of the retained students from the analysis (doing so would exclude a larger proportion of KIPP students and a smaller proportion of the comparison students). To address this, in the matching analysis of math and reading scores we used information on students' past performance to predict (impute) their outcome scores in the years after retention. For more details on this procedure, see the discussion of imputation methods later in this appendix.

Attrition from the sample. As discussed in the previous section of this appendix, we conducted a detailed battery of baseline equivalence tests to determine if there was differential sample attrition for any of the outcomes we examined. Overall, as shown in the baseline equivalence tables presented above, the pattern of sample attrition for KIPP students is similar to the pattern for students in the matched comparison group for all of the middle school outcomes included in our analysis.

The remainder of this appendix presents the additional details regarding the study's propensity score estimation model, matching procedures, imputation model for baseline test scores, and imputation model for the middle school test scores of grade repeaters.

Propensity Score Matching Procedures

The matching procedure consists of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of KIPP students in the sample; and (3) testing the balance of baseline characteristics between our KIPP sample and matched comparison group.

For the first step, we separated the students in each district-level data set[31] into cohorts—grade-by-year groups for each typical KIPP entry grade (5th and 6th) in each year observed in

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

[31] For the purpose of estimating the matching model, we grouped KIPP middle schools together into a single district-wide file when multiple KIPP schools were located in the same district or metropolitan area. After matching

the data. For each cohort of students at a given KIPP school, the pool of eligible comparison students was limited to those in the same district and grade as the KIPP students the year before they first enrolled in a KIPP middle school; comparison students were restricted to those never enrolled in KIPP at any time during middle school. We then performed an iterative propensity score estimation procedure on a combined data set of all cohorts. The dependent variable in this propensity score model is an indicator of whether the student enrolled in a KIPP school in either grade 5 or grade 6.[32] Covariates in the model were selected using an iterative process that identifies the baseline demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. Table F.7 provides an exhaustive list of potential covariates for inclusion in each model.

At a minimum, we required the logistic model to include one year of baseline test scores in both math and reading. The other covariates were iteratively included and tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

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

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

was complete, we estimated impacts using the study's impact model (discussed below) separately for each KIPP middle school in the sample.

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

Next, we calculated propensity scores for KIPP entry. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (without replacement) of comparison group students to treatment group students, separately by cohort, from within the region of common support. In other words, for each KIPP student with a propensity score that fell within the range of propensity scores found among non-KIPP students, we identified the non-KIPP district student whose propensity score was closest to that of the KIPP student.

We then tested the balance of the KIPP group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, special education status, free and reduced price lunch status, and limited English proficiency status). For the matched comparison group sample associated with each KIPP school, we required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no more than one significant difference on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this covariate for the comparison group are not significantly different from the treatment group at the five percent level.[33] If the first round of matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that KIPP school, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[34] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

Impact model and covariates

To obtain impact estimates using this matched sample, we estimated an ordinary least squares (OLS) regression model that considered all math and reading test score data from grades 5–8 to measure students' outcome test scores and incorporated baseline (4th grade) demographic controls including indicators for gender, race/ethnicity, free/reduced-price lunch status, special education status, grade retention in a baseline year, and limited English proficiency status; cohort (year by entry grade); outcome test grade level; and two years of baseline mathematics and reading test scores (3rd and 4th grade for cohorts entering KIPP in grade 5; 4th and 5th grade for cohorts entering KIPP in grade 6). See Table F.8 for a full list of these covariates. The basic form of the model for each school is defined in equation F1:

$$\text{(F1)}\ y_{it} = a + X_i\beta + \delta_1 T1_{it} + \delta_2 T2_{it} + \delta_3 T3_{it} + \delta_4 T4_{it} + grade\_dummies + cohort\_dummies + \varepsilon_{it}$$

[33] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table F.2 through Table F.6, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups for any of the outcomes in this study.

[34] If balance was not achieved in the first round of matching for a given school, under our protocol we would remove the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20). In addition, to address estimation problems in the logistic regression we would consider removing terms identifying exceedingly rare attributes in the treatment group.

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and two years of baseline test scores) of student i; $T1_{it}$ through $T4_{it}$ are binary variables for treatment status in up to four years,[35] indicating whether student i had first enrolled at KIPP one, two, three, or four years previously, as of school year t. For example $T3_{it}$ would be equal to 1 for student i at time t if the student had first enrolled at KIPP at time (t-3), regardless of whether the student was still enrolled at KIPP at time t; otherwise, $T3$ would be equal to 0. The model also include a set of dummy indicator variables for each middle school grade and student cohort in the sample. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ_1, δ_2, δ_3, δ_4, and β are parameters or vectors of parameters to be estimated. The estimated coefficient on the treatment indicator, δ_n represents the cumulative impact of n years of KIPP treatment. Robust standard errors were clustered at the student level since individual students could contribute up to four observations to the analysis sample.

We used the model to separately estimate the impact of each KIPP middle school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP middle schools in the sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

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

Math baseline test score from 1 year prior
Math baseline test score from 2 years prior
Reading baseline test score from 1 year prior
Reading baseline test score from 2 years prior
Gender indicator variable
Set of race/ethnicity indicator variables
Special education status indicator variable
Free or reduced price lunch status indicator variable
Limited English proficiency status indicator variable
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from 1 and 2 years prior are imputed
Set of dummy variables indicating if a student is missing data for demographic variables
Dummy variables indicating whether student repeated grades in either of the two baseline years
Dummy variables for grades 5-8
Dummy variables for each student cohort in the sample

Note: Baseline test scores were imputed when missing. In one jurisdiction data was not available on limited English proficiency status.

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

Imputation for missing baseline data

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

(F2a) $$Yp\_math_{it} = a + X_i\beta + \sum_r \varphi_r Yr\_math_{it} + \sum_{q=3}^{8} \gamma_q Yq\_reading_{it} + \varepsilon_{it}$$

(F2b) $$Yp\_reading_{it} = a + X_i\beta + \sum_r \varphi_r Yr\_reading_{it} + \sum_{q=3}^{8} \gamma_q Yq\_math_{it} + \varepsilon_{it}$$

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

We first estimated equations (F2a) and (F2b) for baseline test scores one and two years prior to KIPP entry using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores (in other words, values of the right hand side variables in equations F2a and F2b) and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student. We only imputed missing baseline test scores for students who have at least one non-missing baseline test score in either math or reading.

Finally, to obtain the imputed baseline test scores used in our benchmark model, we added a stochastic component to the predicted values of Yp_math_{it} and $Yp\_reading_{it}$ obtained from estimating equations (F2a) and (F2b) above. For each student, the stochastic component is randomly selected from the set of all residuals estimated in equations (F2a) and (F2b) for the full sample. The stochastic component is included to ensure that the variance of the imputed baseline test scores is the same as that of the observed values.

While we use these imputed baseline and pre-baseline test scores in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variables were simply treated as being missing from the sample.

To test whether our results are sensitive to this imputation strategy, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores. The results for this smaller sample are nearly identical to our benchmark

impact estimates for the matched middle school analysis (Table F.9). There are no statistically significant differences on any baseline measure and the KIPP impact for each test outcome except year 1 reading remains positive and statistically significant at the 0.01 level, two-tailed test. The magnitude of each impact estimate is very similar to the benchmark estimate as well.

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

	Treatment group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (reading year 1)	-0.102	17,518	-0.084	17,397	-0.018	0.101
Reading scores (reading year 2)	-0.087	14,079	-0.085	13,679	-0.002	0.897
Reading scores (reading year 3)	-0.049	11,318	-0.045	10,837	-0.004	0.761
Reading scores (reading year 4)	-0.007	7,430	-0.035	7,121	0.027	0.149
Math scores (math year 1)	-0.095	17,525	-0.082	17,413	-0.013	0.244
Math scores (math year 2)	-0.070	14,064	-0.082	13,672	0.012	0.413
Math scores (math year 3)	-0.043	11,185	-0.060	10,741	0.017	0.259
Math scores (math year 4)	-0.065	6,887	-0.083	6,737	0.018	0.386
Reading scores (history)	0.022	4,866	-0.013	4,896	0.034	0.112
Math scores (science)	-0.053	8,826	-0.059	8,587	0.006	0.704
	Treatment group		**Comparison group**			
Outcome measure	**Adjusted mean**	**Sample size**	**Mean**	**Sample size**	**Impact estimate**	**p-value**
Reading year 1	-0.105	17,518	-0.111	17,397	0.005	0.426
Reading year 2	-0.008	14,079	-0.113	13,679	0.105**	0.000
Reading year 3	0.062	11,318	-0.092	10,837	0.154**	0.000
Reading year 4	0.076	7,430	-0.085	7,121	0.161**	0.000
Math year 1	-0.051	17,525	-0.109	17,413	0.057**	0.000
Math year 2	0.091	14,064	-0.141	13,672	0.232**	0.000
Math year 3	0.170	11,185	-0.121	10,741	0.291**	0.000
Math year 4	0.136	6,887	-0.131	6,737	0.268**	0.000
History	0.107	4,866	-0.131	4,896	0.238**	0.000
Science	0.084	8,826	-0.166	8,587	0.249**	0.000

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline and outcome tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Reported impacts are an average of equally weighted impact estimates for each KIPP middle school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. Impacts represent the cumulative effect of KIPP after the noted number of years after admission for math and reading scores, not the marginal effect of an additional year. The grade level of middle school exams used for history and science outcomes varied by jurisdiction. We selected the highest middle school grade level where science or social studies was observed for more than one cohort of KIPP students. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Imputation for students repeating a grade

We also impute the math and reading state test scores of students who repeat a grade if they were retained in one of the study's four outcome years. For example, if a student in the treatment group entered KIPP in grade 5 and then repeated grade 6, they would still be in grade 6 (and would take the grade 6 state assessment) at the end of the third follow-up year. Members of their cohort who remained on track would have taken the grade 7 state assessment. Because the grade repeater's grade 6 assessment score would not be comparable to grade 7 scores, we treat this student's year 3 follow-up score as missing and impute its value. To do so, we use the following approach in the math and reading analyses: for each grade repeater, in the year of repetition and subsequent years, we impute the student's z-score on the cohort-appropriate (rather than grade-appropriate) test by setting his or her score equal to the student's standardized score in the last year prior to grade repetition. In this example, we would use the standardized score of the grade repeater on the grade 6 assessment in the second follow-up year (the score from the first time the student took that assessment). In effect, this imputation procedure assumes students maintain the same percentile rank relative to their cohort in the year of grade retention and in all subsequent years. In other words, we assume that each retained student does neither better or worse in relative terms than before retention. If KIPP in fact has a positive impact on retained students, this would cause us to underestimate KIPP's impact. Conversely, if KIPP has a negative impact, this would cause us to overestimate the impact.

