Do Impacts on Test Scores Even Matter? Lessons from Long-Run Outcomes in School Choice Research

ATTAINMENT VERSUS ACHIEVEMENT IMPACTS AND RETHINKING HOW TO EVALUATE SCHOOL CHOICE PROGRAMS

Collin Hitt, Michael Q. McShane, and Patrick J. Wolf

MARCH 2018

AMERICAN ENTERPRISE INSTITUTE

Executive Summary

Almost every major education reform of the past 20 years at both the state and national level has rested on a common assumption: Standardized test scores are an accurate and appropriate measure of success and failure. It has followed that programs or policies that increase student scores on standardized tests are "good" and programs that fail to do so are "bad."

This way of thinking was central to the No Child Left Behind Act. But the same logic was applied elsewhere. The past 20 years has seen explosive growth in school choice programs, and these programs have largely been evaluated based on their impacts on student test scores.[1]

Reading and math tests measure basic skills that almost everyone believes are important. Test scores are convenient to collect. Yet even the most fervent believer in the power of standardized tests agrees that test scores are merely an interim measure. There is no point in increasing test scores for test scores' sake. Increased test scores are supposed to indicate progress toward more important long-term results.

Perhaps the most convincing evidence of this supposed truth comes from Raj Chetty's seminal work that found connections between changes in students' test scores and the likelihood that they would graduate from high school or have children as teenagers and between changes in students' test scores and their earnings in their late 20s.[2] But other education research, especially involving school choice, sows doubt with respect to using test scores as the primary measure of program success.

A growing number of studies are finding that school choice programs can improve high school graduation rates, college attendance, and earnings—without producing gains in test scores. Conversely, studies of other school choice programs have found large short-term test score gains but no lasting benefits in terms of graduation rates or college attainment. Improving test scores appears to be neither a necessary nor sufficient condition for improving the later-life outcomes that truly matter.

We have attempted to collect every experimental and quasi-experimental study of school choice in the US that examines attainment impacts. Most of these studies also examine impacts on test scores. We use an expansive definition of school choice, including private school voucher programs, charter schools, early college high schools, magnet schools, and vocational schools. We compare impacts on test scores to impacts on later attainment outcomes. Our question is, across all studies, do program impacts on test scores predict impacts on later outcomes?

This review is one of the most thorough ever done of the school choice literature. We review every known study that contains participant-effect estimates for *both* student achievement and attainment. We exclude studies that look only at achievement scores. We take a simple analytical approach. We collapse findings into four categories: significantly positive, insignificantly positive, insignificantly negative, and significantly negative. We then map achievement findings against attainment findings.

Using such vote-counting methods, we find that, among these studies, program impacts on achievement are inconsistent, perhaps on balance weakly positive, thus replicating the school choice achievement findings of more sophisticated meta-analyses of the test score effects specifically of vouchers[3] and charters.[4] However, impacts on attainment are much more consistently positive. This pattern itself implies that some programs have produced clearer attainment impacts than achievement impacts, but the pattern of findings is actually more complicated.

Programs that produced no measurable positive impacts on achievement have frequently produced positive impacts on attainment.[5] And on the other

hand, null effects on high school graduation and college attendance have been reported from programs that produced substantial test score gains.[6] Across these studies, achievement impact estimates appear to be almost entirely uncorrelated with attainment impacts.

A school choice program's impact on test scores is a weak predictor of its impacts on longer-term outcomes. Our findings are based on 39 unique impact estimates across studies of more than 20 programs. In the coming months and years, more studies of school choice will be released. Perhaps over time a stronger connection between achievement and attainment impacts will emerge. We suspect not.

This pattern of findings is not unique to choice policies. The growing literature on early childhood education has found that short-term impacts on test scores are inconsistent predictors of later-life impacts. Some of the preschool programs that produced the most impressive improvements in later-life outcomes did so without producing lasting gains on test scores.[7]

Studies of teacher impacts on student outcomes show a similar pattern of results. As with school choice, it has been argued that teacher impacts on test scores should be used for policy purposes: Teachers who produce gains should be rewarded and promoted, and those who do not should be remediated or fired. "Value-added" methods have been developed that produce estimates of teacher-level impacts on test scores. Some researchers have used value-added methods to assess teacher impacts on other noncognitive outcomes.

