

Do Charter School Networks Deflate as They Expand? Trends in the Impacts of KIPP Schools during a Period of Rapid Growth in the KIPP Network

Ira Nichols-Barrer, *Mathematica Policy Research*

Phil Gleason, *Mathematica Policy Research*

Christina Tuttle, *Mathematica Policy Research*

Thomas Coen, *Mathematica Policy Research*

Virginia Knechtel, *Mathematica Policy Research*

Background / Context:

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, 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 1).

Prior studies (Gleason et al. 2014; Tuttle et al. 2013; Angrist et al. 2010; Woodworth et al. 2008) have consistently found that attending a KIPP middle school positively affects student achievement. 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 (Gleason et al. 2014). 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.

Questions remain, however, about the ability of charter school networks like KIPP to remain effective as they expand and serve larger numbers of students. Critics of charter schools argue that their positive effects may be driven by creaming of the most motivated and least disruptive students, either at entry or through selective attrition (Carnoy 2005; Kahlenberg 2011; Miron et al. 2011). To the extent that student selection at entry or through attrition plays a role in helping KIPP achieve positive impacts, this sort of selection could be more difficult to accomplish as the network grows. For example, if KIPP schools are able to be choosy in determining which students to admit, they would presumably be less choosy if they had to fill ten middle schools in a city than if they only had to fill a single school. Thus, if student selection plays a role in KIPP's positive impacts, then these impacts may decline as the network adds schools.

Staffing may also be a challenge in an expanding charter school network. In the above example, KIPP would have to identify and hire ten school principals capable of effectively leading KIPP middle schools in the city, and fill classrooms of these schools with effective teachers. Again, a failure to meet this type of staffing challenge in a growing charter school network could lead the positive impacts of a network like KIPP declining or completely disappearing.

Purpose / Objective / Research Question / Focus of Study:

In this paper, we present estimates of the impacts of KIPP middle schools during a period in which the network greatly expanded. In particular, we measure KIPP impacts between 2005 and 2014, a period in which the number of schools in the network increased from 40 to 140. We are exploring whether there is any evidence of an increase, decline, or stability in KIPP impacts over this period. To the extent that impacts vary over time, we also would like to understand the extent to which variation in network-wide impacts is driven by changes to the composition of the network versus changes over time in the effectiveness of existing schools. In particular, we address the following questions:

1. What is the average impact of KIPP middle schools in the network in each year between 2005 and 2014 on students' achievement on reading and math state assessments after two years of potential exposure to KIPP?
2. To what extent are differences in impacts over time driven by changes to the composition of the KIPP network—new schools coming into the network being more or less effective than existing schools? To what extent are differences in impacts over time driven by changes in the effectiveness of individual schools as they gain experience? Are the most recently opened KIPP middle schools—those opened during the period of expansion between 2011 and 2014 supported by the U.S. Department of Education's Investing in Innovation Scale-Up Grant—more or less effective during their first two years of operation than previously opened KIPP middle schools were during their initial years of operation?

Setting:

The paper focuses on 37 KIPP middle schools open during the 2005-2014 period in 14 KIPP regions and 11 states. Most of the KIPP middle schools in this sample are located in large urban school districts.

Population / Participants / Subjects:

The study sample includes a treatment group of students who entered a KIPP school for the first time in grade 5 or grade 6, and a matched comparison group of students in the same school districts who did not attend KIPP. In total, we analyze a sample of 20,312 treatment students and 20,312 matched comparison students. Nearly all KIPP middle school students are eligible for free- or reduced-price meals, and a substantial majority are either African American or Hispanic.

Intervention / Program / Practice:

The study examines the effectiveness of KIPP middle schools over time. The national network of KIPP middle schools tend to use a number of distinguishing practices (although these practices may be implemented to varying degrees across different KIPP schools and KIPP regions in the network). These practices include: a longer school day; a longer school year; strong student behavior policies with rewards and sanctions; contracts between students, parents, and teachers; substantial autonomy for KIPP principals; and close monitoring of school performance in terms of student achievement.

Research Design:

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

schools.¹ 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).

