



## **What Do We Know About the Outcomes of KIPP Schools?**

**Jeffrey R. Henig, Ph.D.**

**Teachers College, Columbia University**

**The Great Lakes Center for Education Research & Practice**

PO Box 1263

East Lansing, MI 48826

Phone: (517) 203-2940

Email: [greatlakescenter@greatlakescenter.org](mailto:greatlakescenter@greatlakescenter.org)

Web Site: <http://www.greatlakescenter.org>

**November 2008**

One of a series of Policy Briefs made possible by funding from the Great Lakes Center for Education Research and Practice.

EPIC/EPRU policy briefs are peer reviewed by members of the Editorial Review Board. For information on the board and its members, visit: <http://epicpolicy.org/editorial-board>

# What Do We Know About the Outcomes of KIPP Schools?

Jeffrey R. Henig, Teachers College, Columbia University

## *Executive Summary*

Although many consider KIPP (Knowledge is Power Program) to be a prototype charter school operator that merits expansion and replication, few systematic and independent assessments of it are available. A large-scale study incorporating a randomized design has begun, but until that study is complete policymakers need to learn what they can from existing evidence. This brief reviews seven studies that attempt to answer questions concerning whether and to what degree KIPP schools raise student achievement. In weighing the evidence, it is necessary to consider issues raised by two different types of validity. The first is *internal validity*: Are the inferences about cause and effect in a given study well grounded? The second is *external validity*: Which lessons from studies of particular schools might apply more generally across schools in other locations and further in the future, providing useful guideposts for policy?

## **Conclusions**

- The weight of the evidence suggests that students who enter and stay in KIPP schools tend to perform better than similar students in more traditional public schools.
- This does not appear to be attributable to a selective admissions process. KIPP serves minority and high-need students, many of whom performed poorly before they entered the schools. Some unobservable biases may be present in student motivation and support, but except for a tendency to attract more girls than boys, there is as yet no strong observable evidence of a systematic selection bias.
- Where it has been monitored, student attrition is high and seemingly selective. Those who leave KIPP tend to have been performing less well than those who stay, and at least one study suggests that those who leave were lower-performing when they entered. Such attrition, if it were taken into consideration, would reduce the size of gains in reports that simply compare KIPP eighth graders with those in their host districts. However, the evidence does not go so far as to suggest that attrition fully accounts for the observed KIPP advantage.
- Most of the studies are limited to early KIPP schools and students in their first or second year. Studies that follow cohorts over time seem to show that gains persist, but there is no evidence that early gains grow into progressively higher gains in later years.
- Few studies look deeply inside the KIPP process; those that do show that teacher enthusiasm is high but that demands on teachers and leaders are great,

- resulting in high turnover and an unrelieved pressure to find and train new people. The implications for the expansion and sustainability of the KIPP model are still not clear.

## **Recommendations**

- Policy makers at all levels of government should pay attention to KIPP and consider it a possible source of information and guidance for their decisions.
- Although KIPP may yield useful information, policymakers and others should temper their interest in the operation with wariness and realistic expectations. There are significant unanswered questions about how expansion might affect outcomes, especially in relation to the difficulty of sustaining gains dependent upon KIPP's heavy demands on teachers and school leaders. Moreover, it is not realistic to think that the KIPP model is a panacea for distressed systems. It is possible that only a small proportion of students and families will be able to meet the demands KIPP imposes on them; even those enthused when they begin the KIPP regimen tend to leave in high numbers.
- Policymakers, accordingly, should treat KIPP schools as potential tools that may contribute to—but not substitute for—systemic improvement.
- Policymakers should be aware that KIPP has prompted some district interest in longer school days, weeks, and years. However, an extended schedule sometimes brings parental objections as well as potential taxpayer objections to the additional expense. With no strong evidence yet linking extended scheduling to KIPP success, policymakers might best encourage it as a school-level (rather than district-wide) option while concurrently promoting a combination of experimentation and careful analysis of consequences.
- Researchers should help provide better data on patterns of movement in and between charter schools and traditional public schools, including information on why students leave and how their mobility affects student and school-level performance.

# What Do We Know About the Outcomes of KIPP Schools?

Jeffrey R. Henig, Teachers College, Columbia University

When it comes to getting good press, the Knowledge is Power Program (KIPP) charter school operation is second to none. KIPP's identifying features include high academic expectations, parent and student commitment, substantially more time in school, school-level decision-making, and a focus on measurable outcomes. Many hold it up as a model—not just as an interesting exemplar, but as a prototype that others should dissect for lessons about improving both charter schools and traditional school systems. A September 8, 2008, search of newspapers on Proquest Direct uncovered 334 articles mentioning KIPP and charter schools in either their citations or abstract, many with positive claims embedded in the headlines. “A Miracle in the Making?” one asked.<sup>1</sup>

Most of the testimonials are based on school visits, hearsay, or simple comparisons of overall test scores. These should not be totally discounted. The fact that something is supported by conventional wisdom and word-of-mouth does not in itself suggest that it is wrongheaded. But in the era of scientifically based school reform, good press cannot long substitute for objective evidence.

When news reports do cite data in support of the KIPP model, they often draw on the “Report Card” the KIPP Foundation prepares detailing demographics and achievement at its individual schools. The 2007 Report Card, for example, provides data on the 49 schools that were open during the 2006-07 school year.<sup>2</sup> According to this document, about two-thirds of KIPP fifth-grade classes outperform their local districts on state reading and mathematics exams; by eighth grade, this percentage rises to a dramatic 100%. On nationally normed tests, the Report Card indicates that students who stay in KIPP for four years move from the 32<sup>nd</sup> to 60<sup>th</sup> percentile in reading and from the 40<sup>th</sup> to the 82<sup>nd</sup> percentile in math.<sup>3</sup> The KIPP Report Card is highly useful, and the KIPP Foundation should be commended for making more information more readily available and more easily interpreted than do most schools, whether public, charter, or private. But the Report Card does not offer the kind of careful analysis needed to address a range of validity questions that researchers know must be considered in order to determine whether scores and gains can be legitimately attributed to the school. And externally conducted evaluations tend to be given more weight than an organization's own reports.

On its website, the KIPP Foundation indicates: “To date, KIPP has been the subject of five independent reports that have examined KIPP's effect on student achievement. Those reports have found that students make significant academic gains while at KIPP schools.”<sup>4</sup> In this brief, I review these five studies as well as updates for two of them. In addition, I include two studies of similar caliber that KIPP does not mention on its website.<sup>5</sup>

These seven studies put heavy emphasis on test scores, particularly those in reading and mathematics, and their methodological discussion largely focuses

## What Do We Know About the Outcomes of KIPP Schools?

on issues of internal validity (the question of whether inferences about cause and effect are well grounded). In what follows, I first detail the key criteria that I will apply in determining how successfully these studies have addressed the narrow—but important—question of whether the particular KIPP schools seriously studied to date can reasonably claim responsibility for gains in reading and math achievement. Then, I summarize the studies’ key design elements and findings.

In the final section I turn to issues relating to external validity, which asks whether the findings for a particular school might apply more generally across schools in other locations and further into the future, providing useful guideposts for policy. To the extent that the KIPP schools studied so far seem to be working, what lessons relevant to policy and practice can we draw? Can the core elements responsible for good results be isolated? Are they sustainable within the KIPP program as it expands? What do they tell us about the charter school movement overall? Can the successful aspects of the KIPP model be transferred to traditional public schools? Taken as a whole, the existing studies themselves give scant attention to such questions. Yet, these are the most critical to understand before we can move from the existing evidence to confident policy suggestions.

### *Assessing Internal Validity: What to Look For*

Internal validity relates to the capacity of a study to make strong claims about cause and effect. For example, if a study finds that scores went up (or down) at a KIPP school, how likely is it that other factors, rather than KIPP’s program and practices, are responsible? Randomized field trials (RFTs) are widely considered to be the “gold standard” in research design precisely because of their strengths in addressing internal validity. When well-constructed and successfully implemented, RFTs eliminate rival explanations for observed changes by comparing outcomes for two groups—one that experiences the intervention under study, and a second (control) group which differs from the first only in that its members do not experience the intervention.<sup>6</sup>

In March 2008, KIPP announced that it would be working with Mathematica Policy Research, Inc. on a nearly \$4 million study to assess the program’s success. The planned study includes both an experimental component (comparing admission lottery winners to losers at oversubscribed schools) as well as a non-experimental component in which KIPP middle school students will be compared with matched students at nearby traditional schools. When completed, this study will likely provide the most authoritative information, but the final report will not be available for at least five years. Even then, some questions relevant to policymakers and parents may not be answered, since results will not be available for individual schools.