It turns out that teacher impacts on test scores are almost entirely uncorrelated with teacher impacts on student classroom behavior, attendance, truancy, and grades.[8] Likewise, teacher impacts on test scores are uncorrelated with teacher impacts on self-reported noncognitive skills such as grit.[9] Teachers who possess higher noncognitive skills boost the noncognitive skills of their students but not student test scores.[10] In short, the teachers who produce improvements in student behavior and noncognitive skills are not particularly likely to be the same teachers who improve test scores. Our findings suggest that the same appears to be true of schools of choice.

Our findings beg serious questions about using standardized tests as the exclusive or primary metric on which to evaluate school choice programs. If test score gains are neither a necessary nor a sufficient condition for producing long-term gains in crucial student outcomes, then current approaches to accountability for school choice programs are questionable at best. Our findings suggest that focusing on test scores may lead authorities to favor the wrong school choice programs. Focusing on test score gains may lead regulators to favor schools whose benefits could easily fade over time and punish schools that are producing long-lasting gains.

Do Impacts on Test Scores Even Matter? Lessons from Long-Run Outcomes in School Choice Research

ATTAINMENT VERSUS ACHIEVEMENT IMPACTS AND RETHINKING HOW TO EVALUATE SCHOOL CHOICE PROGRAMS

Collin Hitt, Michael Q. McShane, and Patrick J. Wolf

A careful read of the existing literature on school choice presents a paradox. In 2010, a federally funded evaluation of a school voucher program in Washington, DC, found that the program produced large increases in high school graduation rates after years of producing no large or consistent impacts on reading and math scores.[11] Conversely, a recent evaluation of Boston charter schools found no effects on high school graduation and null effects on college attendance after previous evaluations had found remarkably large impacts on reading and math scores.[12]

These findings appear to go against the grain of the current logic model of education policy. Much of the federal and state education policy of the past two decades has been driven by the assumption that test scores are a meaningful and important measure of what children need to know.

Test scores are by far the most popular short-term outcome used in education research and program evaluation.[13] Using short-term outcomes is understandable—stakeholders do not want to wait years, even decades, to know whether a program is effective. In school choice research, the number of studies that examine test score impacts far outnumber studies that examine later-life outcomes such as high school graduation, college attainment, and employment income.[14]

Test scores have become easier and cheaper than ever for researchers to use. Standardized test scores can be administered to any child from third grade on up, efficiently, on a large scale. If the tests are well designed, they reliably measure skills that we think are important, including language arts and mathematics. In the wake of No Child Left Behind, virtually every state already collects test scores on students between third and eighth grade and at least once again during high school. Gaining access to state-collected data is far less labor intensive than collecting one's own data. The ready availability of test scores makes research on math and reading skills convenient—but is such research important?

For research on test scores to actually be meaningful, the following should be true: The impacts that schools have on math and reading skills will change the trajectories of children's lives. Otherwise, why would policymakers and researchers put such emphasis on "student achievement" and "student growth"—measures that are based on test scores?

This assumption seems uncontroversial. It is well-known that childhood test scores and later outcomes are strongly correlated.[15] It seems sensible that boosting reading and math skills would boost later educational attainment. Children who know more will go further in school, right? Yet recent findings from prominent school choice studies present a puzzling picture. Test score effects are disconnected from attainment effects.

These seemingly paradoxical findings motivated this study. To find the best evidence on the question of the connection among school choice, test scores, and later-life outcomes, we have carried out what we believe to be the most expansive review of the scholarly literature on the impact that school choice programs in the United States have had on educational attainment. We have done so to determine the frequency with which the same school choice programs are found to have different impacts on test scores and educational attainment.

We use as broad a definition as possible for school choice. We do so for two reasons. First, we wanted to gather the largest number of studies possible to examine the relationship between achievement and attainment impacts across studies. Second, school choice is bigger than voucher programs and charter schools. Many large districts have embraced a portfolio model of school choice governance, which intentionally offers a wide array of public (and sometimes private) school choices to parents. The diversity of the studies we collected mirrors the diversity of choice options that portfolio school districts attempt to offer.