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

To test the sensitivity of our results to the method used for retained students, in our prior analyses of KIPP schools we estimated KIPP impacts using several alternative approaches to analyzing the test scores of retained students (Tuttle et al. 2013). For example, we conducted a sensitivity test that assigned the test score of a student to the fifth percentile of the jurisdiction's analysis sample in the grades they would have attended under a "normal" grade progression.[36] Even using this conservative approach, in that analysis the KIPP impact estimates remained positive and statistically significant in both reading and math. However, as we might expect, the magnitude of each positive impact was somewhat smaller than under our benchmark approach (the estimates from this sensitivity test were between 0.02 and 0.06 standard deviations smaller in both math and reading).

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

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

APPENDIX G

DETAILED ANALYTIC METHODS: HIGH SCHOOL (MATCHED-STUDENT ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP high schools for students who entered the KIPP network for the first time in grade 9 (our matched- student analysis for high school outcomes). First we present information on the sample and the baseline equivalence of the treatment and matched comparison group for each outcome. Next we discuss the data used in the analysis, and we conclude the appendix with a discussion of the analytic methods used for the analysis.

Detail on sample

Overall, the analysis included 14 different KIPP high schools, nine of which opened in fall 2010 or earlier, meaning we have at least four years of high school data for those schools. Four high schools in our sample opened in fall 2011 and one opened in fall 2013. The 14 high schools represent 9 different states and include 70 percent of the high schools in the KIPP network as of the 2013-14 school year. The total sample of treatment and matched comparison students for each included school is shown in table G.1, along with the number of student cohorts represented in the data.

At the high school level, students served by KIPP have similar characteristics to students at KIPP elementary and middle schools, but differ from other students in the school districts where KIPP high schools in our sample operate. KIPP high school students are more likely to be female, black, and Hispanic than those attending non-KIPP public high schools (Figure G.1).[37] For example, as at the elementary and middle school levels, most KIPP high school students are black (52 percent) or Hispanic (44 percent), compared with 42 and 33 percent among the non-KIPP high school population. KIPP students are also significantly more likely than students at nearby non-KIPP high schools to qualify for free or reduced-price lunch, a proxy for having low family income (84 versus 68 percent). They are slightly (but statistically significantly) less likely to have special education needs than non-KIPP high school students (9 versus 11 percent), while being slightly more likely to be limited English proficiency students (7 versus 6 percent). Finally, KIPP high school students have significantly higher baseline (grade 8) math and reading test scores than non-KIPP students. In particular, on state tests in grade 8, KIPP high school students scored at the 57th percentile in reading and the 60th percentile in math, compared with the 50th percentile in each subject for the non-KIPP high school students. As noted in chapter II, about two-thirds of KIPP high school students attended a KIPP middle school in grade 8, so these grade 8 achievement scores may reflect the influence of KIPP in that year and previous years.

[37] We also compared the baseline (grade 8) characteristics of KIPP high school students to the characteristics of all students who attended a "feeder middle school"—a non-KIPP school where at least one student within the school went on to attend a KIPP high school. The results of that analysis were similar to those presented in Figure III.3.

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

School	Region	Analytic baseline sample			Number of cohorts (years of school data)
		KIPP (N)	Non-KIPP (N)	Total sample size	
KIPP Denver Collegiate High School	Denver, CO	272	272	544	5 (2009-10 to 2013-14)
KIPP DC: College Preparatory	Washington DC	68	68	136	5 (2009-10 to 2013-14)
KIPP Atlanta Collegiate	Atlanta, GA	113	113	226	3 (2011-12 to 2013-14)
KIPP NYC College Prep High School	New York City	78	78	156	4 (2010-11 to 2013-14)
KIPP Memphis Collegiate High	Memphis, TN	149	149	298	3 (2011-12 to 2013-14)
KIPP Pride High School	Eastern NC	110	110	220	8 (2005-06 to 2013-14)[a]
KIPP Delta Collegiate High School	Arkansas Delta	42	42	84	5 (2008-09 to 2013-14) [b]
KIPP Academy Lynn Collegiate High	Massachusetts	58	58	116	3 (2011-12 to 2013-14)
KIPP Austin Collegiate	Austin, TX	124	124	248	5 (2008-09 to 2013-14)[b]
KIPP University Prep High School	San Antonio, TX	86	86	172	5 (2009-10 to 2013-14)
KIPP Houston High School	Houston, TX	125	125	250	9 (2005-05 to 2013-14)[c]
KIPP Northeast College Preparatory	Houston, TX	23	23	46	1 (2013-14)
KIPP Sunnyside High School	Houston, TX	76	76	152	4 (2010-11 to 2013-14)
KIPP Generations Collegiate	Houston, TX	56	56	112	3 (2011-12 to 2013-14)
Total	**14 schools**	**1,380**	**1,380**	**2,760**	

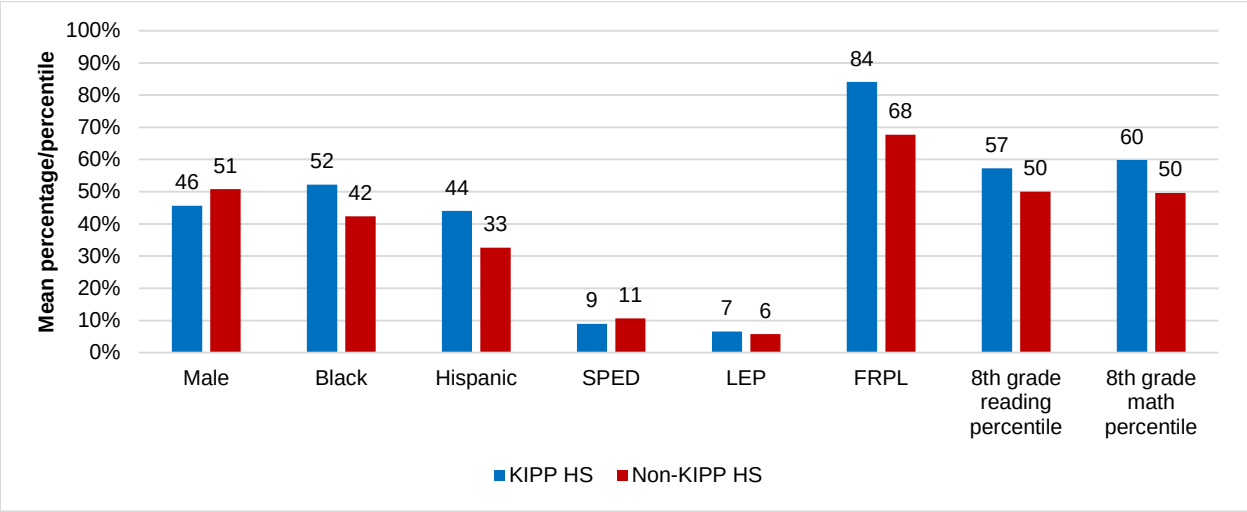
Notes: Test outcomes are drawn from administrative records for each of the first four years following enrollment high school. Treatment students are KIPP high school students who never attended a KIPP middle school, and comparison students are matched based on baseline (grade 8) characteristics.

[a] Data does not include cohort from the 2008-2009 school year

[b] Data does not include cohort from the 2010-2011 school year

[c] Data does not include cohort from the 2011-2012 school year

Figure G.1. Characteristics of KIPP versus non-KIPP high school students



Notes: Sample includes both new entrants and continuing students at the 14 high schools represented in the matched-student sample. Values are proportions unless otherwise indicated. All differences are statistically significant at the 0.05 level, two-tailed test. Each KIPP high school is given equal weight to calculate the overall average and statistical significance.

While the study's matching procedures (described below) ensured that the treatment group (KIPP students) and comparison group (non-KIPP students) were equivalent at baseline, we did not observe outcome measures for every matched student in every outcome sample. To check that treatment and comparison students included in each of our analysis samples are equivalent on observable characteristics, we examined baseline equivalence separately for the study's high school outcomes: math, ELA, science, and social studies exams; and graduating high school within four years. We looked at nine baseline characteristics including baseline reading and math test scores, gender, race, special education, limited English proficiency, free- or reduced price lunch status, and whether the student repeated 8th grade. The following table shows the baseline equivalence of each outcome's analytic sample.

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

Baseline Characteristic	KIPP	Non-KIPP	Difference	Std. Error	# KIPP	# Comparison
Sample for ELA outcome						
Reading scores (z-score)	-0.047	-0.072	0.024	0.052	910	888
Math scores (z-score)	-0.096	-0.086	-0.011	0.049	889	864
Student is male	0.454	0.503	-0.049	0.030	943	918
Student is black	0.530	0.519	0.011	0.021	943	918
Student is Hispanic	0.438	0.436	0.001	0.022	943	918
Special Education	0.061	0.062	-0.001	0.014	943	918
Limited English Proficiency	0.082	0.091	-0.009	0.016	689	649
Free- or reduced-price lunch	0.836	0.847	-0.012	0.019	943	918
Grade Repeaters	0.012	0.016	-0.004	0.006	943	918
Sample for math outcome						
Reading scores (z-score)	0.002	0.001	0.001	0.052	738	703
Math scores (z-score)	-0.084	-0.060	-0.024	0.053	727	692
Student is male	0.446	0.495	-0.050	0.032	763	726
Student is black	0.547	0.540	0.007	0.018	763	726
Student is Hispanic	0.414	0.411	0.002	0.019	763	726
Special Education	0.060	0.071	-0.011	0.015	763	726
Limited English Proficiency	0.083	0.086	-0.003	0.014	566	508
Free- or reduced-price lunch	0.826	0.816	0.010	0.023	763	726
Grade Repeaters	0.015	0.015	0.000	0.008	763	726
Sample for science outcome						
Reading scores (z-score)	-0.081	-0.123	0.042	0.057	680	667
Math scores (z-score)	-0.110	-0.109	0.000	0.049	657	643
Student is male	0.472	0.469	0.004	0.032	697	686
Student is black	0.446	0.447	-0.001	0.024	697	686
Student is Hispanic	0.515	0.496	0.019	0.025	697	686
Special Education	0.051	0.064	-0.013	0.015	697	686
Limited English Proficiency	0.120	0.115	0.005	0.021	443	415
Free- or reduced-price lunch	0.830	0.838	-0.008	0.022	697	686
Grade Repeaters	0.010	0.007	0.003	0.005	697	686
Sample for social studies outcome						
Reading scores (z-score)	0.020	0.012	0.008	0.070	315	308
Math scores (z-score)	-0.039	-0.008	-0.031	0.071	304	297
Student is male	0.480	0.478	0.002	0.043	325	318
Student is black	0.528	0.518	0.010	0.027	325	318
Student is Hispanic	0.448	0.440	0.008	0.027	325	318
Special Education	0.033	0.045	-0.012	0.016	325	318
Limited English Proficiency	0.000	0.000	0.000	0.000	123	100
Free- or reduced-price lunch	0.796	0.793	0.003	0.034	325	318
Grade Repeaters	0.008	0.025	-0.017	0.010	325	318
Sample for 4-year graduation outcome						
Reading scores (z-score)	-0.014	-0.125	0.111	0.067	426	428
Math scores (z-score)	-0.028	-0.155	0.127	0.074	426	426
Student is male	0.485	0.508	-0.023	0.041	440	444
Student is black	0.439	0.420	0.019	0.022	440	444
Student is Hispanic	0.515	0.519	-0.004	0.022	440	444
Special Education	0.050	0.065	-0.015	0.018	440	444
Limited English Proficiency	0.129	0.187	-0.059**	0.022	232	237
Free- or reduced-price lunch	0.789	0.806	-0.018	0.030	440	444
Grade Repeaters	0.007	0.022	-0.015	0.012	440	444

Source: Test outcomes are drawn from state or district administrative records.