In the interim, policymakers and parents face decisions with both social and personal consequences. As Carol Weiss suggests, the pertinent question now is “what to do until the random assigner comes.”<sup>7</sup> Her answer is to make every effort to eliminate credible alternative explanations when a study finds that a particular intervention has produced particular effects. For the studies under consideration here, that means that before anyone accepts the claim that the KIPP

## What Do We Know About the Outcomes of KIPP Schools?

program itself is responsible for gains in student achievement, other possible explanations for the gains must be considered. Three alternative possibilities merit investigation: Might reported gains have been affected by the selection of a particular kind of student for the schools? Might they have been affected by the absence of students who left the schools? Is it possible that the way students were tested and the way test results were reported influenced findings?

### Who Enrolls?

The issue of selection bias has persistently plagued research on various forms of school choice. Early studies indicating that private school students outperformed public school students, even controlling for socioeconomic status, could never quite shake the suspicion that families aggressive enough to pursue a private option had something different about them in terms of motivation and support—suggesting that the family might have been more responsible than the school for the students' academic success.<sup>8</sup> In the case of charter schools, both opponents and proponents have raised the specter of selection bias to challenge findings with which they disagree. Confronting studies that show a charter school advantage, critics argue that families sufficiently motivated to seek out charter schools and to meet the expectations they frequently impose are likely to have unmeasured qualities that will continue to help their children perform well. Confronting studies that suggest charter schools do not perform well, proponents argue that charter schools accept students with greater unmeasured and ongoing disadvantages that prompted the move away from their former school, so that in some cases lower scores result from student characteristics, not the school's program.

In one of the few critical analyses of KIPP, Carnoy et al. focus largely on the issue of selection bias.<sup>9</sup> Using New York State assessment scores from 2002, they compared the fourth-grade reading scores of the class entering the KIPP-Bronx Academy 5<sup>th</sup> grade in 2002 to those of their peers in 31 schools within a two-mile radius. The KIPP-bound students had higher fourth-grade scores than all but one of the other classes; overall, 42% of entering KIPP students had passed the fourth grade reading test versus 28% from the neighboring schools. In addition, through interviews with a small and not necessarily representative sample of teachers in KIPP-feeder schools in NYC, Washington DC, and Houston, Carnoy et al. found indications that teachers may promote KIPP to students they consider more able and motivated—or whose families they believe more likely to be motivated and supportive.<sup>10</sup>

Randomization is the best way to eliminate the possibility that differences in the students who enroll in KIPP, rather than what happens inside a KIPP school, account for any differences in outcomes. While none of the existing studies used a randomized design, all made some effort to assess whether KIPP students are racially, economically, or academically distinguishable. However, these efforts range from *ad hoc* and *post facto* attempts to compare basic descriptive characteristics of student groups to more sophisticated efforts to match individual KIPP students to comparison students based on race, class, and prior

## What Do We Know About the Outcomes of KIPP Schools?

test scores. In general, studies that have more than one year of pre-KIPP student-level performance data are in a better position to disentangle KIPP effects from the likely influence of students' characteristics and prior performance levels when entering the school.

### Who Leaves?

Less attended to, but potentially just as important, is the issue of selection *out* of KIPP schools. Despite survey results showing parents are highly satisfied with the KIPP experience, a substantial number of students either leave during the academic year or fail to return after the summer. Jay Mathews, a *Washington Post* reporter who has closely followed the KIPP story and reported it in mostly positive terms, acknowledges that attrition has become a focal point for attention and concern. "At the KIPP Bridge College Preparatory school in Oakland," he observed, "of the 87 students who enrolled in fifth grade in 2003, 32 later moved out of the area and 30 had parents who decided to remove them from KIPP for other reasons."<sup>11</sup>

If those who leave KIPP schools are disproportionately those who are struggling academically, or whose families lack the supportive attributes that often predict success, then their absence could account for higher test scores or group gains over time. Very few of the existing studies have good data on student attrition to address this rival hypothesis.

### Testing, Measurement, and Proficiency Levels

The existing studies primarily emphasize standardized test scores, particularly in reading and mathematics, as measures of academic achievement. There are reasons to be wary of over-reliance on standardized tests,<sup>12</sup> a point to which I return when I discuss external validity. However, even if standardized test scores were accepted as a legitimate indicator, researchers should attend to some potential internal validity issues. Four questions arise: (a) existing studies vary in the *type* of exam implemented; (b) scores can be for *individual students* or *aggregated*; (c) comparisons can be *cross-sectional* (at a single point in time) or *longitudinal* (across some time span); and, (d) test results can be reported as *scale scores* or *proficiency levels*.

KIPP students all take whatever city or state accountability tests are required in their location. In addition, the KIPP Foundation requires schools to administer a nationally norm-referenced test. Individual schools have discretion about which test to use, but most employ the Stanford Achievement Test (SAT 9 or 10). Both local and national tests have analytical advantages and disadvantages.

Since all students in a district must take the same state- or locally mandated exams, studies that use these tests as measures of achievement can compare performance of KIPP students to similar students in the surrounding neighborhoods. However, state tests vary in difficulty. Critics charge that some states are adopting easier tests, in part in order to more easily meet NCLB

## What Do We Know About the Outcomes of KIPP Schools?

proficiency requirements. These tests may not be able to identify meaningful difference in student performance above some low threshold. In addition, the high-stakes nature of these exams, which can determine whether schools are subject to sanctions, can produce pressure for teachers to “teach to the test”—or even for teachers or principals to cheat.<sup>13</sup> Studies that find KIPP (or district) schools have better achievement outcomes based on these exams may be reflecting test-focused tactics in the schools (limiting time spent on non-tested subjects, for example, or repeated practice on the format of test questions), while at the same time they offer little or no information about the broader kinds of learning presumed to be schools’ real goal.

Nationally normed tests, like the SAT 9 or 10 (9<sup>th</sup> or 10<sup>th</sup> edition), make it possible to compare KIPP students’ performance with the performance of students nationally. They also eliminate the apples-to-oranges comparison problem of state tests, allowing for cross-state comparisons of KIPP schools. If state-mandated tests are too easy, a nationally standardized test can provide a reality check. As we shall see, most of the KIPP studies conducted to date use *either* a state-mandated, criterion-based exam or a nationally normed one, although some have the added discernment provided by employing both.

A second measurement area that raises questions is that some studies used individual student test scores while others used scores aggregated at the grade or school level. Student-level data are much more difficult to obtain; researchers in pursuit of this information depend upon the kindness of strangers to provide it: either KIPP officials or state or district agencies that oversee mandated exams. Such student-level data are far superior, however, from the standpoint of sorting out effects of the KIPP experience from factors relating to family, neighborhood, and prior learning.

A third measurement area needing attention is whether test scores for KIPP and comparison groups include gains made over time or are simply cross-sectional snapshots. As researchers try to rule out the possibility that student characteristics at KIPP entry (rather than the KIPP program) account for gains in performance, studies that rely on multiple tests administered to the same students and schools are better able than simple cross-sectional comparisons to rule out the possibility that differences are due to students’ achievement and learning curves that may have existed prior to enrollment. Most useful are studies in which researchers have access to students’ gain scores during the year or two before they enter KIPP.