Data Collection and Analysis:

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

Findings / Results:

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

¹ 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.

earlier years than recent years (Figure 2). 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 newly-opened KIPP schools supported by the network's Investing in Innovation scale-up grant from the Department of Education, 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 3). 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.

Conclusions:

In sum, we find evidence that the KIPP network of middle schools was able to sustain a pattern of statistically significant and positive impacts throughout a period of rapid growth. Although the magnitude of these positive impacts declined over several of the years in our data, the size of these impacts recovered in the most recent years we observe for this sample of middle schools. For the KIPP network as a whole, this pattern of impacts suggests that KIPP middle schools continued to produce positive impacts on student achievement outcomes even as the organization faced the pressures of growth in terms of student recruitment, hiring new principals and teachers, and managing turnover in the network's workforce over time.

Appendices

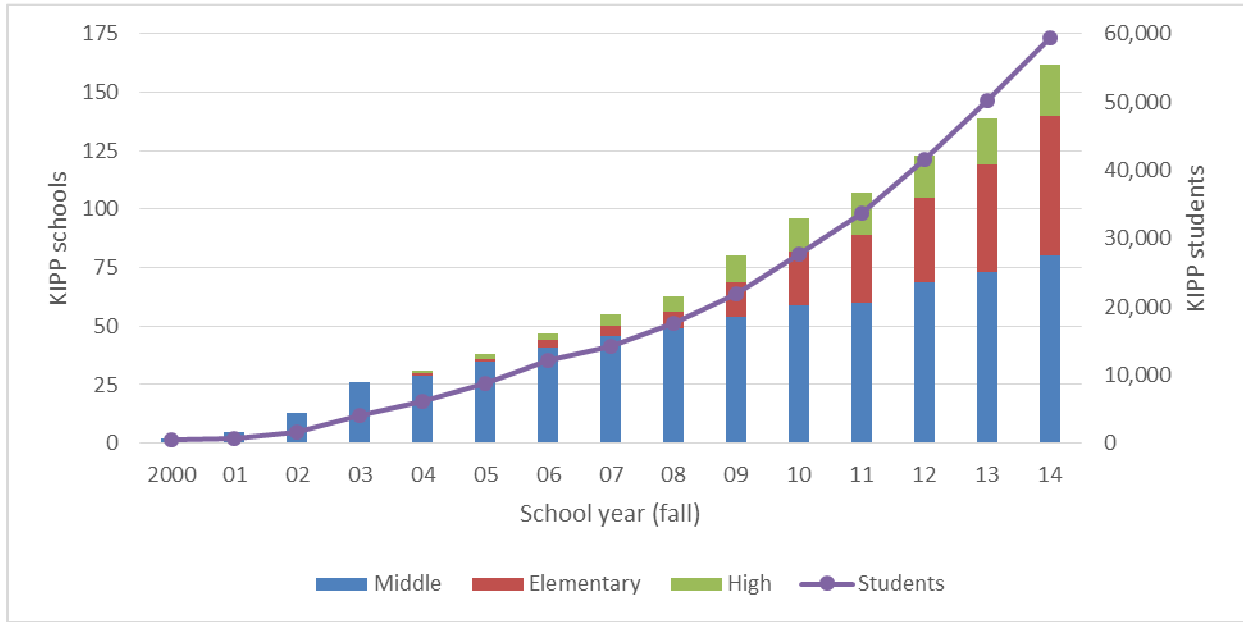
Appendix A. References

- 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.
- 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.
- 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.

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.