Notes: All values in this table are based on non-imputed data and the sample for which we have the outcome test score. Values are proportions unless otherwise indicated. Due to rounding, the value reported in the

"Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to the local jurisdiction of each high school. One jurisdiction did not provide data on limited English proficiency, reducing the sample size for that indicator.

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

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

None of the outcome samples show a statistically significant difference between the treatment and comparison group with respect to baseline test scores in reading and math. In addition, for four of the five outcome samples there were no significant differences for any of the demographic characteristics we examined. The only exception was the four-year graduation outcome, where the treatment group includes fewer limited English proficiency students compared to the control group (by 5.9 percentage points).

Detail on data

For the matched-student comparison group analysis, we used de-identified, longitudinally-linked student-level data from jurisdictions (states or districts) hosting at least one KIPP school and able to provide student-level records at the time of data collection. The variables from jurisdictions' administrative data systems included: test scores in ELA, mathematics, social studies, and science (where high school scores represent the primary outcome and middle school scores represent a key matching variable and baseline covariate); demographic characteristics, used for matching and as baseline covariates; and schools attended and dates of enrollment, identifying students' exposure to KIPP. Within each jurisdiction, we requested data for all school years beginning with two years prior to the KIPP high school's first year (to capture baseline data) through the 2013–14 school year. Of the 14 high schools in our sample, we obtained data from districts for four of the schools. For the other ten we obtained records from the state in which the high school was located and then limited the data to the district (or districts) from which the KIPP school drew students.

Importantly, the test outcomes at the high school level are different than for statewide grade-specific exams administered during middle school. A few states include statewide assessment tests for high school students in particular grades (as in middle school) but other states have end-of-course exams that students take after they complete specific courses, such as algebra I or biology. For end-of-course exams, students may complete the exam in a variety of different grades and years, depending on their course progression and grade progression pattern. For our analysis, outcome scores from these exams were limited to the test from the first year each student completed a given end-of-course exam (that is, we disregarded scores from retests if a student was retained and took the same exam in multiple years). To make the analysis of state high school test scores comparable across states and districts, all raw test scores were converted to z-scores defined relative to the distribution of scores in each year, subject, and jurisdiction. That is, for each jurisdiction associated with a given KIPP school, we calculated the difference between each student's raw score and the mean score recorded in that year across all high school grades in that subject, and then divided the difference by the standard deviation of raw scores in the jurisdiction in that year and subject. Thus, each z-score reflects a student's achievement level relative to the average student in the relevant cohort and jurisdiction (in terms of the number of standard deviations above or below the mean).[38]

For some jurisdictions and subjects, there were multiple different end-of-course exams available for a given academic subject. In these cases, we selected the exam where the test-taking pattern among KIPP high school students was most similar to the pattern in the matched

[38] By definition, the distribution of student z-scores in high school has a mean of 0 and standard deviation of 1 for each subject (math, ELA, science, and social studies).

comparison group in terms of the number of test takers and the timing of the exam. Table G.3 summarizes the outcome exams used in each jurisdiction.

The high school analysis also includes an outcome indicator for graduating within four years after entering 9th grade. There are two important limitations to the study's graduation outcome. First, our propensity score matching approach relies on the assumption that the pre-KIPP characteristics observed in our data (middle school test scores and demographic attributes) fully capture attributes that are associated both with selection into KIPP and the outcomes of interest. In the case of graduation, this assumption may be somewhat questionable because some attributes that are likely to be correlated with graduation, such as parental involvement or students' commitment to achieving long-term goals, may not be fully captured in our data. A second important limitation is that the administrative data used in this study only identifies graduation for the students who remain in the data through grade 12. In other words, the graduation indicator we use cannot distinguish between dropouts (who did not graduate) and students who transferred to private school or to a different school district (who may or may not have graduated high school). Both of these groups of students are classified as "non-graduates" in the analysis. This data limitation is a potential source of bias in the graduation impact estimates. If KIPP students transfer to other districts or private schools after grade 9 at a greater rate than comparison students, our impact estimates would be more negative than the true impact of KIPP. Conversely, if the comparison group transfers out of district more often during high school than the treatment group, our impact estimates would be more positive than the true impact of KIPP. The baseline equivalence results presented above for the four-year graduation indicator provides evidence that differential attrition between KIPP and non-KIPP students does not bias the impact estimate. (Only the Limited English Proficiency indicator showed a statistically significant difference between the two groups).

There could also be a missing data issue for the test outcomes. For a variety of reasons, some students may not have valid data in the year when a given outcome was measured. For example, some students may transfer to a jurisdiction outside of our data catchment area, while others may transfer to local private schools or drop out of school altogether. In a small number of cases, students may simply have missing variable values in a given year or subject. We categorize these cases when students disappear from the analytic sample as out-of-district transfers. If KIPP students transfer out-of-district at a different rate than matched comparison students, it could undermine the validity of impact estimates. To check this, we examined baseline equivalence of the sample for each of our four academic outcome measures (math, ELA, science, and social studies). As noted above, there were no statistically significant differences on any of the baseline characteristics for each test outcome sample. All variables assessed for baseline equivalence are also controlled for in our impact regression model.

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

High school	Data entity	Math	ELA	Science	Social studies
KIPP Denver Collegiate High School	CO Dept. of Education	CSAP grade 9 (2010-11), TCAP grade 9 (2012-14)	CSAP grade 9 (2010-11), TCAP grade 9 (2012-14)	CSAP grade 10 (2010-11), TCAP grade 10 (2012-13), ACT science (2014)	n.a.
KIPP DC: College Preparatory	OSSE (Washington D.C.)	CAS 10th grade	CAS 10th grade	n.a.	n.a.
KIPP Atlanta Collegiate	Atlanta Public Schools	Geometry	Lit Comp	Biology	U.S. History
KIPP NYC College Prep High School	NYC Dept. of Education	Integrated Algebra Regents	ELA Regents	Living Environment Regents	n.a.
KIPP Memphis Collegiate High	Shelby County Schools	Algebra II	English I	Biology	U.S. History
KIPP Pride High School	NC Dept. of Public Instruction/NCERDC	Algebra I	English I (1999-2012), English II (2013-2014)	Biology	U.S. History
KIPP Delta Collegiate High School	AR Dept. of Education	Geometry	Literacy	n.a.	n.a.
KIPP Academy Lynn Collegiate High	MA Dept. of Education	MCAS grade 10	MCAS grade 10	MCAS Biology grade 10	n.a.
KIPP Austin Collegiate	Texas Education Agency	TAKS grade 11	TAKS grade 11	TAKS grade 11	TAKS grade 11
KIPP University Prep High School	Texas Education Agency	TAKS grade 11	TAKS grade 11	TAKS grade 11	TAKS grade 11
KIPP Houston High School	Texas Education Agency	TAKS grade 11	TAKS grade 11	TAKS grade 11	TAKS grade 11
KIPP Northeast College Preparatory	Texas Education Agency	n.a.	English I	Biology	n.a.
KIPP Sunnyside High School	Texas Education Agency	TAKS 9 (2011); Algebra I (2012-14)	TAKS 9 (2011); Reading I (2012-13); English I (2014)	Biology	World Geography
KIPP Generations Collegiate	Texas Education Agency	n.a.	TAKS 9 (2011); Reading I (2012-14)	Biology	World Geography

Notes: Abbreviations for data entities correspond to the following: OSSE = Office of the State Superintendent of Education; NCERDC = North Carolina Education Research Data Center. Abbreviations for exams correspond to the following: TAKS = Texas Assessment of Knowledge and Skills; TCAP = Transitional Colordado Assessment Program; CSAP = Colorado Student Assessment Program; CAS = Comprehensive Assessment System; MCAS = Massachusetts Comprehensive Assessment System

Different analytic sample attrition might occur when students are missing one or more baseline or pre-baseline test scores. To address this we imputed missing baseline data, ensuring that all students with at least one recorded baseline test score remain in the sample. For a detailed discussion of our imputation methods, see the discussion of imputation methods below.

Detail on analytic methods

This study relied on a matched comparison group design that used "nearest neighbor" matching to identify a similar comparison student for each treatment student entering a KIPP high school in grade 9. The validity of our matched comparison group design depends on the ability to eliminate or minimize differences in key characteristics between students who enter KIPP and students in the comparison group who remain in non-KIPP public schools.[39] Our approach achieved this in two ways. First, we used student-level data that included a rich set of student characteristics and multiple years of baseline (prior to KIPP entry) test scores. We used this information to identify a matched comparison group of students who are similar to KIPP students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in middle school. By matching on more than one year of baseline test score data, we accounted for achievement levels at the time when students applied to KIPP high schools as well as pre-KIPP trends in student achievement. After we identified the matched comparison group, the second feature of our approach estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between KIPP students and comparison students. Specifically, the impact estimates adjust for any differences between KIPP students and the matched comparison group pertaining to demographic characteristics or students' prior two years of math and reading test scores.

The combination of propensity-score matching and OLS accounts for differences in observed baseline characteristics and achievement scores between KIPP students and comparison students (in other words, the differences associated with initial selection into KIPP schools). But it remains possible that KIPP students and comparison students differ in unobserved ways that may affect later test scores. However, previous studies have suggested that applying a combination of propensity-score matching and OLS, as we did here, can succeed in replicating experimental impact estimates in certain contexts (Cook et al. 2008; Bifulco 2012; Furgeson et al. 2012; Tuttle et al. 2013; Fortson et al. 2015).

There are several other threats to the validity of these impact estimates that we addressed, including students moving from KIPP middle schools to other district schools and attrition from the sample.

Attrition from KIPP Schools. The fact that some students depart KIPP schools and return to non-KIPP schools in the surrounding district before the end of 12th grade could potentially introduce selection bias if not appropriately handled. If lower-performing student tend to exit KIPP high schools before graduation, an analysis that only includes the persistently enrolled KIPP students in the treatment group would positively bias the estimated impact of KIPP schools (that is, make KIPP impacts look more positive than they actually are). We addressed this

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

problem by permanently assigning to the treatment group any student who can be found in the records as enrolling at KIPP for the first time in grade 9, regardless of whether the student remained in a KIPP high school or transferred elsewhere before the end of high school.[40] In other words, a student who enrolled at KIPP in 9th grade for the 2009–10 school year but left KIPP after completing 10th grade in the 2010–11 school year is included in the treatment group for all four years he or she appears in the data (from 2009–10 to 2012–13, inclusive). By including all students observed as entering a KIPP school in grade 9, regardless of whether they stay through the end of grade 12, we avoid the problem of overstating the effect of KIPP. Instead, this approach is likely to produce a conservative estimate of KIPP's full impact on students during the years they actually attended KIPP schools.