In the following section, we briefly describe our search strategy for gathering studies and our screening methods for including studies in our overall review. In the section thereafter, we describe the studies included by school choice type: private school voucher programs, open enrollment programs, charter schools, selective enrollment schools, career and technical schools, inclusive STEM schools, early college high schools, and small schools of choice. We then present our aggregate findings and discuss technical considerations to keep in mind. Finally, we discuss the important implications of our findings for current policy and future research.

Methods

The goal of our systematic literature search was to identify studies that examined the effect of school choice on reading and math scores, high school graduation, college attendance, and completion of a four-year college degree. We limited our search to studies set in the United States.

We searched the following terms on Google Scholar:

- "school choice" + "attainment"
- "school choice" + "graduation"
- "charter school" + "attainment"
- "charter school" + "graduation"
- "school voucher" + "attainment"
- "school voucher" + "graduation"

We then screened the titles of the first 200 studies returned for each search term. If the title appeared in any way relevant, the study was logged for a review of its abstract. The abstract review process essentially screened for a sufficiently rigorous study design and relevant outcomes.

Any study not eliminated during the abstract review was given a full reading, whereon we determined the research methods used and the outcomes of interest. Only quantitative studies were included in our analysis. Studies with data flaws—such as major problems with attrition from the study—were eliminated, as were studies that failed to report point estimates of effect sizes on test scores and attainment outcomes. Studies that passed these standards were included in our main analysis.

Studies selected for inclusion in our main analysis then formed the basis for a secondary round of searches. We searched the citations *in* each study, beginning with a title search as outlined above, and the citations *of* each study, as indicated by Google Scholar. We also searched the entire publication history of the author of every included study. Finally, we searched the entire databases of research centers that supported or published the studies selected, including the American Institutes of Research, the Brookings Institution, the Everyone Graduates Center

at Johns Hopkins University, Mathematica Policy Research, Manpower Demonstration Research Corporation (MDRC), the National Bureau of Economic Research, the Rand Corporation, the School Choice Demonstration Project at the University of Arkansas, and SRI International. In each instance that a title was identified through the secondary search, the entire process described above was repeated.

This meta-analysis focuses on the relative effect that individual school choice programs have had on achievement and attainment. During our search, we identified several rigorous studies that examined attainment impacts but not achievement impacts.[21] We identified many more studies that examined only achievement impacts. Ultimately, a study (or a series of studies) was included only if it provided both achievement and attainment impacts.

In total, we included studies of more than 20 programs, which provided 39 unique estimates of the impact that school choice programs have had on achievement and attainment. Findings from each study were recorded using the following guidelines.

Most studies treated test scores as a continuous variable of scale scores or percentile scores. Some studies treated test scores as a binary variable of students scoring above or below a given proficiency point. Some studies did both. We analyze the impacts on continuous test scores, whenever available.

With respect to high school graduation, some studies examined multiple windows of time for high school or college graduation (e.g., four or six years). We always used estimates taken from the longest time frame examined, in a given study.

In three instances, studies examined impacts on high school dropout rates instead of high school graduation rates, typically because sufficient time had not yet passed to observe actual graduation rates. Dropout rates and graduation rates are naturally correlated highly, though they are not perfect substitutes. When only dropout rates were included in a study, we used those effects as proxies for impacts on high school graduation.

Our goal is to compare the effects on reading and math scores to effects on high school graduation, college attendance, and completion of a four-year college degree. Ideally, we would simply examine the correlation between program effect sizes on each outcome. But several complications arise when examining effect sizes.

Not all studies express results in effect sizes. For example, a regression table may report an effect on test score percentiles. To transform these results into an effect size that is comparable across studies, information is needed about the standard deviation for that variable across the sample and ideally across just the control or comparison group as well. Not all studies included this information.

Furthermore, even if sufficient information were available to compute effect sizes, one would also have to take into account the precision of the point estimates when examining the correlation between effect sizes. The precision of a given estimate determines whether a finding is considered statistically significant. Not all studies provide exact information on the precision of their estimates—that is, rather than reporting standard errors or p-values, some studies simply indicate with stars whether an observed effect is above or below a given significance threshold.