A final issue relating to testing and internal validity concerns the distinction between comparing scaled scores themselves and comparing the percentage of students who meet some proficiency level based on the scores. State and national policies put heavy priority on raising the percentage of students scoring above a state-defined proficiency threshold, and as a result data on proficiency levels are more readily available than scaled scores. However, schools can focus their attention on students just below the proficiency cut-off (widely known as “bubble students”), significantly raising proficiency rates without improving overall student performance levels much—or at all.<sup>14</sup>

## What Do We Know About the Outcomes of KIPP Schools?

**Table 1. Empirical Studies of KIPP Outcomes: Design Features**

Author	Date	Number of schools	Geographic focus	Level of analysis	Longitudinal?	Comparison group
Doran & Drury	2002	3	DC; Gaston, NC; Houston	student	First year gains	district aggregate; national norms
	<b>Major Conclusion:</b> "Test scores improved at impressive rates....these gains were reflected across demographic subgroups and exceeded those achieved by these same students in the year prior to their enrollment."					
Ross & others	Multiple	1	Memphis	student	Four years; four cohorts	matched students at nearby and similar schools
	<b>Major Conclusion:</b> Year 4 results "revealed fairly positive outcomes," with authors speculating that leadership instability had disrupted progress.					
EPI	2005	24	National (In states with KIPP schools; 24 of 31)	school level	Single year (Fall to Spring or Fall to Fall)	national norms
	<b>Major Conclusion:</b> "KIPP schools post substantially greater gains than what is considered normal"					
SRI Int'l	2006 & 2008	5	San Francisco Bay Area	student level for 3 of 5 schools	3 years	propensity score matched comparison group for 3 of 5 schools
	<b>Major Conclusions:</b> "in most grades and cohorts, KIPP students make above-average progress compared with national norms, but with considerable variation across schools, grade levels, cohorts, and subjects." Matched student level analysis finds positive effects for first year, but attrition and in-grade retention make it impossible to tell whether this advantages carries to later grades.					
Augenbick, Palaich & Associates	2006	1	Denver	school level	First year	national norms; same school prior year; other similar schools in district
	<b>Major Conclusions:</b> "Despite significant staff turnover and extreme public scrutiny and pressure, [the school] ended its first year in a better place than when it started out." "Despite the jumps from 2005-2006, Cole College Prep continues to score below the district average, especially in Reading for both 7th and 8th grades."					
Musher, Musher et al.	2005	1	Houston KIPP Academy	school level	2 cohorts (school's first), 5th through 7th grade	national norms
	<b>Major Conclusion:</b> Scores on Woodcock-Johnson tests in Reading, mathematics and writing improved dramatically (about 1.8 years for each academic year) for both the first and second cohorts.					
Mac Iver & Farley-Ripple	2007	1	Baltimore	student	Four years four cohorts	own prior achievement; students at feeder schools
	<b>Major Conclusions:</b> Attrition "was not trivial," and students who left were more likely to be male and had lower test scores. In most cases the KIPP advantage was statistically significant even when students who subsequently left the program were retained as part of the experimental group.					

## *Review of the Studies*

In this section I review the basic design features and key findings of the seven external studies of KIPP. Summary descriptions of the studies' key design features and major conclusions are available in Table 1 (preceding).

### **New American Schools: A Study of Three KIPP Schools**

In 2002 Doran and Drury evaluated achievement gains in three schools (KIPP DC/KEY Academy in Washington, DC; Gaston College Preparatory in Gaston, North Carolina; and 3D Academy in Houston, Texas).<sup>15</sup> The study examined fifth-grade performance only and focused on the start-up year (2001-2)—a time when things might be rocky but also when things might go better because the school is smaller and enthusiasm presumably high. Test score data available to the researchers varied from site to site. For the Houston analysis, researchers had both nationally normed Stanford 9 and state criterion-referenced (TAAS) scores, not only for the fifth grade but for the two prior years at the students' home schools. In Gaston, they had state test scores only, but also for grades 3 & 4 (pre-KIPP) as well as end-of-grade 5 (after one year at KIPP). For DC, researchers had only Stanford 9 scores, and only a fall and spring sequence.

The study concluded that KIPP students' test scores "improved at impressive rates," greater than those same students achieved in the year before entering KIPP, and greater than the performance of their respective districts. The gains applied overall and to various subgroups. The largest achievement gains were in the DC school, where the average normal-curve equivalent (NCE) gain was 12.13 in reading and 23.54 in math. These scores, however, were based on fall-to-spring comparisons rather than the spring-to-spring pattern upon which the national Stanford 9 norms are calculated, a factor that might inflate results because it does not reflect the typical (and often large) summer fall-off. Because grade 3 and 4 scores were not available in DC, we do not know whether these students might have been exhibiting such gains even before they enrolled.

In Gaston, where there were both pre- and post-data,<sup>16</sup> KIPP students made gains in proficiency, but the patterns in reading and math were different. Pre-KIPP reading proficiency was declining prior to fifth grade, and then rose after the same students experienced the KIPP program in fifth grade. In math, KIPP students made progress toward proficiency but at a slower rate than they had been making previously. The authors suggest that the slower rate of improvement in math may be due a ceiling effect: over 80% of the KIPP fifth graders were proficient—by North Carolina standards—even before they enrolled in the school, leaving little room for improvement (and suggesting a relatively low state proficiency threshold).

In Houston, KIPP students did better than other district students and improved over time on the state TAAS reading and math tests, but they also had been scoring higher even before they came to KIPP. For example, 88% of KIPP fifth graders were proficient compared to 84% of other district fifth graders, but 87% of those same KIPP student already had been deemed proficient in grade 3

## What Do We Know About the Outcomes of KIPP Schools?

and 85% in grade 4. As seems likely in North Carolina, the low proficiency thresholds on the state tests create potential ceiling effects, and mask what might be important information regarding different performance levels among students considered proficient. Clear signs of the limitations of the TAAS include the fact that 100% of white KIPP students had been deemed proficient in both reading and math even before they started fifth grade.<sup>17</sup>

In fact, the scaled SAT 9 exams show clearly that KIPP students had been making above-average gains in both reading and math before they started fifth grade. For example, KIPP students gained 25.33 points in math (vs. 22.6 for the nationally normed group) and 26.56 (vs. 14.9) in reading, but the year before enrolling these students had gained 33.98 (vs. 24.7) in math and 25.44 (vs. 23.3) in reading). This indicates that, at least in the case of this one Houston school, some form of entry-level selection bias may have been taking place.

### **KIPP: Diamond Academy, Memphis**

The second independent study listed on the KIPP site is a case report on KIPP Diamond (Daring Individual Achievers Making Outstanding New Dreams) in Memphis. The report listed on the web site is an unpublished Year 2 assessment,<sup>18</sup> but the researchers, from the University of Memphis' Center for Research in Education Policy (CREP), have also produced other reports on the school.<sup>19</sup> To take advantage of the longer time series, I focus on their Year 4 report, although I draw on the others for some details and to highlight how the passage of time may matter.

Like the three KIPP schools in the previous study, KIPP Diamond Academy (KIPP: DA) was a relatively early KIPP school, and in Year 1 operated only at a single grade. The school opened in 2002 with a fifth-grade enrollment of 49. The Year 4 study includes the original cohort (notably down to just 16 by eighth grade), as well as three other cohorts at various grade levels who had been at the school for one, two, or three years.

This study stands out not only for its longitudinal range but for its detailed cataloguing of implementation specifics. Rather than simply assessing outcome changes associated with attending a KIPP school (an ill-defined intervention because it's unclear what specific characteristics of the school might account for the change), CREP employed multiple instruments to look inside the black box of what, in this instance at least, being a KIPP school means. In addition to test scores, yearly reports include: results from a School Observation Measure (SOM) based on observers' records of activities during 15-minute observations in 10 to 12 randomly selected classrooms; a 49-item school climate inventory; teacher and parent surveys; focus groups; and a set of program implementation benchmarks.

After one year of the school's operation, the CREP analysis found strong signs of favorable "school climate" and of teacher and parent satisfaction, along with generally positive indications of student achievement. "School climate" was measured along seven dimensions (Collaboration, Environment, Expectations, Instruction, Involvement, Leadership, and Order), and in Year 1 KIPP: DA scored more than one full point better (on a scale of 1 to 5) than national norms on every

## What Do We Know About the Outcomes of KIPP Schools?

point of comparison.<sup>20</sup> Teachers were almost uniformly positive in their responses. To assess achievement gains, students were individually matched (on demographics and test scores) with fifth graders in nearby schools. There were no differences in achievement between KIPP and the control group on pre-test assessments (both norm- and criterion-based), but KIPP students outperformed the comparison group in language arts, reading, and math post-tests (there was no difference on a writing examination).

Based on the Year 1 study, the researchers deemed the achievement outcomes “noteworthy,” especially in light of the fact that the school “had to overcome multiple challenges in its first year” and that “the consensus by scholars and researchers of school reform [is] that school change takes multiple years to produce implementation success.”<sup>21</sup> The generally upbeat assessment was tempered slightly by the observations that “Given the mobility and burn-out of many urban teachers, an important question for future research concerns the degree to which the increased work and time demands of KIPP: DA impact the sustainability and scale-up of the program.”<sup>22</sup>

By Year 4, “school climate” had deteriorated on all seven dimensions, although overall it remained positive compared to national norms on the same climate scale. Teacher surveys showed a similar pattern of declining satisfaction with the school extending across all five dimensions tested with that instrument (Capacity/resources; Focus; Outcomes; Pedagogy; Support). Year 4 parents, nonetheless, were almost universally positive in their assessment.