Attrition from the sample. As discussed in the previous section of this appendix, we conducted a detailed battery of baseline equivalence tests to determine if there was differential sample attrition for any of the outcomes we examined. Overall, as shown in the baseline equivalence tables presented above, the pattern of sample attrition for KIPP students is similar to the pattern for students in the matched comparison group for all of the high school outcomes included in our analysis.

The remainder of this appendix presents the additional details regarding the study's propensity score estimation model, matching procedures, and imputation model for baseline test scores.

Propensity Score Matching Procedures

The matching procedure consists of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of KIPP students in the sample; and (3) testing the balance of baseline characteristics between our KIPP sample and matched comparison group.

For the first step, we separated the students in each district-level data set into cohorts—grade-by-year groups for the KIPP high school entry grade (grade 9) in each year observed in the data. For each cohort of students at a given KIPP school, the pool of eligible comparison students was limited to those in the same district and grade as the KIPP students the year before they first enrolled in a KIPP high school; comparison students were restricted to those never enrolled in KIPP at any time during elementary or middle school. We then performed an iterative propensity score estimation procedure on a combined data set of all cohorts. The dependent variable in this propensity score model is an indicator of whether the student enrolled in a KIPP school in grade 9.[41] Covariates in the model were selected using an iterative process that identifies the baseline

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

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

demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. Table G.4 provides an exhaustive list of potential covariates for inclusion in each model.

At a minimum, we required the logistic model to include one year of baseline test scores in both math and reading. The other covariates were iteratively included and tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

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

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

Next, we calculated propensity scores for KIPP entry. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (without replacement) of comparison group students to treatment group students, separately by cohort, from within the region of common support. In other words, for each KIPP student with a propensity score that fell within the range of propensity scores found among non-KIPP students, we identified the non-KIPP district student whose propensity score was closest to that of the KIPP student.

We then tested the balance of the KIPP group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, special education status, free and reduced price lunch status, and limited English proficiency status). For the matched comparison group sample associated with each KIPP school, we required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no more than one significant difference on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this

covariate for the comparison group are not significantly different from the treatment group at the five percent level.[42] If the first round of matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that KIPP school, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[43] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

Impact model and covariates

To obtain impact estimates using this matched sample, we estimated an ordinary least squares (OLS) regression model for each examined outcome (test scores in math, ELA, science, and social studies, as well as high school graduation). The model incorporated baseline (8th grade) demographic controls including indicators for gender, race/ethnicity, free/reduced-price lunch status, special education status, grade retention in a baseline year, and limited English proficiency status; cohort (year by entry grade); and two years of baseline mathematics and reading test scores (7th and 8th grade). See Table G.5 for a full list of these covariates. The basic form of the model for each school is defined in equation G1:

$$(G1) \qquad y_{it} = a + X_i\beta + \delta treat_{it} + cohort\_dummies + \varepsilon_{it}$$

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and two years of baseline test scores) of student i; $treat_{it}$ is a binary variables for treatment status indicating whether student i had entered KIPP in grade 9. The model also include a set of dummy indicator variables for each student cohort in the sample. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ and β are parameters or vectors of parameters to be estimated. The estimated coefficient on the treatment indicator, δ represents the impact of the KIPP high school on the observed outcome. Robust standard errors were clustered at the student level.

We used the model to separately estimate the impact of each KIPP high school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP high schools in the sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

[42] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table G.2, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups for any of the outcomes in this study.

[43] If balance was not achieved in the first round of matching for a given school, under our protocol we would remove the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20). In addition, to address estimation problems in the logistic regression we would occasionally remove terms identifying exceedingly rare attributes in the treatment group.

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

Math baseline test score from 1 year prior
Math baseline test score from 2 years prior
Reading baseline test score from 1 year prior
Reading baseline test score from 2 years prior
Gender indicator variable
Set of race/ethnicity indicator variables
Special education status indicator variable
Free or reduced price lunch status indicator variable
Limited English proficiency status indicator variable
Set of math and reading imputation dummies indicating whether math and reading baseline test scores from 1 and 2 years prior are imputed
Set of dummy variables indicating if a student is missing data for demographic variables
Dummy variables indicating whether student repeated grades in either of the two baseline years
Dummy variables for each student cohort in the sample

Note: Baseline test scores were imputed when missing. In one jurisdiction data was not available on limited English proficiency status.

Imputation for missing baseline data and retained students

Here we explain in greater detail how our analysis handled missing data when students were missing baseline or pre-baseline test score data in grade 7 or 8. Our analyses used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores; imputation was completed separately by treatment status. This imputation process involved estimating the following model:

(G2a) $$Yp\_math_{it} = a + X_i\beta + \sum_r \varphi_r Yr\_math_{it} + \sum_{q=3}^{8} \gamma_q Yq\_reading_{it} + \varepsilon_{it}$$

(G2b) $$Yp\_reading_{it} = a + X_i\beta + \sum_r \varphi_r Yr\_reading_{it} + \sum_{q=3}^{8} \gamma_q Yq\_math_{it} + \varepsilon_{it}$$

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

We first estimated equations (G2a) and (G2b) for baseline test scores one and two years prior to KIPP entry using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores (in other words, values of the right hand side variables in equations G2a and G2b) and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student. We only imputed

missing baseline test scores for students who have at least one non-missing baseline test score in either math or reading.

Finally, to obtain the imputed baseline test scores used in our benchmark model, we added a stochastic component to the predicted values of Yp_math_{it} and $Yp\_reading_{it}$ obtained from estimating equations (G2a) and (G2b) above. For each student, the stochastic component is randomly selected from the set of all residuals estimated in equations (G2a) and (G2b) for the full sample. The stochastic component is included to ensure that the variance of the imputed baseline test scores is the same as that of the observed values.

While we use these imputed baseline and pre-baseline test scores in our analysis of KIPP's impacts, none of the imputed values were included in the tests of baseline equivalence discussed earlier in this appendix. For the analysis of baseline equivalence, students missing data on a given variables were simply treated as being missing from the sample.

To test whether our results are sensitive to this imputation strategy, we estimated our benchmark model using the subsample of students with complete baseline test score data—that is, we dropped students with missing baseline scores from the sample and compared the KIPP students for whom we did not impute scores to matched comparison students for whom we did not impute scores. The results for this smaller sample are nearly identical to our benchmark impact estimates for the matched-student high school impact analysis (Table G.6). There are no statistically significant differences on any baseline measure and the KIPP impact in ELA, math, and science remains positive and statistically significant while impacts in social studies and high school graduation are positive but not statistically significant. The magnitude of each impact estimate is nearly identical to the benchmark estimate as well.

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

Baseline measure (analyzed outcome for this sample)	Treatment group		Comparison group			
	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	-0.060	887	-0.089	861	0.028	0.591
Math scores (math)	-0.084	725	-0.059	691	-0.026	0.631
Reading scores (social studies)	0.000	304	0.001	297	-0.002	0.982
Math scores (science)	-0.111	656	-0.109	643	-0.002	0.973
Reading scores (4-year graduation)	-0.014	426	-0.125	426	0.111	0.099
Math scores (4-year graduation)	-0.028	426	-0.155	426	0.127	0.088
Free and reduced-price lunch status (4-year graduation)	0.806	426	0.809	426	-0.003	0.924
	Treatment group		**Comparison group**			
Outcome measure	**Adjusted mean**	**Sample size**	**Mean**	**Sample size**	**Impact estimate**	**p-value**
ELA achievement	0.111	887	-0.065	861	0.175**	0.000
Mathematics achievement	0.236	725	-0.036	691	0.273**	0.000
Science achievement	0.105	656	-0.223	643	0.328**	0.000
Social studies achievement	-0.134	304	-0.149	297	0.015	0.795
Four-year high school graduation	0.707	426	0.672	426	0.035	0.355

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

APPENDIX H

DETAILED ANALYTIC METHODS: HIGH SCHOOL (MATCHED-SCHOOL ANALYSES)

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

This appendix presents detailed information about the study's analysis of the impacts of KIPP high schools on student achievement for continuing students. Specifically, these results are estimates of the additional benefit of attending a KIPP high school among students who attended a KIPP middle school on achievement in reading, language, and math as measured by TerraNova test scores, as well as impacts on nonacademic outcomes. First we present information on the sample and baseline equivalence of the treatment and matched comparison group for each outcome. Next we discuss the data used in the analysis, and we conclude with a discussion of the analytic methods used for the analysis.

Analysis sample

Since the majority of students attending KIPP high schools also attended a KIPP middle school, our challenge was to identify a credible and rigorous comparison group for these students. We developed two different but complementary approaches, each with an associated sample of schools and students, to compare outcomes for KIPP middle school students who had an option to attend a KIPP high school with those who did not have an option to attend a KIPP high school. Those comparison students attended a wide variety of other high schools, including private, magnet, boarding, traditional public, or other charter high schools (discussed in more detail below). Our results are analogous to "intent-to-treat" findings, since not all students with the option to attend a KIPP high school do so. However, rates of enrollment in KIPP high schools among KIPP middle-school graduates in our sample are generally high where the option is available, at 70 percent overall and ranging from 59 to 83 percent across feeder KIPP middle schools. Effects on students actually enrolling in KIPP high schools would be larger than the "intent-to-treat" impacts on students with the opportunity to enroll.

For the first model (Same KIPP Middle School, Adjacent Cohorts), we focus on a set of KIPP high schools in their first year of operation. The treatment group in this model includes 8th-grade KIPP middle school students who had the option to attend the local KIPP high school in the first year of its operation. The comparison group includes the previous cohort of 8th-grade KIPP students from the same middle school who did not have the option to attend the local KIPP high school because it had not yet opened. Because sample sizes are too small to restrict the comparison group, we do not employ student-level matching in this model. Rather, we compare later outcomes of an entire class of KIPP middle school graduates in one cohort with the later outcomes of the entire classroom of KIPP middle school graduates from the previous cohort. There are five high schools (served by six feeder middle schools) included in this analysis, yielding an analytic sample of 467 students (229 treatment, 238 comparison).