We take the following approach in comparing attainment findings to achievement findings, which allows us to maximize the number of studies in our analysis while also taking into consideration the size and precision of each estimate. We code each finding in one of four ways: positive and statistically significant, positive and statistically insignificant, negative and statistically insignificant, and negative and statistically significant. It is debatable whether the data, coded as such, should be treated as nominal or ordinal variables. This distinction makes little difference to our findings.

Our coding scheme may seem rigid. One can criticize the exactitude that we are imposing on the data (achievement and attainment results must match regarding both direction and significance). We believe this exactitude is justified by the fact that the conclusions one draws about whether school choice "works" depends on the direction and significance of the effect parameter. In other words, we are testing whether the results of the hypothesis tests are similar

Methodology Screen

Schools of choice are attended by students, *by choice*. Researchers can easily observe test scores and educational-degree attainment after students enroll in schools of choice. The first fundamental question of program evaluation is: What would those outcomes have looked like if students had not attended that school?

This question is difficult to answer because students who select to attend schools of choice are obviously different—by virtue of their choice of school—than students who elected to or who had no choice but to attend school elsewhere. This creates a problem called selection bias—a problem that affects everything from educational to pharmaceutical research. Students who select schools of choice may naturally score differently on tests and graduate at different rates than students who do not exercise school choice, in ways that researchers cannot observe.

Various methods exist for addressing selection bias, some more rigorous than others. For a study to be included in our meta-analysis, it must have used one of the following research designs.

Random Assignment. In an ideal research setting, students are randomly selected to participate in a program. This approach is often considered the gold standard of education research because it eliminates the problem of selection bias.[16] Students are offered enrollment (or not) in a program due to random chance, rather than their own pluck, intelligence, desperation, or other factor not captured in the data. Indeed, in school choice programs, random assignment is often used.[17] Schools of choice frequently receive more applications than there are seats available, and state laws usually require in these cases that a random lottery be used to admit students.

Comparing lottery winners to lottery losers produces intent-to-treat estimates (ITT). In practice, however, many students who win school choice lotteries do not take up the offer, and some students who lose the lottery occasionally find a way to enroll in the school of choice anyway. There are various methods that adjust for noncompliance with lottery results, which produce treatment-on-treated (TOT) or local average treatment effects (LATE).[18] Whenever possible, we record the TOT or LATE effects, rather than ITT.

Regression Discontinuity Design (RDD). Some schools do the opposite of accepting students by lottery. They select students based on prior performance in school. In the case of magnet schools, for example, students must achieve a given test score to be accepted. Those applicants just above the cutoff point are eligible to attend, whereas those just below the cutoff point are not. This creates a natural, lottery-like condition due to two factors. First, the cutoff is arbitrary. Second, the minor differences in test scores just above and just below the cutoff are driven mostly by random measurement error in testing. The cutoff score creates an artificial discontinuity in the likelihood that students will be accepted to a given school. Regression discontinuity studies examine the performance of students just above and just below the performance cutoffs. Regression discontinuity studies cannot directly say how well students in the very top or the very bottom of the test score distribution would perform if attending magnet schools. These studies are the closest approximation to random assignment, in terms of rigor, but results must also be interpreted as valid only for students who tested near the cutoff point.[19]

As in lottery-based studies, some students above the test score cutoff choose to attend school elsewhere, and those below the cutoff occasionally find a way to attend the selective schools. Again, TOT or LATE estimates can be produced in RDD studies, and these adjusted estimates are what we record for our meta-analysis, when available.

(*continued from the previous page*)

Instrumental Variables. The methods outlined above exploit random factors that affect the likelihood that a student will attend a school of choice. Other near-random factors may affect where students choose to attend school. For example, when a new school of choice opens, students will face different travel times to the school, based on where they live. Even minor differences in travel time for students from the same neighborhood might affect their relative likelihood of attending that school, even if their minor differences in home location would not otherwise have made a difference on academic outcomes. Such near-random shocks to the likelihood that students attend a school of choice can be used to produce estimates of the effect of attending that school, in ways that eliminate selection bias, using instrumental variables methods. Generally, for a near-random event to be validly used in an instrumental variables model, it must be relevant to the student's selection of a school, and it must have an effect on the outcomes of interest *only through* its effects on the likelihood of a student attending that school of choice.[20]

Matching Methods. Often, none of the research conditions above can be met. In such instances, researchers often match students attending schools of choice to another group of non-choice students considered similar at baseline.