The researchers attribute the general ebb in enthusiasm to leadership changes: the popular first principal had to step down for medical reasons, and the school then experienced a period of interim leadership until a new principal was hired in spring 2005. Despite this instability, student achievement was generally positive. The researchers assessed achievement using multivariate analysis of covariance of the number of correct answers, controlling for the students’ own pre-charter school scores, and chi-square analyses of proficiency levels for KIPP versus control students. The analysis revealed “fairly positive outcomes.”<sup>23</sup> Seven of eight overall grade-level comparisons favored KIPP although, with low sample sizes, most were not statistically significant. The 16 eighth graders who remained from the initial cohort did not significantly outperform their non-KIPP comparisons in the number of correct answers but were more likely to reach proficiency in reading. As after Year 1, the CREP researchers saw encouraging signs. Although school climate and teacher attitudes had deteriorated, they expressed an expectation that the newly appointed principal would turn things around.

### **EPI Study of 24 schools**

In 2005, Educational Policy Institute, Inc. (EPI) released *Focus on Results*, a report commissioned by the KIPP Foundation.<sup>24</sup> This study may have done more than any other to create the impression that there is strong empirical evidence showing that KIPP works. Jay Mathews, the *Washington Post*’s

## What Do We Know About the Outcomes of KIPP Schools?

respected and widely read education columnist, summarized the study in an article headlined “Study Finds Big Gains For KIPP; Charter Schools Exceed Average.”<sup>25</sup>

The EPI study analyzed the performance of students, all of whom started KIPP in fifth grade in fall 2003. The cohorts came from 24 schools, encompassing all KIPP schools that administered the SAT 9/10 to their fall 2003 fifth grade students and a follow-up exam to the same students the next spring or fall. Seven KIPP schools were not included based on this testing criterion; EPI provided no data and no indication of whether the excluded schools may have differed in other important ways. The data were delivered to EPI by KIPP in cohort form. There were no student-level data. The exams of all students in the fall 2003 cohort were compared to that same cohort the following spring or fall, but there is no certainty that the same students took the exams at both points; some students probably left the schools, and others may have entered. (Issues and data concerning attrition bias are discussed later in this brief.) Nor was there a comparison group other than the national sample to which the exam was normed. “Although this study design does not allow for claims of causality,” the authors acknowledged, “we feel confident that the following findings are reflective of what is happening in KIPP schools.”<sup>26</sup>

The KIPP cohorts started out below NCE means on the initial test but, at most schools, improved substantially. Reading scores improved relative to the national means in all but one of the 18 schools with fall-to-spring cohorts; the average increase was 10.1 points. The scores also increased in the schools with fall-to-fall testing, although these increases were substantially smaller, which could reflect a summer fall-off or could be associated with changes in the specific students enrolled in a cohort from the end of one academic year to the beginning of the next.

### **San Francisco Bay Area KIPP Schools**

The fourth independent report included on the KIPP website is a first-year study by SRI International of five KIPP schools in the San Francisco Bay area.<sup>27</sup> In September 2008, SRI released a second report that extended the analysis to include three years of data.<sup>28</sup> The study focuses on five KIPP middle schools, all serving grades 5-8. The oldest opened in 2002, the newest in 2004. Each school began with a single grade serving about 80 students; by 2006-2007 each served a full complement of grades 5-8, with enrollment ranging from 239 to 328. Achievement data included results on both the SAT 10 and the California Standards Test (CST), and for three of the five schools the researchers were able to analyze individual student-level CST scores compared to a matched group of students from the surrounding district. In addition to compiling test scores, researchers conducted interviews (of teachers and school leaders), administered surveys (of teachers and students), and gathered structured observations.

There was no evidence of race or class selection bias in enrollment; the five schools all had higher proportions of minority students than their surrounding districts, and four also had more low-income students. The KIPP schools, though, did tend (four of the five) to enroll smaller proportions of English learners and

## What Do We Know About the Outcomes of KIPP Schools?

special education students. Four of the five schools also had notably higher proportions of female students than their districts.

The earlier (2006) SRI study offered a strongly positive assessment of achievement, concluding that “the overall performance of students performing at a proficient level or above is consistently higher for KIPP than for comparable schools in the districts—in some cases dramatically so.”<sup>29</sup> The tenor of the findings from the more recent multi-year SRI report is more modulated: “in most grades and cohorts, KIPP students make above-average progress compared with national norms, but with considerable variation across schools, grade levels, cohorts, and subjects.”<sup>30</sup> The student-level analysis with a matched comparison group found statistically significant effects on achievement for fifth graders and for the subset of sixth graders who entered at that grade level rather than in fifth, as is the norm.

However, researchers were unable to determine whether achievement was positively affected in grades 7 and 8 because of three complicating factors. First, attrition rates were very high: 60% of students who started KIPP in fifth grade had left before the end of eighth grade. Second, attrition was selective: students who left tended to be the ones who entered with lower scores and who made the least progress while in KIPP. Third, researchers found it difficult to control for the fact that compared with the surrounding districts, the KIPP schools had a relatively high rate of students forced to repeat a grade due to poor performance.

Like the Memphis study, this one looked inside the black box to see what the KIPP experience actually entails. Interviews and surveys with school leaders and teachers showed high levels of buy-in to the KIPP approach, but also indicated the stress of the long school day and school year. Teacher turnover was high. Of 121 teachers hired since 2003-4, only 43 remained at the beginning of the 2007-08 school year; median tenure was only two years. The combination of turnover and expansion means that school leaders never escape the pressure to aggressively recruit and train new teachers, leading at least some to wonder about the sustainability of their enterprise.

### **Denver: A KIPP “Transition” School**

Cole Middle School in Denver was closed by the state in August 2004 because of poor student performance. KIPP was selected that November to take over the operation of the school, renamed Cole College Prep. A report by Augenblick, Palaich & Associates (APA) reviews the performance of the school in the 2005-06 academic year,<sup>31</sup> during which Cole operated as a “KIPP transition school.” This designation meant Cole was not considered an official KIPP school until it had the chance to recruit its own fifth grade class (which it did in 2007). During the 2005-6 year the school served seventh and eighth graders who had enrolled prior to the KIPP takeover. This unusual status limits generalization from the Cole case. Moreover, the study’s methodology is less sophisticated than some of the others discussed here. However, as the report notes, KIPP’s willingness to take on the Cole challenge rested in part on its national leadership’s decision that, in the context of NCLB, there was a growing need for this kind of takeover work.

## What Do We Know About the Outcomes of KIPP Schools?

The organization was interested in helping to fill this gap. Thus the Cole experience might provide useful lessons for what to expect in cases like this, which may become more frequent in the future.

The APA report included surveys of teachers and parents, less formal discussions with students, and an analysis of enrollment, but the empirical centerpiece was an assessment of changes in school-level performance between 2005 and 2006 on the SAT 10 and the Colorado Student Assessment Program (CSAP) exams in multiple subject areas.

The percentage of students deemed proficient or advanced on the state exam increased substantially during the first year under KIPP's oversight. Between 2005 and 2006, the percent of seventh- and eighth graders deemed proficient doubled or tripled in almost all tested subjects (reading, math, science, and writing). The one exception was for eighth grade reading, which did not change. Even those increases, however, left the school with only mediocre results compared with other Denver schools serving similar populations. For example, Cole College Prep seventh graders improved from 10% to 24% proficient in reading compared to the seventh grade cohort from the prior year—but this improvement pales when compared to a 40% improvement rate for the district overall, and it placed Cole only at the middle of the pack among six comparison schools with similar demographics. APA's analysis of the SAT 10 scores focused only on students who took the exam in both the fall and spring; for this subset of 90 students (compared to a total of 115 who had taken the exam in the fall), scaled scores increased in all subjects (language, math, reading, science, and social studies). For seventh graders the gains were greatest in math and reading; for eighth graders in science and math. Again, despite this progress, performance levels remained poor when compared to national norms.<sup>32</sup>

### **Houston KIPP Academy**

Musher et al., in a 2005 study not included on KIPP's list of independent reports, provide a longitudinal analysis of achievement gains for the first two cohorts in the Houston KIPP Academy. Approved in 1995, it was the first charter school in Houston Independent School District and one of the very first KIPP schools in the nation.<sup>33</sup> The analysis uses subtests from the Woodcock-Johnson Tests of Achievement-Revised (WJ-R), a nationally normed exam. A second part of the study looks at the performance of a later cohort of students from the school on the Texas Assessment of Knowledge and Skills (TAKS), introduced in 2003 to replace the TAAS examination, which had been criticized for setting too low a standard. The study's TAKS analysis uses scores averaged at the school level and does not control for prior performance trends or any student characteristics.