For the second model (Same Cohort, Matched Middle Schools), the treatment group includes 8th-grade KIPP middle school students in 2008-09 who had the option to attend the local KIPP high school. The comparison group includes 8th-grade KIPP students from different middle schools in 2008-09 in regions with no KIPP high school open at the time. To define a sample that was equivalent at baseline (grade 8), we first identified a set of comparison KIPP middle schools that most resembled the feeder KIPP middle schools on the basis of average school-level characteristics (race/ethnicity, baseline achievement on a nationally-normed test, and baseline test instrument—either the SAT-10 or the MAP). Then, within these matched sets of schools, we conducted student-level propensity score matching to identify the individual comparison student who was the closest match to each treatment student on the basis of gender,

race/ethnicity, baseline achievement in reading and math, and whether the student was old for his/her grade at baseline. There are five high schools included in this analysis, served by six feeder middle schools. Together with students from five matched KIPP middle schools without a high school option, the analytic sample comprises 550 students (275 treatment, 275 comparison).

Across the two models, we estimate impacts for eight unique KIPP high schools and 933 students (464 treatment, 469 control). Two schools (University College Prep in San Antonio and DC College Prep) are included in both models, which allows us to compare estimates across models. A key difference between the models is that the Same KIPP Middle School, Adjacent Cohorts model examines only schools serving their first cohort of students, whereas the Same Cohort, Matched Middle Schools model includes a combination of new and more established high schools. The feeder middle school and studied cohort for each included school in the combined sample is shown in table H.1.

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

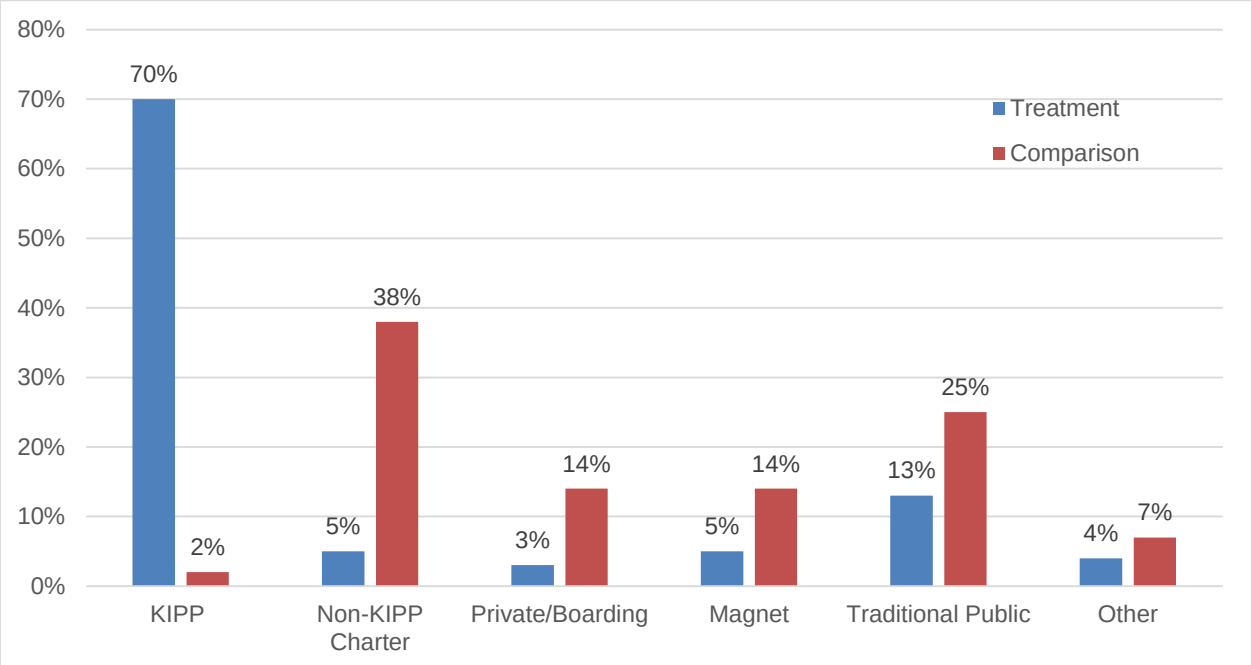
KIPP High School	Region	Year opened (cohort studied)	Same KIPP middle school, adjacent cohorts: KIPP feeder MS	Same cohort, matched middle school:	
				KIPP feeder MS	KIPP matched MS
Austin Collegiate	KIPP Austin	2008 (1st)	Austin College Prep	--	--
DC College Prep	KIPP DC	2009 (1st)	KEY Academy	KEY Academy, AIM Academy	Ascend, Philadelphia CS
Dubois Collegiate	KIPP Philadelphia	2010 (1st)	Philadelphia Charter	--	--
King Collegiate	KIPP Bay Area	2007 (3rd)	--	Summit Academy	Adelante, LA Prep, San Francisco Bay
Newark Collegiate	KIPP New Jersey	2007 (3rd)	--	TEAM Academy	Ascend, Philadelphia CS
NYC College Prep	KIPP NYC	2009 (1st)	KIPP Academy Middle (NYC), STAR	--	--
San Jose Collegiate	KIPP Bay Area	2008 (2nd)	--	Heartwood	Adelante, LA Prep, San Francisco Bay
University College Prep	KIPP San Antonio	2009 (1st)	Aspire	Aspire	Adelante, LA Prep, San Francisco Bay

Notes: Treatment students are KIPP middle school students who attended schools with an available KIPP high school, and are matched with comparison students at schools that did not have a KIPP high school option available.

As with the other grade levels, it is important to understand differences in the types of schools attended by the treatment and comparison group students to help provide context for the two groups' experiences in school and thus any impacts we observe (Figure H.1). For example, if all comparison group students attended traditional public schools, then the impact estimate would capture the effects of KIPP high schools versus these traditional public high schools. If most comparison students attended private or boarding schools, by contrast, the interpretation of the impact estimate would be different.

Using data from our matched-school sample of high school students, we investigate the types of high schools attended by KIPP middle school students, both when they have the option to attend a KIPP high school and when they do not. In the treatment group—those that have the option of a KIPP high school—the majority of students (70 percent) attend the KIPP high school. Another 13 percent attend traditional public schools and the rest are fairly evenly distributed between non-KIPP charter high schools, magnet schools, and private or boarding schools. The distribution of high school types attended is much different in the absence of a KIPP high school option, where the largest group of comparison students attend a non-KIPP charter school (38 percent) and another 14 percent attend magnet schools, another option involving choice. A much larger percentage attend private schools than do so when a KIPP high school is an option (14 percent versus 3 percent), but less than twice as many attend a traditional public school (25 percent versus 13 percent). These patterns suggest that KIPP high schools enroll students who would have otherwise enrolled in a wide variety of school types.

Figure H.1. Type of high schools attended by KIPP middle school students



Notes: Sample refers to the students in the matched-school analysis of KIPP high schools. Proportions reflect the schools students reported attending during grade 11, based on KIPP alumni records and a study-administered survey.

To check that each of our outcome samples are equivalent on observable characteristics, we examined baseline equivalence separately for each of the study's matched samples. We looked at six baseline, or 8th grade, characteristics including reading and math test scores; gender, race, special education, limited English proficiency, and free- or reduced price lunch status; and whether the student repeated a grade in the baseline year. The following tables (Tables H.2 through H.4) show the baseline equivalence of each analytic sample.

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

Baseline characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (z-score)	0.366	0.254	0.112	0.118	464	469
Math scores (z-score)	0.743	0.647	0.056	0.248	464	469
Student is male	0.450	0.443	0.008	0.879	464	469
Student is black	0.502	0.503	-0.001	0.982	464	469
Student is Hispanic	0.425	0.433	-0.008	0.855	464	469
Student is old for grade	0.207	0.213	-0.006	0.859	464	469

Note: All values in this table are based on the sample for which we have test outcome data. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution.

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

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

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

Baseline characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (z-score)	0.425	0.357	0.068	0.409	208	213
Math scores (z-score)	0.772	0.800	-0.028	0.718	208	213
Student is male	0.447	0.413	0.034	0.483	208	213
Student is black	0.519	0.521	-0.002	0.969	208	213
Student is Hispanic	0.476	0.408	0.068	0.164	208	213
Student is old for grade	0.250	0.254	-0.004	0.934	208	213

Note: All values in this table are based on the sample for which we have test outcome data. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution.

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

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

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

Baseline characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (z-score)	0.131	0.140	-0.009	0.926	256	256
Math scores (z-score)	0.568	0.561	0.007	0.955	256	256
Student is male	0.453	0.465	-0.012	0.858	256	256
Student is black	0.488	0.488	0.000	1.000	256	256
Student is Hispanic	0.383	0.453	-0.070	0.279	256	256
Student is old for grade	0.172	0.180	-0.008	0.871	256	256

Note: All values in this table are based on the sample for which we have test outcome data. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "Treatment" and "Comparison" columns. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution.

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

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

For the combined analysis sample, among the students with non-missing outcome data, there were no statistically significant differences between the treatment and control group on any baseline characteristics. Assessing baseline equivalence separately for the two different models, the treatment and comparison groups in the Same Cohort, Matched Middle School model are equivalent, which is consistent with the model's design due to the matching process. In the Same KIPP Middle School, Adjacent Cohorts model, some differences between groups are closer to being significant (but none that are significant at the 5% level), reinforcing the importance of controlling for these characteristics in the impact model.

Data

Baseline test score data is from KIPP-administered math and reading tests from grade 8. For the Same Cohort, Matched Middle School analysis, the KIPP middle schools either administered the Measures of Academic Progress (MAP) or the Stanford Achievement Test (SAT-10), and the schools were divided into two groups by baseline test score type for this analysis. For the Same Middle School, Adjacent Cohorts analysis the KIPP-administered baseline test was used unless the KIPP middle school changed tests between the two cohorts included in our analysis. In the cases where different tests were used across the two cohorts for a particular KIPP middle school, data on the state-administered math and reading tests for grade 8 were used for the baseline test scores.

The primary outcome measure for the achievement analysis is the study-administered TerraNova assessment (Form G, Level 21/22) in reading, language, and math, administered in 2011, 2012, or 2013 (depending on the cohort and treatment group) in treatment and comparison students' third year after grade 8 (typically grade 11). We measure students' performance on the TerraNova assessment with z-scores that were standardized to capture student achievement relative to that of a nationally representative norming population.

We analyzed the impact of KIPP high schools on student behavior and attitudes covering three broad areas: student motivation and engagement, education goals and college preparation,

and school experiences and satisfaction. Data was collected from student survey conducted between March and July in 2012, 2013, or 2014 (depending on the cohort and treatment group), the spring of the students' fourth year after completing 8th grade at a KIPP middle school (grade 12 for most students).

The student survey included multiple items capturing the same underlying construct we wished to measure in many cases, so we created indices that summarize students' responses on related data items. (See Appendix B for more detail.)

Analytic methods

The Same KIPP Middle School, Adjacent Cohorts analysis examines students at KIPP high schools in their first year of operation, with a treatment group of 8th-grade KIPP middle school students who had the option to attend the local KIPP high school in the first year of its operation, and a comparison group composed of students in the previous cohort of 8th-grade KIPP students from the same middle school who did not have the option to attend the local KIPP high school because it had not yet opened.