The most prevalent matching method used in school choice research is propensity score matching. All known characteristics for a student—demographics, prior test score, prior school of attendance, home address, etc.—are used to find a student attending another school who possesses identical or nearly identical background characteristics. This is done for every possible student in a school of choice. Naturally, in terms of background characteristics, the matched-comparison group will on average be nearly identical to the group of students attending the school of choice. Differences that emerge between the two groups on later outcomes are used to estimate program impacts.

Statistical Modeling to Control for Differences. Finally, some school choice researchers include in their study sample choice and comparison students with observable differences but then use control variables in statistical models to adjust for those differences. Along with matching, statistical modeling makes an important assumption that the variables included in the analysis are sufficient to control for all relevant differences between students who do and do not attend a given school of choice. This assumption is almost certainly untrue, and as a result, matching and modeling methods are likely more susceptible to selection bias than other research designs. Even with their limitations, matching and modeling methods take more care to isolate the possible *causal impact* that schools of choice have on outcomes than do the performance ratings used in state K–12 accountability systems or in portfolio management systems. We include studies using matching or modeling approaches as the least rigorous analyses in our meta-analysis.

if achievement or attainment are the chosen outcomes in a given study. If they are not, as we generally show, then the selection of outcome is crucial to the assessment of school choice. We would make the case that attainment is a more important outcome.

Summary of Studies

There are several policy options for offering school choice. The studies we collect cover several of these mechanisms.

One mechanism is private school vouchers, in which students are awarded vouchers to cover tuition at any one of numerous private schools of their choosing. Another mechanism is public school open enrollment programs, in which students can choose to attend whichever local public school they want. These might be labeled macro choice programs.

Another mechanism is creating specific schools of choice. We present studies of what might be labeled micro choice programs sorted into the following categories, which are not entirely mutually exclusive: charter schools, selective enrollment high schools, career and technical schools, early college high schools, inclusive STEM schools, and new small schools of choice.

Private School Choice. The most controversial form of school choice has long been private school vouchers, in which students receive financial assistance to attend private schools. Publicly funded voucher programs are a relatively new phenomenon. For years, privately funded but publicly available voucher programs existed as demonstration projects in many cities—including New York City, San Antonio, Charlotte, and Cleveland—whose purpose was to pilot a model for what school choice supporters hoped would become a blueprint for publicly funded programs in the future. Wisconsin created the nation's first publicly funded urban voucher program in 1990. The first and so far only federally funded school voucher program was created in 2003 to serve students in Washington, DC.

In New York City in 1997, a privately funded school voucher program was created after a proposal from the archdiocese of New York was rejected that would have sent students from the worst-performing public schools to Catholic schools at public expense. The private program, naturally limited in size, was massively oversubscribed. Any elementary school student in the city with family income below 270 percent of the poverty line was eligible to apply for the vouchers, which were awarded by lottery.

In the ensuing years, multiple research teams exploited the random assignment of vouchers to students to conduct evaluations of the program's impact on standardized test scores. Whereas heated methodological debates ensued regarding the program's impacts on African American students, there was a consensus that the program overall had positive but insignificant impacts on test scores.[22] Researchers from Harvard and the Brookings Institution years later were able to explore the program's impacts on college attendance, finding that the program had an overall positive but insignificant impact on college attendance rates and college graduation rates.[23]

The Milwaukee Parental Choice Program was created by the State of Wisconsin in 1990. The only evaluation of the long-term effects of the program was conducted by researchers at the School Choice Demonstration Project at the University of Arkansas. They used a matching design, in which voucher program participants were matched to demographically similar students from the same neighborhoods with similar baseline test scores. Separate studies were conducted on the program's impacts on achievement and attainment.[24] Only one cohort of students—those attending eighth grade in 2006–07—were present in both analyses. The results for this cohort of students are included in this meta-analysis: The program had a negative and significant impact on English language arts (ELA) scores, a negative but insignificant impact on math scores, and a positive but insignificant effect on high school graduation rates.