Both the first and second entering classes were near grade level when they were first tested (in the second month of their initial year), but they made sharp progress thereafter. Both cohorts gained an average of about 1.8 years each during their fifth- through seventh-grade years. By the end of seventh grade, those students still in the program were performing at about the 10th-grade level, with the sharpest gains in mathematics. The analysts note, but somewhat underplay, the

## What Do We Know About the Outcomes of KIPP Schools?

substantial attrition. The first cohort shrank from 59 to 37 students between fall of fifth grade and the end of seventh; the second cohort shrank from 55 to 39. The authors acknowledge that “some attrition occurred,” but they dismiss it as a major factor because there was not much attrition during the first year or between sixth and seventh grade for the second cohort and there were strong academic gains in those years as well. Nonetheless, it is unclear from this study whether selective attrition might be a rival explanation for the sustained and large gains that it reports.

The TAKS analysis, performed for a later cohort of students at the same school, provides a comparison with other schools in the district and state. KIPP students at all grade levels (5-8) were more likely to receive a passing grade than the average student in Houston and Texas, and performed decisively better than students at five Houston schools serving similar low-income populations. The strongly phrased conclusion was that “KIPP was the only low-income neighborhood school in Texas at which 100 percent of eighth-grade students passed all components of TAKS.”<sup>34</sup> But this perhaps should be tempered a bit by the facts that selective attrition may have played a role and that the TAKS cut-off, for passing is generally low (70% of the state’s eighth graders passed all components of the exam that year).

### **Baltimore KIPP Ujima Village**

The KIPP website also does not include a link for a June 2007 report on the Baltimore KIPP Ujima Village Academy, authored by Martha Mac Iver and Elizabeth Farley-Ripple.<sup>35</sup> This study resembles the Memphis and Bay Area studies in its longitudinal design, incorporating four cohorts entering fifth grade in 2002, 2003, 2004, and 2005. The study analyzes the trajectory of test performance (Maryland School Assessment) of the KIPP students over time and compares it to that of students at feeder schools. Researchers paid particular attention to the role of transfers and attrition.

Across the four cohorts, KIPP students were similar to those in the comparison group in gender, race, free-lunch status, and attendance. They were less likely to have special education status; this was true in all cohorts, with the difference being statistically significant for the first two cohorts. Entering KIPP fifth graders had higher fourth-grade test scores, with the differences reaching statistical significance for cohort 2 in both math and reading, and for cohort 3 in math.

KIPP students generally outperformed the comparison group, even when controlling for gender, special education status, and fourth-grade test scores. Substantial attrition again emerged as an important factor to consider, and this study goes further than the others in assessing whether attrition might in itself account for the apparent KIPP advantage. Of the 79 fifth graders in the initial cohort, 11 left during the first year and only 49 were still at the school at the end of eighth grade. Students who left did not have significantly different fourth-grade achievement, but they did have lower scores as fifth graders, indicating that at least some of the attrition may have been selective in a way that would raise test

## What Do We Know About the Outcomes of KIPP Schools?

scores over time. Significantly, the researchers found that the KIPP advantage still held even when they included students who transferred out as part of the experimental group (referred to as “KIPP ever”). For example, the first cohort (entering 2002) had shown no significant difference from the comparison group in either reading or math entering scores. As fifth graders, this cohort outperformed the comparison group in math but not reading. As sixth graders, those who stayed in KIPP for the whole year outscored the comparison group in reading, and those who attended even part of the year outscored the comparison group in math. In grades 7 and 8, KIPP ever students did better in both reading and math.

### **The Weight of the Evidence**

Elsewhere I have argued against the temptation to look for “killer studies,” stand-alone studies that are so strong in design and so clear and decisive in their findings that they resolve policy questions once and for all.<sup>36</sup> In most instances, citizens and policy-makers have no choice but to make inferences based on a body of empirical research that includes studies of various imperfect designs, context-specific findings, and problematic measures. Rather than focusing on the one or two studies with the best research designs and discarding information available from the others, it makes more sense to take full advantage of the collective nature of the research enterprise, looking for broad patterns and central tendencies. Findings from studies with stronger designs warrant greater attention, and when strong studies coalesce around findings that differ from those of weaker ones, they carry more authority even if they are fewer in number.

There are some differences in findings among the studies reviewed above, and each has methodological soft points, some acknowledged and some not. At this point we can declare no definitive answer about whether and to what degree KIPP is ratcheting up performance or—a more modest contribution—providing a new alternative to a subset of families and children who are attracted to and willing to stick to its demanding regime. There is little indication from these studies that KIPP is systematically biasing student selection by race, class, or past performance as it recruits and enrolls students, although KIPP tends somewhat to enroll more females and it remains possible that there are additional unmeasured differences distinguishing KIPP entrants from others. What is emerging as the bigger threat to internal validity is the issue of substantial attrition, especially among students who are not performing as well as others in the schools. It is apparent that such attrition is occurring and likely that it is exaggerating the findings of relative gains made by KIPP cohorts, although at this point it does not appear that accounting for attrition would eliminate the KIPP edge.

Overall, the balance of the evidence suggests something positive happening at these sites, although not of the magnitude and cumulative nature that some popular accounts suggest. Policy makers at all levels of government are right to be attentive to KIPP and to consider it a possible source of information and guidance for the decisions they face. But once they are paying attention, what does the research tell them they should do?

***External Validity: Drawing Lessons  
Across Schools, Place, and Time***

Moving from findings to policy requires making inferences about whether the gains demonstrated in most of these studies are reliable, sustainable, and likely to apply in other contexts. Such questions shift us from issues of internal validity—an arena in which social science has clearly defined and reasonably consensual criteria—to external validity, where the ground is less certain. In this final section I discuss external validity as it relates to the selection of schools, maturation effects, and implications for other valued outcomes from schools. Taken as a whole, the studies reviewed here are less well developed in addressing these issues than they are in considering internal validity. They offer some hints and suggestions but leave many questions unanswered.

**Selection of Schools: Quality & Context**

From a narrow program-evaluation standpoint, it may be enough to know whether a particular set of KIPP schools is doing well. Charter authorizers and parents in a given community, for instance, may be satisfied to know that their local KIPP school is doing well. But officials considering KIPP involvement in their community, wondering whether a charter applicant is a good bet if it proposes to use KIPP-like methods, or deciding whether to increase charter levels in their jurisdiction, need to judge whether the schools studied so far represent the likely results if transferred to a new context or expanded more generally.

These seven studies raise several issues relating to possible school selection bias. The first relates to potential quality bias in schools selected for study. Schools confident they are doing well are more likely than struggling schools to expose themselves to scrutiny. Selection bias related to quality also could occur in several other situations. Some schools might not be studied because they have failed to collect or properly maintain requisite test score data. Or, turnover in leadership might interfere with getting approval to include a school in a study. Or, a school might be closed or dropped from the KIPP alliance because of operating or staffing or performance problems in their early years.<sup>37</sup> If the KIPP schools not studied are systematically different from those that have been studied, drawing lessons for policy is a riskier enterprise.

A second kind of selection bias has to do with context. Advocacy groups often square off about how much schools can accomplish without supportive communities, but few deny that the relative presence or absence of community buy-in, fiscal support, social capital, and civic capacity can make a difference in organizational performance. If early KIPP schools tended to locate in communities with favorable climates, their outcomes may not be a good indication of what will come as the operation expands. This is not just a question of whether KIPP is serving relatively high-need populations, because the enrollment data in these reports clearly indicate that KIPP is serious about its stated goal of working with minority and high-need populations. But interactions between school and community can be complex. A school that is invited in by

## What Do We Know About the Outcomes of KIPP Schools?

indigenous organizations—a community organization, a leading politician, a respected social service organization, a church—can have very different outcomes than one that enters a community with an indifferent or hostile reception. Racial and ethnic differences between school staff and students might also be a factor. So might local levels of per-pupil funding, patterns of redevelopment or urban disinvestment, and population growth and decline.

Making inferences about whether the successes of studied schools will translate into other contexts requires good information about the contexts, which the existing studies discuss only in broad stroke—if at all. They provide much better information on the characteristics of their enrollees than they do on the surrounding areas. It also is possible that early KIPP schools have been operating in especially hostile settings; after all, the early political responses to charter schools were often wary in a way that is moderating over time. If that's the case, the results achieved by this set of schools might understate the general potential of the KIPP model. Without more detailed information however, we have little leverage on these types of questions.