The Same Cohort, Matched Middle School analysis relied on a matched comparison group design that used "nearest neighbor" matching to identify a similar comparison student for each treatment student. The validity of our matched comparison group design depends on our ability to eliminate or minimize differences in key characteristics between students at KIPP middle schools with a KIPP high school option and KIPP middle school students in the comparison group who lacked such an option.[44] We used student-level data that included student characteristics and baseline test scores to identify a matched comparison group of students who are similar to treatment students in terms of observed demographic characteristics and—most importantly—baseline test scores measured while they were in middle school. By matching on baseline test scores, we accounted for achievement levels for students prior to entering high school.

After we identified the matched comparison groups for both analyses, we estimated impacts using ordinary least squares (OLS) regressions that control for any remaining baseline differences between treatment students and comparison students. Specifically, the impact estimates adjust for any differences between treatment students and the comparison group pertaining to demographic characteristics or students' prior math and reading test scores.

The combination of matching, either with adjacent cohorts or propensity-score matching, and OLS accounts for differences in observed baseline characteristics and achievement scores between treatment students and comparison students (in other words, the differences associated with having access to a KIPP high school). But it remains possible that treatment students and comparison students differ in unobserved ways that may affect later test scores. However, previous studies have suggested that quasi-experimental methods that are similar in some respects to the approach we use here can succeed in replicating experimental impact estimates in

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

certain contexts (Cook et al. 2008; Bifulco 2012; Fortson et al. 2012; Furgeson et al. 2012; Tuttle et al. 2013).

The remainder of this appendix presents the additional details regarding the study's propensity score estimation model, matching procedures, and imputation model for baseline data.

Propensity Score Matching Procedures (Same Cohort, Matched Middle Schools Analysis)

The matching procedure for the Same Cohort, Matched Middle Schools analysis consists of three steps: (1) determining the covariates to be included in the matching model, and estimating the matching model; (2) calculating propensity scores for sample members and selecting a matched comparison group based on these scores being close to those of treatment students in the sample; and (3) testing the balance of baseline characteristics between our treatment group and matched comparison group.

For the first step, we selected a group of comparison schools by matching potential comparison KIPP middle schools to the feeder KIPP middle schools, using aggregate student-level characteristics including race, ethnicity, and average achievement in grade 8, as well as state (where possible). We then separated the feeder KIPP middle schools into two groups based on the type of test administered in grade 8 (baseline) at that school: either the MAP or SAT-10. For each group, the pool of eligible comparison schools (and thus students) was limited to those administering the same baseline tests as the treatment students. We then performed an iterative propensity score estimation procedure on the data sets for each test group. The dependent variable in this propensity score model is an indicator of whether the student had access to a KIPP high school. Covariates in the model were selected using an iterative process that identifies the baseline demographic characteristics and test score variables, higher-order terms, and interaction terms that resulted in the best fit of the logistic model. Table H.5 provides an exhaustive list of potential covariates for inclusion in each model.

At a minimum, we required the logistic model to include baseline test scores in both math and reading. The other covariates were iteratively included and tested for whether they improved the fit of the logistic model. For this purpose only, we used a cut-off p-value of 0.20, instead of the traditional 0.05, to test for the significance of the covariates. If a potential covariate had a p-value of 0.20 or lower, it was retained in the matching model; it was dropped if its p-value exceeded 0.20.

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

Observed and imputed (when missing) math and reading baseline test scores from one year prior (always included)
Second and third order observed and imputed (when missing) values of math and reading baseline test scores from one year prior
Set of math and reading imputation dummies indicating whether math and reading baseline test scores are imputed
Demographic variables (gender, race/ethnicity, and whether the student was old-for-grade)
Interactions of baseline test scores and all available demographic variables
Interactions of gender and race/ethnicity variables
Interactions of old-for-grade indicator and race/ethnicity variables

Next, we calculated propensity scores for having the option to attend a KIPP high school. For any given sample member, the propensity score was based on the values for that individual of the variables included in the propensity score model multiplied by the estimated coefficients from the model. We then performed nearest neighbor matching (with replacement) of comparison group students to treatment group students, separately for each baseline test type (that is, only matching treatment and comparison group members who took the same type of baseline test), from within the region of common support. In other words, for each treatment student with a propensity score that fell within the range of propensity scores found among potential matched comparison students, we identified the comparison student whose propensity score was closest to that of the treatment student.

We then tested the balance of the treatment group and the matched comparison group by conducting a test of the significance of differences between the two groups in their baseline test scores and other demographic variables (race/ethnicity, gender, and old-for-grade). We required the baseline test scores of treatment students and comparison students to be balanced in both math and reading; we also required there to be no significant differences on any of the other demographic characteristics listed above. We consider a covariate to be balanced when the means of this covariate for the comparison group are not significantly different from the treatment group at the five percent level.[45] If the first round of matching did not identify a comparison group meeting these criteria, we adjusted the propensity score estimation model for that baseline test score group, re-estimated a new set of propensity-scores, obtained a new matched comparison group, and tested for balance between the treatment group and the new matched comparison group.[46] These steps were iterated until we obtained a matched comparison group that achieved balance with the treatment group according to our criteria.

[45] The What Works Clearinghouse standards require baseline test scores between treatment and control groups to differ by no more than 0.25 of a standard deviation if used as control variables in estimating equations. As shown in Table H.2-H.4, no baseline test scores in either subject differ by more than 0.25 of a standard deviation between treatment and control groups for any of the outcomes in this study.

[46] If balance was not achieved in the first round of matching for a given school, under our protocol we would remove the variable or interaction term with the least statistical significance (that is, the variable or interaction term that was closest to our p-value cutoff of 0.20).

Imputation for missing baseline data for propensity score models

Here we explain in greater detail how our propensity score models handled missing data when students were missing baseline test score data.[47] Our propensity score matching procedure used data sets with imputed baseline test scores created by conducting single stochastic regression imputation for missing baseline test scores; imputation was completed separately by site. Variables in the imputation model included baseline math and reading test scores, demographic characteristics (gender, race/ethnicity, and whether the student was old for grade). Treatment status not part of the imputation model because imputation is performed separately for the treatment group and then the comparison group.

We first estimated the imputation model using those students in our sample who have non-missing scores on these tests. For students with missing values for a given test, we used that student's demographic characteristics and other non-missing test scores and multiplied them by the estimated coefficients from the model. This gave us a predicted value of the missing test score for that student.

While we use these imputed baseline test scores in our propensity score matching model, none of the imputed values were included in our benchmark impact analysis or tests of baseline equivalence discussed earlier in this appendix. In these cases, students missing data on a baseline test score were simply treated as being missing from the sample.

Impact model and covariates

To obtain impact estimates for the two analysis samples, as well as for both samples combined, we estimated an ordinary least squares (OLS) regression model for each examined outcome (test scores in reading, language, and math, as well as the survey-based outcomes discussed in Appendix B). The models were run on the pooled sample of schools, separately for the two different analysis samples, as well as in a pooled model combining both analysis samples. The model incorporated baseline (8th grade) demographic controls including indicators for gender, race/ethnicity, and whether the student was old for grade[48]; and baseline mathematics and reading test scores (8th grade). See Table H.6 for a full list of these covariates. The basic form of the model is defined in equation H1:

$$\text{(H1)} \qquad y_{it} = \alpha + X_i\beta + \delta treat_{it} + indicator\_variables + \varepsilon_{it}$$

where y_{it} is the outcome test score for student i in school year t; α is the intercept term; X_i is a vector of characteristics (demographic controls and baseline test scores) of student i; $treat_{it}$ is a binary variables for treatment status indicating whether student i had access to a KIPP high school. The model also include a set of dummy indicator variables, indicating either the KIPP middle school attended (for the Same KIPP Middle School, Adjacent Cohorts model only), or the

[47] Imputed baseline test scores were used for the propensity score matching procedure. We tested whether the impact results changed when imputed baseline test scores were included as covariates, and the results were the same under both specifications.

[48] The cutoff for when a student was considered old for grade varied as was calculated based on the admission cutoff date for each location.

baseline test score type (for the Same Cohort, Matched Middle School group), and interactions of those indicators with the baseline test score variables. ε_{it} is a random error term that reflects the influence of unobserved factors on the outcome; δ and β are parameters or vectors of parameters to be estimated. The estimated coefficient on the treatment indicator, δ represents the impact of access to a KIPP high school on the observed outcome. Robust standard errors were clustered at the student level.

We used the model to separately estimate the impact for each KIPP high school in the sample. To calculate the average KIPP impact, the impact estimate for each KIPP school was given an equal weight. The standard error of the mean impact across all KIPP high schools in the sample uses the pooled student-level variance of school-specific impact estimates for each outcome sample.

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

Math baseline test score
Reading baseline test score
Gender indicator variable
Set of race/ethnicity indicator variables
Old-for-grade indicator variable
Set of math and reading dummies indicating which type of test was used for the baseline test score, and interactions of those indicators with baseline math and reading test scores

APPENDIX I

DETAILED TABLES FOR WHAT WORKS CLEARINGHOUSE REVIEW

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

APPENDIX I.1 LOTTERY-BASED RANDOMIZED CONTROLLED TRIAL IMPACTS

In this section, we provide detailed tables to facilitate the review of the lottery-based academic achievement impacts following WWC standards and procedures. We do not include tables for exploratory outcomes on student behavior and attitudes.

Lottery-based impacts of KIPP elementary schools

In Table I.1, we provide the number of students randomly assigned to the intervention and comparison groups and the number of students with data on each outcome. In Table I.2 we present our benchmark, intent-to-treat (ITT) impact estimates for the elementary school sample. Due to the nature of the study design, we are unable to demonstrate the baseline equivalence of the study groups on mathematics or reading achievement.

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

	Full randomized sample		Analytic sample	
Sample definition	Intervention	Comparison	Intervention	Comparison
Woodcock-Johnson III Calculation	290	334	176	195
Woodcock-Johnson III Applied Problems	473	624	282	370
Woodcock-Johnson III Letter-Word Identification	473	624	281	370
Woodcock-Johnson III Passage Comprehension	473	624	280	368

Notes: At schools where the sample randomly assigned by lottery had a treatment-control group size ratio greater than 2:1 (or less than 1:2), a random subsample was drawn from the students randomly assigned by admissions lottery to receive an admission offer or not. This random subsample is reflected in the "full randomized sample" columns. Analytic sample columns reflect the number of students with non-missing outcome scores for each outcome. The full randomized sample and analytic sample for the Calculation test are smaller because that test was administered only to second graders, and only five of the eight study school samples were comprised of students who entered in kindergarten and were thus in second grade in the third year, when the test was administered. (Sample members in the remaining three schools entered admissions lotteries for age 3 pre-kindergarten, or PK3, and were in kindergarten in the third year.)