The Opportunity Scholarship Program in Washington, DC, was created by an act of Congress in January 2004. Low-income students living in the District of Columbia were eligible to receive $7,500 vouchers to attend private schools. Total funds were limited, and demand was high, so vouchers were awarded by lottery. A federally funded evaluation led by the School Choice Demonstration Project at the University of Arkansas exploited the random assignment of vouchers to evaluate the program's impacts on achievement scores and high school attainment.[25] The program had positive but insignificant impacts on both language

arts and math scores and significant positive impacts on high school graduation rates.

Public School Open Enrollment Programs. Some school districts have moved away from systems in which high school attendance is determined solely by neighborhood of residence and toward a system in which students choose from existing high schools. Open enrollment programs effectively make so-called neighborhood schools into schools of choice, without necessarily affecting changes in the schools' management structure or curricular focus. A consistent feature of other school choice programs explored in this meta-analysis is the freedom that schools have to pursue a particular focus, hire preferred teachers and staff, or set their own budgets. Such liberties are not necessarily available to schools in open enrollment programs.

Chicago has long had a system of open enrollment at the high school level. Some schools are oversubscribed, and lotteries are used to determine admissions. Researchers at the University of Chicago exploited these admissions lotteries to evaluate the impact of being admitted to one's school of choice on achievement tests and degree attainment.[26] Broadly, the schools included in this analysis are old and large—unlike the newer and smaller schools studied in other work contained in this meta-analysis. Winning a lottery to attend one's school of choice had negative but insignificant effects on language arts and reading scores and a significant negative impact on high school graduation.

Charlotte-Mecklenburg School District in North Carolina moved to a system of open enrollment for its high schools in 2002. Students registering for high school were asked to rank their preferred schools of choice. As in Chicago, schools that were oversubscribed used lotteries to admit students. Researchers have examined the impact of winning a lottery to attend one's top school of choice.[27] Winning a lottery to one's top school of choice had a negative but significant effect on language arts and math scores, a positive but insignificant impact on high school graduation, a negative but insignificant impact on college enrollment, and a positive and significant impact on completing a four-year college degree.

Charter Schools. Public charter schools are by far the fastest-growing choice option in the United States. The precise definition of charter schools varies by state, but broadly charter schools are privately operated public schools of choice that have far more operational autonomy than traditional public schools. In most states, charter schools are free from the hiring, procurement, and collective bargaining rules that are imposed on traditional public schools. One nearly universal requirement of charter schools is that they be open to all students. In the case that charter schools are oversubscribed, they typically must admit students by lottery.

For high school students, charter schools had a significant and positive effect on language arts and math scores and a positive but insignificant effect on high school graduation.

New York City is home to hundreds of charter schools, most of which are oversubscribed and must admit students by lottery. Researchers at the National Bureau of Economic Research used lottery results to conduct a random-assignment evaluation of charter school impacts on achievement tests and high school graduation.[28] For high school students, charter schools had a significant and positive effect on language arts and math scores and a positive but insignificant effect on high school graduation.

Subsequent evaluations for individual New York City charter school networks have since been released, covering different time frames. The Harlem Children's

Zone is home to the Harlem Promise Academies, a network famous for the extra social services that its students receive. The Promise Academies are oversubscribed and admit students by lottery, which has allowed for a random-assignment evaluation of the schools' impacts.[29] The network had a positive but insignificant effect on language arts, a positive and significant impact on math scores, a positive but insignificant impact on high school graduation, and a positive but insignificant impact on college attendance.

Likewise, Boston has a large number of students enrolled in charter schools. There, schools are often oversubscribed as well, and enrollment lotteries are used to admit students. Exploiting these lotteries, researchers at the Massachusetts Institute of Technology and elsewhere have determined that Boston charter schools have positive and significant effects on language arts, positive and significant impacts on math scores, negative but significant impacts on high school graduation rates, and positive but insignificant impacts on college attendance rates.[30]

Chicago too has a sizable charter schools sector. An evaluation team at the Rand Corporation and Mathematica used a combination of matching and instrumental variables methods to estimate the impact of attending a charter high school.[31] Essentially, students attending charter schools in eighth grade were followed into high school—some attended charter schools in high school, and others did not. The two groups were compared to assess the effects of attending a charter high school. The authors estimated that charter high schools in Chicago had a positive but insignificant impact on language arts, a positive and significant impact on math scores, a positive and significant impact on high school graduation, and a positive and significant impact on college attendance.