Relying on randomization to select research sites would have addressed the selection bias issue, but none of the studies has done that. Even better would be to include the universe of existing schools, making sampling unnecessary. The EPI study comes close to this, including 24 of 31 schools in operation during 2003-4. It excluded schools that failed to provide data to the KIPP Foundation or that used an exam other than the Stanford Achievement Test, however, and the researchers made no effort to track down missing data from individual schools.<sup>38</sup> Schools different enough to opt for a different testing instrument than the one favored by the central administration may differ in other ways also, including performance. Moreover, while the EPI study is relatively strong on cross-sectional coverage, compared with several other studies it is limited in its lack of student-level data and in its reliance on a single year of test-score gains. The rest of the studies look at a single school or a very small number of schools and provide little or no information about why those schools were selected or how they compare with other KIPP schools.

### **Maturation of Schools and Systems**

As is often the case with new policy initiatives, the earliest studies of KIPP focus on first-generation reforms in their early years. Two types of selection bias issues relating to changes over time (maturation) potentially come into play. The first relates to *organizational maturation* at the school level. Schools in the early years of operation may do worse (for example, because of a range of challenges associated with starting up and the students' own transition) or better (for example, because of early enthusiasm, exceptional pioneer leaders, philanthropic and public start-up grants).<sup>39</sup> The second relates to *systemic maturation*: as the overarching KIPP management system (and charter school phenomenon itself) matures, the performance of later cohorts of schools might differ systematically from that of early cohorts. Here, too, the selection bias in principle could work in either direction. Later cohorts might perform

## What Do We Know About the Outcomes of KIPP Schools?

systematically better than the pioneers. Such improvement might occur, for example, under any of the following circumstances: if KIPP learns more about what works, if it institutionalizes its recruitment and training systems, if increased experience allows charter authorizers to improve application reviews and to better orient providers to community needs, or if parents learn more about distinctions among charters and do a better job of selecting the charters that will best meet their children's needs. Later cohorts might do worse, however, under a different set of possible circumstances: if scaling up creates problems in finding qualified employees or leads to bureaucratization and calcification, if geographic expansion entails working with less accommodating communities, if competition increases from other charter management organizations, if the philanthropic support per school declines, if public funding fails to keep pace, or if governmental oversight and regulation imposes new costs.

The studies of Ujima Village in Baltimore, KIPP: DA in Memphis, and the five Bay Area schools provide the best longitudinal analyses. The results in each case were generally positive, in the sense that various comparisons with control populations often suggested relative gains and only rarely negative ones. But small sample sizes, complicated by substantial attrition, left the possibility that some of the positive results were attributable to chance or to differential patterns of transferring out. In Memphis, researchers attributed the somewhat disappointing results to leadership discontinuity in 2005-6. In Baltimore, researchers found more consistently positive results, especially in mathematics; yet, because some entering cohorts had higher math scores than comparison students in feeder schools, selection bias at entry may account for some, if not all, of the gains. Moreover, attrition was substantial and students who left had lower scores than those who stayed; therefore, bias in attrition may account for some portion of the gains, although again not all. The Baltimore study also excluded students who were retained in grade, a practice more common in KIPP than among the feeder schools. The Bay Area study also verified that attrition and in-grade retention are substantial factors complicating longitudinal analysis.

While longitudinal patterns remain an area needing much more research, the existing literature raises doubts about the most optimistic scenario (promoted by some charter advocates), that small gains in early years will be followed by larger gains as incoming students shed their bad habits and poor skills in the new environment and as the schools and systems hurdle difficult start-up challenges and really hit their stride.<sup>40</sup> To the contrary, in contrast to the predicted slingshot effect in which early momentum generates exponential gains at later grades, there are some indications that early gains are at least in part the result of a catch-up phenomenon, after which relative growth rates attributable to KIPP may level off to some extent.

### **Other Outcomes**

Reflecting contemporary discourse, the studies reviewed here place primary emphasis on academic achievement, defined in terms of standardized test scores and almost exclusively in reading and mathematics. Historically, though,

## What Do We Know About the Outcomes of KIPP Schools?

American schools have been expected to serve a wide range of functions. In the academic arena, schools have been expected not only to teach basic skills in reading and mathematics but also to teach other academic subjects, such as science, history, civics, and foreign languages. At various times, some of these have even been perceived as urgent needs in the face of national and international challenges. Moreover, academic achievement at times has been conceptualized much more broadly than raising test scores, including things like reducing dropout rates, preparing more students for higher education, and creating lifelong learners. The purposes of schooling also have been seen as extending beyond academics, often including such tasks as promoting values relating to good citizenship, teaching and providing good personal and public health habits (for example, teaching hygiene and physical education and providing eye examinations and inoculations), cultivating business skills (typing and accounting classes) as well as family and parenting skills, and providing cultural enrichment. Many might agree that academic learning in math and reading are highest priority, because they are foundational. But public concern in many communities about the perceived narrowing of the curriculum attributable to NCLB and state accountability initiatives suggests that there continues to be a constituency who conceptualizes education far more broadly.

To the extent that such other outcomes are also important, assessment of initiatives like KIPP should include attention to what the gains in math and reading test scores might cost in other important areas. Of the studies to date, only the APA Denver Cole School study and the Musher et al. Houston study had enough information on other content areas to assess whether gains in reading and mathematics came at the expense of subjects like science and social studies.<sup>41</sup> Neither found clear evidence that this was the case. At this point, none of the existing studies can offer insight into longer-term academic growth or broader dimensions of life success—something that may be possible if the longitudinal studies continue to follow the early cohorts.

The SRI study of Bay Area schools, however, includes systematic information on how instructional time was used. While there is indeed a substantial amount of time devoted to reading and math skills, the much longer school days (9.5 hours; over 50% more time than most public schools) and summer and Saturday classes make it possible to increase attention to reading and math without skimping on other subjects. Indeed, while many traditional public schools are cutting back on so-called enrichment courses, such as in music, dance, theater, sports, and the like, it appears that the Bay Area KIPP schools use their extra time to make room for a wide variety of activities. Given this information, if a study were to include outcome measures tapped to these types of nonacademic activities, KIPP might fare well. But no such studies have yet been done.

### **Inside the Box: What Is KIPP?**

The potential problem of selection based on quality is exacerbated by the fact that KIPP deliberately empowers school-level leaders to make a wide range of decisions according to local context and need. Because KIPP schools do not

## What Do We Know About the Outcomes of KIPP Schools?

offer a uniform curriculum or pedagogical style, when they do have good results it is hard to know what elements may be responsible. This may very well be a sensible pedagogical policy, but it makes it hard for policymakers to apply lessons from KIPP to other charters or traditional public schools.

The Memphis and Bay Area studies provide the most detailed information on what actually occurs inside the black box of the KIPP experience. These suggest there is substance behind KIPP's public emphasis on high quality and motivated teachers, more time and focus on learning, and high expectations for students, parents and staff. The Memphis study, though, shows that maintaining focus and enthusiasm can be difficult if there is turnover, even at a school that starts off well, and the Bay Area study shows that there can be substantial variation in the inculcation of the KIPP culture even across as few as five schools in a single metropolitan area.

The fact that there is variation in what happens at the school level is both an issue of concern and an opportunity for better knowledge. The concern is that the KIPP effect demonstrated so far might depend on a supply of personnel that is finite. If KIPP, like McDonalds or Wal-Mart, comprised a set of precisely defined activities, it would be safer to assume that the model could be replicated from place to place, but the KIPP model depends on leaders astute enough to make good on-the-spot decisions and not simply follow a script. The opportunity lies in the fact that variation in the elements makes it more feasible that future research might link such variation to performance, gaining more insight into the question of what, specifically, about KIPP's approach seems to matter. This, in turn, could help answer the question of whether the KIPP model could be successfully "injected" into traditional public schools, constrained as they are by rules, regulations, contracts, politics, and fiscal conditions. None of the research to date has traveled down this path.

### *Final Thoughts and Recommendations*

#### **Final Thoughts**

Despite being around for almost a decade and a half, KIPP is still relatively under-studied. We need more and richer studies covering a broader slice of the KIPP sites, expanding the range of outcomes considered, and most importantly continuing the process of assessing how the model bears up under the strains of expansion and sustainability over time. Even the strongest studies to date lack the internal validity power of an experimental model. Nonetheless, by using student-level data, matching comparison groups based on demographics and prior test scores, and beginning to wrestle with the complications imposed by high attrition and in-grade retention, they have provided some reasonably sound evidence relevant to the question of whether KIPP independently affects academic achievement. They remain vulnerable to methodological challenges, and critics will argue that despite their efforts they have not yet ruled out the possibility that higher scores are due to preexisting differences in the students and families that seek out KIPP and stick to its demanding regimen. That all of these studies show

## What Do We Know About the Outcomes of KIPP Schools?

positive effects is meaningful, though, and this increases the probability that as more and better studies accumulate they will not overturn this verdict entirely.