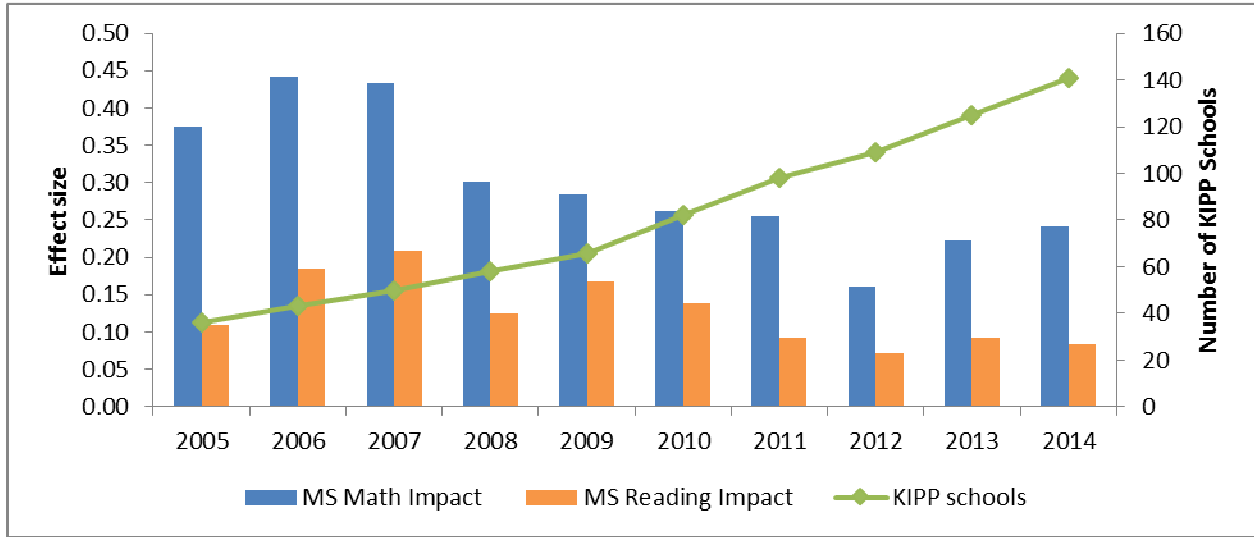
Appendix B. Tables and Figures

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



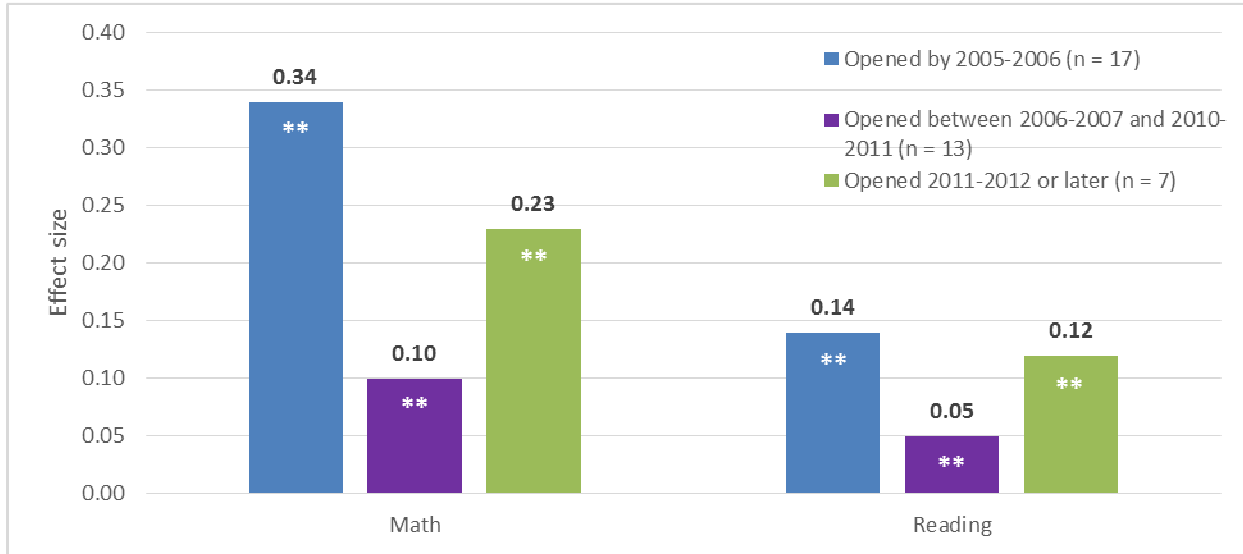
Source: KIPP Foundation.

Figure 2. 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.

Figure 3. 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.