Likewise, Florida has a large charter school sector. The research team that conducted the evaluation of charter high schools in Chicago conducted an evaluation using identical methods in Florida.[32] Charter high schools in Florida were estimated to have positive but insignificant effects on language arts and math scores, a positive and significant effect on high school graduation rates, and a positive and significant impact on college attendance. The positive impacts were followed by positive impacts on earnings, in the only charter school study that examines workplace outcomes.

The Seed Charter School is a network of residential prep schools of choice. Researchers at MDRC exploited admissions lotteries at the network's DC campus to conduct a random-assignment evaluation of the school's impacts.[33] The school was initially found to have sizable, positive impacts on language arts and math scores. However, findings from the longer-term evaluation underscore the importance of tracking impacts over time. After three years, Seed attendance had a negative but insignificant impact on language arts, a positive but insignificant impact on math, and negative but insignificant impacts on high school graduation. The fade-out of achievement effects at Seed are particularly relevant to our overall meta-analysis.

In Los Angeles, a unique combination of education and medical researchers sought to evaluate the impact of five charter high schools that had received high ratings in California's state accountability system.[34] The schools used lotteries to admit students, which allowed the researchers to conduct a random-assignment evaluation of the schools' impacts. The schools had positive and significant effects on language arts and math scores and significantly lowered dropout rates. The schools also positively affected self-reported health-related behaviors.

Texas has one of the largest and oldest charter school sectors in the country. A small number of the state's charter high schools were included in an evaluation of the Texas High School Redesign Initiative. The evaluation, conducted by SRI International and released by the Texas Department of Education, used a school-level matching design, in which charter students were compared in performance to students attending demographically similar schools.[35] The charter schools had positive and significant impacts on language arts and math scores. The schools also had reduced dropout rates, but the effect was insignificant.

A separate evaluation of Texas charter schools was conducted in 2016 by Will Dobbie and Roland Fryer, using propensity score matching methods.[36]

Results were reported separately for "no excuses" and "other" charter schools. "No excuses" charter schools in Texas produced significant gains in ELA and math scores and in high school graduation rates. The "other" kinds of Texas charter schools also produced significant gains in high school graduation rates, despite having negative but significant impacts on ELA and math scores.

Undoubtedly the most famous charter school network in the country is the Knowledge Is Power Program (KIPP). The vast national network of KIPP schools has for years been the subject of an ongoing evaluation by Mathematica researchers. The network's large, positive effects at the middle school level are well-documented.[37] KIPP more recently expanded by opening high schools. Researchers at Mathematica used matching methods to evaluate the impact of KIPP high schools on achievement and attainment.[38] One component of the evaluation examined the effect of attending a KIPP high school on students who had previously attended a KIPP middle school, some of whom did not attend KIPP high schools. Among these students, attending a KIPP high school had positive but insignificant effects on language arts and math scores while significantly reducing high school dropout rates. Another component of the evaluation examined high school freshman attending a KIPP school for the first time compared to other high school students who had also never previously attended KIPP schools. For these students, KIPP high schools had positive and significant effects on language arts and math scores and positive but insignificant effects on high school graduation rates.

KIPP is not the only large charter management organization (CMO) in the United States. In 2012, an evaluation Mathematica Policy Research led examined outcomes across a large number of CMOs.[39] A component on the analysis examined the impacts at CMOs that operated high schools. Enrollment lotteries were used to calculate experimental impacts. Impacts for *both* achievement and attainment were reported for three pseudonymous groups. "CMO 2" had positive but insignificant impacts on ELA scores and positive and significant impacts on high school graduation and college attendance. "CMO 5" had significant and positive impacts on ELA and math scores and positive but insignificant impacts on high school graduation. "CMO 6" had negative and insignificant impacts on ELA and negative but significant impacts on math and high school graduation.

Early College High Schools (ECHS). ECHSs are small schools of choice typically located on college campuses. ECHS students receive college credit and can even complete college degrees while still in high school. Classes are sometimes taken with high school instructors and at other times with college professors, alongside normal college students. Some early colleges are operated by traditional public school districts, and others are operated as charter schools.