The intense nature of debates about charter schooling, however, does not hinge on whether some charter schools do a good job. Increasingly, research is converging on the finding that some do well and others not.<sup>42</sup> Ideological and partisan stakes run high because charter schooling has been framed as a test case of rival systems, one anchored in government and democratic oversight and the other in markets and family choice. Behind the heat and steam it is easy to lose sight of the more mundane, but potentially more valuable, lessons towards which the emerging evidence may point. Recruiting better teachers and school leaders, spending more time in school, motivating families, unhinging school assignment from strict attendance zones: are these what matters? If so, perhaps we can turn down the flame of clashing systems and confront more directly the question of how these can be best institutionalized.

### **Recommendations**

In light of the above, following are recommendations based on this review of the best available evidence to date.

- Policy makers at all levels of government should pay attention to KIPP and consider it a possible source of information and guidance for their decisions.
- Although KIPP may yield useful information, policymakers and others should temper their interest in the operation with wariness and realistic expectations. There are significant unanswered questions about how expansion might affect outcomes, especially in relation to the difficulty of sustaining gains dependent upon KIPP's heavy demands on teachers and school leaders. Moreover, it is not realistic to think that the KIPP model is a panacea for distressed systems. It is possible that only a small proportion of students and families will be able to meet the demands KIPP imposes on them; even those enthused when they begin the KIPP regimen tend to leave in high numbers.
- Policymakers, accordingly, should treat KIPP schools as potential tools that may contribute to—but not substitute for—systemic improvement.
- Policymakers should be aware that KIPP has prompted some district interest in longer school days, weeks, and years. However, an extended schedule sometimes brings parental objections as well as potential taxpayer objections to the additional expense. With no strong evidence yet linking extended scheduling to KIPP success, policymakers might best encourage it as a school-level (rather than district-wide) option while concurrently promoting a combination of experimentation and careful analysis of consequences.
- Researchers should help provide better data on patterns of movement in and between charter schools and traditional public schools, including information on why students leave and how their mobility affects student and school-level performance.

## What Do We Know About the Outcomes of KIPP Schools?

### *Notes and References*

- <sup>1</sup> Jay Mathews (2006, April 2). "A Miracle in the Making/KIPP turns its efforts toward elementary schools," *The Washington Post*, W20.
- <sup>2</sup> KIPP Foundation (2007). *KIPP: Report Card 2007*. Retrieved Sept. 15, 2008, from <http://reportcard.kipp.org> (Registration req'd.)
- <sup>3</sup> KIPP Foundation (2007). *KIPP: Report Card 2007*, 6. Retrieved Sept. 15, 2008, from <http://reportcard.kipp.org>. (Registration req'd)
- <sup>4</sup> Retrieved August 31, 2008, from <http://www.kipp.org/01/resultsofkippsch.cfm>.
- <sup>5</sup> The omission of these two studies was probably due merely to oversight. No implication is intended here of selective omission, and the omitted studies are not notably more critical of KIPP than those that were listed.
- <sup>6</sup> For a wide-ranging but generally enthusiastic account of the advantages of RFTs, see:  
Mosteller, F. & Boruch, R. eds. (2002). *Evidence Matters: Randomized Trials in Education Research*. Washington DC: Brookings Institution.  
For a discussion of some of the trade-offs between RFTs and other strong designs, particularly in the context of charter school research, see  
Betts, J. & Hill, P.T. (2006). *Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines*. National Charter Schools Research Project, Retrieved Sept. 15, 2008, from [http://www.ncsrp.org/cs/csr/view/csr\\_pubs/5](http://www.ncsrp.org/cs/csr/view/csr_pubs/5)
- <sup>7</sup> Weiss, C. (2002) "What to Do Until the Random Assigner Comes," in Mosteller, F. & Boruch, R. eds. (2002). *Evidence Matters: Randomized Trials in Education Research*. Washington DC: Brookings Institution.
- <sup>8</sup> Early studies finding a private school advantage include:  
Coleman, J. S. & Hoffer, T. (1987). *Public, Catholic, and Private Schools: The Importance of Community*. New York: Basic Books  
Chubb, J. E., & Moe, T.M. (1990). *Politics, Markets, and America's Schools*. Washington, DC: Brookings Institution Press.  
More recently at least two major studies have raised questions about the private school advantage, finding it depends on type of private schools and grade level and that under some circumstances it disappears or reverses.  
Braun, H., Jenkins, F. & W. Grigg, W. . (2006). Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling (NCES 2006-461). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office  
Lubienski, C. & Lubienski, S.T. (2006). School Sector and Academic Achievement: A Multi-Level Analysis of NAEP Mathematics Data. *American Educational Research Journal* 43 (4): 651-698.
- <sup>9</sup> Carnoy, M., Jacobsen, R., Mishel, L. & Rothstein, R. (2005). *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement*. New York: Teachers College Press.
- <sup>10</sup> Carnoy, M., Jacobsen, R., Mishel, L. & Rothstein, R. (2005). *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement*. New York: Teachers College Press. The authors take care not to conclude that such selection biases, if real, fully account for any edge in test scores that may be observed. What they do challenge is some proponents' claims that KIPP achieves powerful test score gains despite systematically recruiting students with who are academically more disadvantaged than those in traditional public schools serving the same neighborhoods.
- <sup>11</sup> Mathews, J. (2007, April 24). "Looking at KIPP, Coolly and Carefully." *Washingtonpost.com*. Retrieved Oct. 29, 2008, from [http://www.washingtonpost.com/wp-dyn/content/article/2007/04/24/AR2007042400558\\_pf.html](http://www.washingtonpost.com/wp-dyn/content/article/2007/04/24/AR2007042400558_pf.html)

## What Do We Know About the Outcomes of KIPP Schools?

- <sup>12</sup> For a broad indictment, see Alfie Kohn, (2000). *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools*. Heinemann.
- <sup>13</sup> Under NCLB, schools that fail to make adequate progress at some point must be reconstituted, and increasingly districts are looking to test scores as a way to assess teachers and principals and in some cases to allocate financial rewards. On the likelihood that high stakes testing can lead to exaggerated test score gains, see: Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge MA: Harvard University Press, Chapter 10.
- Also Nichols, S.L. and Berliner, D.C. (2007). *Collateral Damage How High-Stakes Testing Corrupts America's Schools*. Cambridge, MA: Harvard Education Press.
- <sup>14</sup> Critics of mayoral control of New York City schools, for example, have presented evidence to show that focusing on levels instead of scores exaggerates the overall proficiency gains and narrowing of educational achievement gaps claimed by the Bloomberg/Klein administration.
- Green, E. (2008, August 5). "'Achievement Gap' in City Schools Is Scrutinized Slight Gains in English Are Reported." *New York Sun*.
- See also Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge MA: Harvard University Press, Chapter 8.
- <sup>15</sup> Doran, H.C., & Drury, D.W. (2002). *Evaluating Success: KIPP Educational Program Evaluation: New American Schools*.
- <sup>16</sup> Although restricted to the state test (no Stanford 9).
- <sup>17</sup> Worth noting because it seems incongruous, but is not specifically commented upon in the report, is that the report shows fully 100% of the KIPP special education students as having reached proficiency by the time they completed the KIPP fifth grade. Figure 6 of the report shows that 0% of these special education students had been proficient in third-grade reading (50% math), rising to 50% in fourth grade (falling to 33% math). The report does not provide the n for the various subgroups, but these changes are too large to be believable unless they are being generated by a very small number of students.
- <sup>18</sup> McSparrin-Gallagher, B. & Ross, S.M. (2005, January). *Analysis of Year 2 (2003-2004) Student Achievement Outcomes for the Memphis KIPP DIAMOND Academy*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.
- <sup>19</sup> Ross, S.M., McDonald, A.J., Alberg, M. & McSparrin-Gallagher, B. (2007). Achievement and Climate Outcomes for the Knowledge is Power Program in an Inner-City Middle School. *Journal of Education for Students Placed at Risk* 12 (2): 137-165.
- McDonald, A.J., Ross, S.M., Abney J., & Zoblotsky, T. (2008, March). *Urban School Reform: Year 4 Outcomes for the Knowledge is Power Program in an Urban Middle School*. Paper presented at the annual meetings of the American Educational Research Association, New York, NY.
- <sup>20</sup> The national norms were based on both middle schools and high schools.
- <sup>21</sup> Ross, S.M., McDonald, A.J., Alberg, M. & McSparrin-Gallagher, B. (2007). Achievement and Climate Outcomes for the Knowledge is Power Program in an Inner-City Middle School. *Journal of Education for Students Placed at Risk* 12 (2), 159.
- <sup>22</sup> Ross, S.M., McDonald, A.J., Alberg, M. & McSparrin-Gallagher, B. (2007). Achievement and Climate Outcomes for the Knowledge is Power Program in an Inner-City Middle School. *Journal of Education for Students Placed at Risk* 12 (2), 161.
- <sup>23</sup> McDonald, A.J., Ross, S.M., Abney J., & Zoblotsky, T. (2008). *Urban School Reform: Year 4 Outcomes for the Knowledge is Power Program in an Urban Middle School*. Paper presented at the annual meetings of the American Educational Research Association, New York, NY (March): 45