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

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Woodcock-Johnson III Calculation	Mathematics achievement	0.482	176	0.200	195	0.282**	0.009
Woodcock-Johnson III Applied Problems	Mathematics achievement	0.044	282	-0.030	370	0.074	0.156
Woodcock-Johnson III Letter-Word Identification	Alphabetics	1.005	281	0.757	370	0.248**	<0.001
Woodcock-Johnson III Passage Comprehension	Reading comprehension	0.185	280	-0.038	368	0.223**	0.001

Notes: Test scores are from Woodcock-Johnson subtests administered by the study in spring of the third year following admissions lotteries. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using information from the Woodcock-Johnson national norming sample. Impacts and group means are presented in z-score units. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact, which is estimated controlling for baseline characteristics. Impacts on achievement are estimated using the sample with non-missing outcome data for each measure, but missing baseline characteristics information is imputed using multiple imputation by chained equations. Imputation is conducted separately by treatment group, using standard commands in Stata to generate 20 imputed datasets, calculate and combine impact estimates, and adjust robust standard errors accounting for the number of imputations and the variability of estimated impacts across imputations. The imputation model includes all baseline characteristics that the impact model controls for. The student is the unit of assignment and unit of analysis.

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

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

Lottery-based impacts of KIPP middle schools

In Table I.3, we provide the number of students randomly assigned to the intervention and comparison groups and the number of students with outcome data for each outcome. In Table I.4 we present our benchmark, intent-to-treat (ITT) impact estimates for the middle school sample.

Tables I.3-I.4 are based on samples including imputed data. In Tables I.5 and I.6, we present information on the middle school samples using only the sample of students with non-imputed data. The analysis with non-imputed data is supplemental—our preferred estimates are those presented in table I.4.

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

	Full randomized sample		Analytic sample	
Sample definition	Intervention	Comparison	Intervention	Comparison
Math, Year 1	436	421	313	294
Math, Year 2	436	421	287	268
Math, Year 3	386	369	233	222
Reading, Year 1	436	421	314	294
Reading, Year 2	436	421	291	272
Reading, Year 3	386	369	234	224

Notes: At schools where the sample randomly assigned by lottery had a treatment-control group size ratio greater than 2:1 (or less than 1:2), a random subsample was drawn from the students randomly assigned by admissions lottery to receive an admission offer or not. This random subsample is reflected in the "full randomized sample" columns. Analytic sample columns reflect the number of students with non-missing outcome scores for each outcome. Randomized sample sizes and analysis sample sizes are smaller for year three because one of the sample schools did not provide any outcome data in the third year. Since admissions lotteries randomly assign students to intervention or comparison conditions at the student level, this school is removed from both the full randomized and analytic samples for year three.

Table I.4. Impact of offer of admission to KIPP middle school (ITT)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Math, Year 1	Mathematics achievement	-0.120	313	-0.217	294	0.097*	0.046
Math, Year 2	Mathematics achievement	-0.002	287	-0.245	268	0.244**	<0.001
Math, Year 3	Mathematics achievement	0.010	233	-0.165	222	0.176**	0.008
Reading, Year 1	General literacy achievement	-0.233	314	-0.260	294	0.026	0.584
Reading, Year 2	General literacy achievement	-0.156	291	-0.337	272	0.181**	<0.001
Reading, Year 3	General literacy achievement	-0.134	234	-0.277	224	0.143*	0.014

Notes: Test scores are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using statewide means and standard deviations provided in assessment technical documentation. Impacts represent the cumulative effect of KIPP, not the marginal effect of an additional year. Impacts and group means are presented in z-score units. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact, which is estimated controlling for baseline characteristics. Impacts on achievement are estimated using the sample with non-missing outcome data for each measure, but missing baseline characteristics information is imputed using multiple imputation by chained equations. Imputation is conducted separately by treatment group, using standard commands in Stata to generate 20 imputed datasets, calculate and combine impact estimates, and adjust robust standard errors accounting for the number of imputations and the variability of estimates across imputations. The imputation model includes all baseline characteristics that the impact model controls for. The student is the unit of assignment and unit of analysis.

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

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

Table I.5. Baseline equivalence for lottery-based middle school impacts (sample with non-imputed data)

	Intervention group		Comparison group			
Baseline measure, outcome year	Mean	Sample size	Mean	Sample size	Difference	p-value
Math Pretest, Year 1	-0.091	266	-0.183	263	0.092	0.239
Math Pretest, Year 2	-0.032	242	-0.190	237	0.158*	0.050
Math Pretest, Year 3	-0.043	194	-0.154	192	0.111	0.201
Reading Pretest, Year 1	-0.165	268	-0.243	263	0.078	0.281
Reading Pretest, Year 2	-0.132	246	-0.293	242	0.161*	0.031
Reading Pretest, Year 3	-0.112	196	-0.266	194	0.154	0.067

Notes: Test scores are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using statewide means and standard deviations provided in assessment technical documentation. Differences between treatment and control groups, and group means, are presented in z-score units. The comparison group mean is the unadjusted mean pretest z-score in the comparison group, while the treatment group mean is equal to the comparison group mean plus the difference between groups. The model used to estimate the difference between groups includes school and grade indicators so that intervention-comparison differences are measured between students in the same school and grade. Baseline equivalence is assessed separately for the analysis sample for each outcome year and subject, so that the sample used to assess baseline equivalence for a given outcome year and subject is identical to the sample used to estimate the impact for that outcome year and subject. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in two "Mean" columns.

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

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

Table I.6. Impact of KIPP middle schools (lottery-based estimates for sample with non-imputed data)

Outcome measure	WWC domain	Intervention group		Comparison group		Impact estimate	p-value
		Adjusted mean	Sample size	Mean	Sample size		
Math, Year 1	Mathematics achievement	-0.091	266	-0.212	263	0.121*	0.016
Math, Year 2	Mathematics achievement	-0.003	242	-0.260	237	0.257**	<0.001
Math, Year 3	Mathematics achievement	-0.046	194	-0.161	192	0.115	0.081
Reading, Year 1	General literacy achievement	-0.249	268	-0.261	263	0.012	0.822
Reading, Year 2	General literacy achievement	-0.148	246	-0.351	242	0.204**	<0.001
Reading, Year 3	General literacy achievement	-0.163	196	-0.290	194	0.127	0.053

Notes: Test scores are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Test scores are standardized into z-scores with mean zero and a standard deviation of one, using statewide means and standard deviations provided in assessment technical documentation. Impacts and group means are presented in z-score units. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact, which is estimated controlling for a pre-test (taken in the spring before the lottery) in the same subject as the outcome. Impacts on achievement in each outcome year after the lottery are estimated using the sample of students with a non-missing outcome score for that subject and year, and a non-missing pre-test score from the baseline year in that subject. All regressions use robust standards errors. The student is the unit of assignment and unit of analysis.

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

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

APPENDIX I.2 MATCHED STUDENT IMPACTS USING NON-IMPUTED SAMPLES

In this section, we provide detailed tables to facilitate the review of the matched-student academic achievement and attainment impacts following WWC standards and procedures.

Matched-student impacts of KIPP middle schools (supplemental analysis)

In this section we present supplemental information on the quasi-experimental matched-student estimates of the impact of KIPP middle schools. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated only on the sample of students with non-imputed data. In Table I.7, we present baseline equivalence information for the sample of students in the matched-student analysis with non-missing baseline and outcome data. In Table I.8, we present estimates of the impact of KIPP middle schools using the same sample. In Table I.9, we present baseline equivalence results for the sub-sample of KIPP middle schools that opened in fall 2011 or later (new middle schools). In Table I.10, we present estimates of the impact of KIPP middle schools using the same sample of new middle schools.

Table I.7. Baseline equivalence for matched-student middle school impacts (sample with non-imputed data)

	Intervention group		Comparison group			
Baseline measure (outcome sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (reading year 1)	-0.102	17,518	-0.084	17,397	-0.018	0.101
Reading scores (reading year 2)	-0.087	14,079	-0.085	13,679	-0.002	0.897
Reading scores (reading year 3)	-0.049	11,318	-0.045	10,837	-0.004	0.761
Reading scores (reading year 4)	-0.007	7,430	-0.035	7,121	0.027	0.149
Math scores (math year 1)	-0.095	17,525	-0.082	17,413	-0.013	0.244
Math scores (math year 2)	-0.070	14,064	-0.082	13,672	0.012	0.413
Math scores (math year 3)	-0.043	11,185	-0.060	10,741	0.017	0.259
Math scores (math year 4)	-0.065	6,887	-0.083	6,737	0.018	0.386
Reading scores (history)	0.022	4,866	-0.013	4,896	0.034	0.112
Math scores (science)	-0.053	8,826	-0.059	8,587	0.006	0.704

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in two "Mean" columns. Data shown in this table do not include imputed values for any baseline variables.

Table I.8. Impact of KIPP middle schools (matched-student estimates for sample with non-imputed data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading year 1	General literacy achievement	-0.105	17,518	-0.111	17,397	0.005	0.426
Reading year 2	General literacy achievement	-0.008	14,079	-0.113	13,679	0.105**	<0.001
Reading year 3	General literacy achievement	0.062	11,318	-0.092	10,837	0.154**	<0.001
Reading year 4	General literacy achievement	0.076	7430	-0.085	7,121	0.161**	<0.001
Math year 1	Mathematics achievement	-0.051	17,525	-0.109	17,413	0.057**	<0.001
Math year 2	Mathematics achievement	0.091	14,064	-0.141	13,672	0.232**	<0.001
Math year 3	Mathematics achievement	0.170	11,185	-0.121	10,741	0.291**	<0.001
Math year 4	Mathematics achievement	0.136	6,887	-0.131	6,737	0.268**	<0.001
History		0.107	4,866	-0.131	4,896	0.238**	<0.001
Science	Science achievement	0.084	8,826	-0.166	8,587	0.249**	<0.001

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. Reported impacts are an average of equally weighted impact estimates for each KIPP middle school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. Impacts represent the cumulative effect of KIPP after the noted number of years after admission for math and reading scores, not the marginal effect of an additional year. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. The grade level of middle school exams used for history and science outcomes varied by jurisdiction. We selected the highest middle school grade level where science or social studies was observed for more than one cohort of KIPP students. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Table I.9. Baseline equivalence for matched-student middle school impacts (sample with non-imputed data) for new KIPP middle schools

	Intervention group		Comparison group			
Baseline measure (outcome sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (reading year 1)	-0.323	1,195	-0.238	1,165	-0.085*	0.021
Reading scores (reading year 2)	-0.301	586	-0.242	532	-0.059	0.289
Math scores (math year 1)	-0.272	1,196	-0.195	1,170	-0.078*	0.036
Math scores (math year 2)	-0.225	587	-0.223	536	-0.002	0.973

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. Due to rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in two "Mean" columns. Data shown in this table do not include imputed values for any baseline variables.