## What Do We Know About the Outcomes of KIPP Schools?

- <sup>24</sup> Educational Policy Institute. (2005). *Focus on Results: An Academic Impact Analysis of the Knowledge is Power Program (KIPP)*. Prepared for the KIPP Foundation and available at their website: Retrieved Sept. 15, 2008, from <http://www.kipp.org/01/independentreports.cfm>
- <sup>25</sup> Mathews, J. (2005, Aug. 11). "Study Finds Big Gains For KIPP; Charter Schools Exceed Average." *The Washington Post*, A14
- <sup>26</sup> Educational Policy Institute. (2005). *Focus on Results: An Academic Impact Analysis of the Knowledge is Power Program (KIPP)*. Prepared for the KIPP Foundation, 7. Retrieved Sept. 15, 2008, from <http://www.kipp.org/01/independentreports.cfm>
- <sup>27</sup> David, J.L, Woodworth, K., Grant, E., Guha, R., Lopez Torkos, A. & Vicki M. Young. (2006). *Bay Area KIPP Schools: A Study of Early Implementation First Year Report 2004-05*. Report by SRI International, Commissioned by The William and Flora Hewlett Foundation (March 2005).
- <sup>28</sup> Woodworth, K., David, J.L, Guha, R., Wang, H., & Lopez Torkos, A. (2008). *San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement, Final Report*. Menlo Park CA: SRI International.
- <sup>29</sup> David, J.L, Woodworth, K., Grant, E., Guha, R., Lopez Torkos, A. & Vicki M. Young. (2006). *Bay Area KIPP Schools: A Study of Early Implementation First Year Report 2004-05*. Report by SRI International, Commissioned by The William and Flora Hewlett Foundation (March 2005): 62.
- <sup>30</sup> Woodworth, K., David, J.L, Guha, R., Wang, H., & Lopez Torkos, A. (2008). *San Francisco Bay Area KIPP Schools: A Study of Early Implementation and Achievement, Final Report*. Menlo Park CA: SRI International: 10.
- <sup>31</sup> Anderson, A. B. & DeCesare, D. (2006). *Opening Closed Doors: Lessons form Colorado's First Independent Charter Schools*. Augenblick, Palaich, & Associates; report prepared for the Donnell-Kay Foundation & Piton Foundation. Augenblick, Palaich and Associates is a privately owned company that, since 1983, has "worked with state legislatures, education agencies and boards of education, district and county-level leaders, business leaders, foundations and numerous other groups interested in improving our country's education system." <http://www.apaconsulting.net>
- <sup>32</sup> "For example, in the subjects where 8<sup>th</sup> graders did best—math and science—CCP students were ranked below the 40<sup>th</sup> percentile nationally at the end of the school year. And in language [where there was the least improvement] these same students ranked in the 20<sup>th</sup> percentile nationally ...". Anderson, A. B. & DeCesare, D. (2006). *Opening Closed Doors: Lessons form Colorado's First Independent Charter Schools*: Augenblick, Palaich, & Associates; report prepared for the Donnell-Kay Foundation & Piton Foundation. pp. 24-5. KIPP subsequently dropped Cole, indicating that they were unable to place an appropriate principal there. See Mathews, J. (2007). "Looking at KIPP, Coolly and Carefully." *The Washington Post* (April 24, 2007). It is not clear whether that indicates a policy decision by the KIPP organization to be wary about involvement with reconstituting failed schools.
- <sup>33</sup> Musher, K.K, Musher, D.M., Graviss, E.A., & Strudler, R.M. (2005). Can an Intense Educational Experience Improve Performance on Objective Tests? Results from One Charter School. *The Educational Forum*. v. 69 (Summer): 352-366.
- <sup>34</sup> Musher, K.K, Musher, D.M., Graviss, E.A., & Strudler, R.M. (2005). Can an Intense Educational Experience Improve Performance on Objective Tests? Results from One Charter School. *The Educational Forum*. v. 69 (Summer): 361.
- <sup>35</sup> Mac Iver, M.A., & Farley-Ripple, E. (2007). *The Baltimore KIPP Ujima Village Academy, 2002-2006: A Longitudinal Analysis of Student Outcomes*: Johns Hopkins University Center for Social Organization of Schools, prepared for The Abell Foundation.
- <sup>36</sup> See Henig, J. (2008). *Spin Cycle: How Research is Used in Policy Debates: The Case of Charter Schools*. NY: Russell Sage Foundation/Century Foundation.

## What Do We Know About the Outcomes of KIPP Schools?

<sup>37</sup> In a couple of reasonably high profile cases KIPP has dropped schools because of performance or leadership problems. This appears to have been the case, for example, in the case of the Sankofa school in Buffalo, where KIPP discontinued its association, after sixth graders “finished the year with math scores that were eight percentile points lower than those same students achieved when they arrived at Sankofa's fifth grade two years before.” and also suffered a nine percentile point drop in language arts.” Mathews, J. (2007, April 24). “Looking at KIPP, Coolly and Carefully,” Washingtonpost.com. Retrieved Oct. 29, 2008, from [http://www.washingtonpost.com/wp-dyn/content/article/2007/04/24/AR2007042400558\\_pf.html](http://www.washingtonpost.com/wp-dyn/content/article/2007/04/24/AR2007042400558_pf.html)

The *Post's* Jay Mathews indicates that this, as of April 2007, was the sixth school in total to have left the network. In some other instances, factors other than poor performance or leadership problems were cited, including problems getting space to expand or, in the case of a Chicago school, because of problems attributed to the fact that the school was opened as a contract rather than charter school and dissatisfaction with this as a model.

<sup>38</sup> Educational Policy Institute. (2005). *Focus on Results: An Academic Impact Analysis of the Knowledge is Power Program (KIPP)*. Prepared for the KIPP Foundation. Retrieved Sept. 15, 2008, from <http://www.kipp.org/01/independentreports.cfm>. Information on excluded schools is found in the Appendix, p. 15.

<sup>39</sup> For a discussion and review of some of the relevant empirical literature vis a vis charter schools, see Henig, J. R., (2008). *Spin Cycle: How Research is Used in Policy Debates: The Case of Charter Schools*. NY: Russell Sage/Century Foundation.

<sup>40</sup> Howell and West raise this issue in their critique of a report by the AFT that claimed evidence that public schools outperformed charters: “Having just hired new staff and teachers, implemented new curricula, and acquired building facilities, new schools often face considerable start-up problems.”

West M.R. & Howell, W.G. (2005, Winter) “The Grey Lady Wheezing,” *Education Next* (vol. 5, no. 1).

The conservative Heritage Foundation, in summarizing a 2004 study by Caroline Hoxby, put it this way: “Further, Hoxby’s study shows that charter school students’ gains in academic achievement, relative to their public-school counterparts, tend to increase as the charter schools mature. In other words, charter schools are pulling away from public schools in terms of performance.” Retrieved Oct. 29, 2008, from [http://www.heritage.org/Research/Education/wm622.cfm#\\_ftn1](http://www.heritage.org/Research/Education/wm622.cfm#_ftn1)

<sup>41</sup> Ross et al. include an analysis of test scores in writing and language arts in addition to reading and math. Despite the Memphis KIPP: Diamond’s stated emphasis on writing across the curriculum, they did not find positive program effects, but neither did they find evidence that writing scores declined due to undue evidence on reading and mathematics.

<sup>42</sup> Henig, J. (2008). *Spin Cycle: How Research is Used in Policy Debates: The Case of Charter Schools*. NY: Russell Sage Foundation/Century Foundation: Chapter 5.