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

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

Table I.10. Impact of new KIPP middle schools (matched-student estimates for sample with non-imputed data)

	Intervention group		Comparison group			
Outcome measure	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading year 1	-0.220	1,195	-0.269	1,165	0.049*	0.028
Reading year 2	-0.123	586	-0.243	532	0.120**	0.000
Math year 1	-0.188	1,196	-0.228	1,170	0.039	0.069
Math year 2	-0.063	587	-0.289	536	0.226**	0.000

Notes: Test scores are standardized within each middle school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. Reported impacts are an average of equally weighted impact estimates for each KIPP middle school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. Impacts represent the cumulative effect of KIPP after the noted number of years after admission for math and reading scores, not the marginal effect of an additional year. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Matched-student impacts of KIPP high School on new entrants in Grade 9 (supplemental analysis)

In this section we present supplemental information on the quasi-experimental matched-student estimates of the impact of KIPP high schools for new entrants in grade 9. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated only on the sample of students with non-imputed data. In Table I.11 we present baseline equivalence information for the sample of students in the matched-student analysis with non-missing baseline and outcome data. In Table I.12, we present estimates of the impact of KIPP high schools using the same sample.

Table I.11. Baseline equivalence for matched-student high school impacts (sample with non-imputed baseline data)

	Intervention group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	-0.060	887	-0.089	861	0.028	0.591
Math scores (math)	-0.084	725	-0.059	691	-0.026	0.631
Reading scores (social studies)	0.000	304	0.001	297	-0.002	0.982
Math scores (science)	-0.111	656	-0.109	643	-0.002	0.973
Reading scores (4-year graduation)	-0.014	426	-0.125	426	0.111	0.099
Math scores (4-year graduation)	-0.028	426	-0.155	426	0.127	0.088
Free and reduced-price lunch status (4-year graduation)	0.806	426	0.809	426	-0.003	0.924

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Because of rounding, the value reported in the “Difference” column may differ slightly from the difference between the values reported in two “Mean” columns. Data shown in this table do not include imputed values for any baseline variables.

Table I.12. Impact of KIPP high schools for new entrants in grade 9 (matched-student estimates for sample with non-imputed baseline data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
ELA achievement	General literacy achievement	0.111	887	-0.065	861	0.175**	<0.001
Mathematics achievement	Mathematics achievement	0.236	725	-0.036	691	0.273**	<0.001
Science achievement	Science achievement	0.105	656	-0.223	643	0.328**	<0.001
Social studies achievement	Social studies achievement	-0.134	304	-0.149	297	0.015	0.795
Four-year high school graduation	Completing school	0.707	426	0.672	426	0.035	0.355

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Matched-student cumulative impacts of KIPP middle and high schools (supplemental analysis)

In this section we present supplemental information on the quasi-experimental matched-student estimates of the cumulative impact of KIPP middle and high schools. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated only on the sample of students with non-imputed data. In Table I.13 we present baseline equivalence information for the sample of students in the matched-student analysis with non-missing baseline and outcome data. In Table I.14, we present estimates of the cumulative impact of KIPP middle and high schools using the same sample.

Table I.13. Baseline equivalence for matched-student cumulative middle and high school impacts (sample with non-imputed baseline data)

	Intervention group		Comparison group			
Baseline measure (analyzed outcome for this sample)	Mean	Sample size	Mean	Sample size	Difference	p-value
Reading scores (ELA)	0.065	2,072	0.100	1,929	-0.035	0.261
Math scores (math)	-0.032	1,516	0.037	1,414	-0.069	0.083
Reading scores (social studies)	0.117	788	0.109	707	0.007	0.871
Math scores (science)	0.065	1,946	0.133	1,636	-0.067	0.078
Reading scores (4-year graduation)	0.037	1,025	0.055	1,008	-0.018	0.658
Math scores (4-year graduation)	0.019	1,025	0.080	1,008	-0.061	0.138
Free and reduced-price lunch status (4-year graduation)	0.883	1,025	0.892	1,008	-0.009	0.562

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Baseline tests are from statewide assessments collected through administrative records requested from each state or jurisdiction in the sample. The outcome sample for each baseline characteristic is noted in parentheses next to the baseline measure. Sample means are calculated separately for each KIPP school, and the average reported assigns an equal weight to each of the school-level means. No baseline differences are significant at the 0.05 level, two-tailed test. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in two "Mean" columns. Data shown in this table do not include imputed values for any baseline variables.

Table I.14. Cumulative impact of KIPP middle and high schools (matched-student estimates for sample with non-imputed baseline data)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
ELA achievement	General literacy achievement	0.381	2,072	0.085	1,929	0.295**	<0.001
Mathematics achievement	Mathematics achievement	0.341	1,516	-0.001	1,414	0.341**	<0.001
Science achievement	Science achievement	0.417	788	0.001	707	0.417**	<0.001
Social studies achievement	Social studies achievement	0.180	1,946	-0.087	1,636	0.267**	<0.001
Four-year high school graduation	Completing school	0.785	1,025	0.651	1,008	0.134**	<0.001

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are from end-of-course (e.g., algebra) or end-of-grade (e.g., grade 10 mathematics) high school exams collected through administrative records that were requested from each state or jurisdiction in the sample. High school graduation is a binary variable. Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors and the student is the unit of assignment and unit of analysis. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

APPENDIX I.3 MATCHED-SCHOOL IMPACTS

In this section, we provide detailed tables to facilitate the review of the matched-school academic achievement impacts following WWC standards and procedures. We do not provide tables for exploratory outcomes on student behavior and attitudes. These estimates differ from those presented in the body of the report (our preferred estimates) in that they are estimated separately for the sample of schools in our adjacent cohorts and matched-middle schools models. Both models compare outcomes for KIPP middle school students who had an option to attend a KIPP high school with those who did not have an option to attend a KIPP high school.

Matched-school impacts of opportunity to attend a KIPP high school (adjacent cohorts model)

The first model (Same KIPP Middle School, Adjacent Cohorts), focuses on a set of KIPP high schools in their first year of operation. The treatment group in this model includes 8th-grade KIPP middle school students who had the option to attend the local KIPP high school in the first year of its operation. The comparison group includes the previous cohort of 8th-grade KIPP students from the same middle school who did not have the option to attend the local KIPP high school because it had not yet opened. Because sample sizes are too small to restrict the comparison group, we do not employ student-level matching in this model. In Table I.15 we present baseline equivalence information for the sample of students in the matched-school analysis for the adjacent cohorts sample with non-missing baseline and outcome data. In Table I.16, we present estimates of the impact of an opportunity to attend a KIPP high school using the same sample.

Table I.15. Baseline equivalence for matched-school impacts of opportunity to attend a KIPP high school (adjacent cohorts sample)

Baseline Characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (reading achievement, language achievement)	0.425	0.357	0.068	0.409	208	213
Math scores (mathematics achievement)	0.772	0.800	-0.028	0.718	208	213

Note: Baseline tests are either the SAT-10 or the MAP, depending on what was offered in that jurisdiction. They are presented in z-score units. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution. The relevant outcome sample for each baseline characteristic is noted in parentheses. All values in this table are based on the sample for which we have test outcome data. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

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

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

Table I.16. Impact of opportunity to attend a KIPP high school (adjacent cohort estimates)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading achievement	Reading comprehension	0.220	208	0.230	213	-0.010	0.844
Language achievement	General literacy achievement	0.081	208	0.081	213	0.000	0.995
Mathematics achievement	Mathematics achievement	0.026	208	-0.056	213	0.082	0.217

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are study-administered TerraNova assessments (Form G, Level 21/22). Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. The student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

Matched-school impacts of opportunity to attend a KIPP high school (matched middle schools model)

In the second model (Same Cohort, Matched Middle Schools), the treatment group includes 8th-grade KIPP middle school students in 2008-09 who had the option to attend the local KIPP high school. The comparison group includes 8th-grade KIPP students from different middle schools in 2008-09 in regions with no KIPP high school open at the time. To define a sample that was equivalent at baseline (grade 8), we first identified a set of comparison KIPP middle schools that most resembled the feeder KIPP middle schools on the basis of average school-level characteristics (race/ethnicity, baseline achievement on a nationally-normed test, and baseline test instrument—either the SAT-10 or the MAP). Then, within these matched sets of schools, we conducted student-level propensity score matching to identify the individual comparison student who was the closest match to each treatment student on the basis of gender, race/ethnicity, baseline achievement in reading and math, and whether the student was old for his/her grade. In Table I.17 we present baseline equivalence information for the sample of students in the matched-school analysis for the matched middle schools sample with non-missing baseline and outcome data. In Table I.18, we present estimates of the impact of an opportunity to attend a KIPP high school using the same sample.

Table I.17. Baseline equivalence for matched-school impact of opportunity to attend a KIPP high school (matched middle schools sample)

Baseline Characteristic	Treatment	Comparison	Difference	P-value	# KIPP	# Comparison
Reading scores (reading achievement, language achievement)	0.131	0.140	-0.009	0.926	256	256
Math scores (mathematics achievement)	0.568	0.561	0.007	0.955	256	256

Note: Baseline tests are either the SAT-10 or the MAP, depending on what was offered in that jurisdiction. They are presented in z-score units. Z-scores are standardized with a mean of 0 and a standard deviation of 1 relative to a nationally representative distribution. The relevant outcome sample for each baseline characteristic is noted in parentheses. All values in this table are based on the sample for which we have test outcome data. Because of rounding, the value reported in the "Difference" column may differ slightly from the difference between the values reported in the "KIPP" and "Comparison" columns.

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

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

Table I.18. Impact of opportunity to attend a KIPP high school (matched-middle school estimates)

		Intervention group		Comparison group			
Outcome measure	WWC domain	Adjusted mean	Sample size	Mean	Sample size	Impact estimate	p-value
Reading achievement	Reading comprehension	0.255	256	0.095	256	0.160*	0.026
Language achievement	General literacy achievement	0.069	256	-0.054	256	0.123	0.152
Mathematics achievement	Mathematics achievement	0.071	256	-0.065	256	0.136	0.099

Note: Test scores are standardized within each high school jurisdiction with a mean of 0 and a standard deviation of 1. They are presented in z-score units. Outcome tests are study-administered TerraNova assessments (Form G, Level 21/22). Reported impacts are an average of equally weighted impact estimates for each KIPP high school in the sample—using regressions of the relevant outcome variable on a treatment indicator and other covariates and adjusting for students' baseline test scores in reading and math and students' demographic characteristics. All regressions use robust standards errors. The comparison group mean is the unadjusted mean outcome z-score in the comparison group, while the treatment group adjusted mean is equal to the comparison group mean plus the impact estimate. The student is the unit of assignment and unit of analysis. Data shown in this table do not include imputed values for any baseline or outcome variables.

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

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

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

www.mathematica-mpr.com

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

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

MATHEMATICA
Policy Research