The American Institutes for Research and SRI International conducted a random-assignment study of ECHS impacts, exploiting admissions lotteries at 10 ECHSs from five states.[40] The schools had a positive and significant effect on language arts, a positive but insignificant effect on math scores, and positive and significant impacts on high school graduation, college attendance, and college graduation.

SRI International also conducted an evaluation of early colleges for the Department of Education in Texas, a state where ECHSs are relatively numerous and popular. The study used a school-level matching design, in which students at ECHSs were compared to students at demographically similar schools.[41] Texas ECHSs had positive but insignificant effects on language arts and math scores and on high school graduation rates.

North Carolina has also heavily embraced ECHSs. Research on the state's ECHSs is the most rigorous research done so far. In an evaluation led by the Serve Center at the University of North Carolina at Greensboro, researchers have used enrollment lotteries to examine experimental impacts of schools.[42] Early colleges in North Carolina have a positive and significant impact on ELA scores but a negative and insignificant impact on math scores. Much larger positive and significant impacts were shown for high school graduation, college attendance, and the completion of a four-year college degree.

A separate analysis of North Carolina ECHSs was conducted using matching methods.[43] Findings agree with Julie Edmunds and her colleagues on high school graduation and ELA impacts, but not with respect to math scores. We note these findings here. However, in keeping with our methods screen, we include only the estimates from Edmunds and her colleagues—who use a stronger research design—for our overall analysis.

Selective Enrollment High Schools (SEHS). Variously known as "exam schools" or "magnet schools," SEHSs are typically seen as elite public schools—and they are some of the most sought-after schools of choice in many cities. They attract some of the highest-achieving students from throughout their school districts. Students are admitted, at least in part, based on test scores.

The test-based admissions component allows researchers to use an RDD to evaluate the school's effectiveness. This exploits the fact that students who score barely above or barely below the cutoff are essentially identical in prior achievement—their differences in scores are more likely due to the imprecision of the tests rather than a true difference in ability.

Separate research teams in Boston, Chicago, and New York City have used the admissions cutoffs at highly popular SEHSs to obtain plausible estimates of school effects on achievement and attainment.

In Boston, SEHSs had negative but insignificant impacts on language arts and math scores and positive but insignificant impacts on college enrollment.[44]

In the evaluation of Chicago SEHSs, language arts and math were pooled into a combined score on the ACT, which Illinois long has used as its standardized test for high school students. SEHSs in Chicago had negative but significant effects on ACT scores and high school graduation rates and positive but insignificant effects on college attendance.[45] We note this finding here but do not include this study in our overall analysis because ELA and math scores are not disaggregated.

In New York City, researchers did not present the pooled effects of attending a SEHS. The effects were reported separately for three of the city's SEHSs: Brooklyn Tech, Bronx Science, and Stuyvesant.[46] Brooklyn Tech had a positive but insignificant effect on language arts scores, a positive and significant effect on math scores, a positive but insignificant effect on high school graduation, a negative but insignificant effect on college attendance, and a negative but insignificant effect on college graduation. Bronx Science had a negative but insignificant effect on language arts scores, a positive but insignificant effect on math scores, a positive and significant effect on high school graduation, a positive but insignificant effect on college attendance, and a negative but insignificant effect on college graduation. Stuyvesant had a negative but insignificant effect on language arts scores, a positive but insignificant effect on math scores, a positive but insignificant effect on high school graduation, a negative but insignificant effect on college attendance, and a negative but insignificant effect on college graduation.

Career Academies and Vocational-Technical Schools. Vocational schooling is a well-known and relatively uncontroversial form of educational choice for high school students. As a school choice option, vocational schooling exists basically in two forms: either as career academies or career and technical schools.

Career academies are typically organized as schools within a school, colocated on a high school campus but with their own autonomy and identity. They are attended by choice. Admission to career academies is often in high demand, and some schools admit students by lottery. Researchers at MDRC exploited these admissions lotteries at nine career academies across the country to conduct a random-assignment evaluation of the schools' impacts on test scores, attainment, and earnings.[47] Attending a career academy had a negative but insignificant effect on ELA scores, a positive but insignificant effect on math scores, and positive but insignificant effects on high school graduation, college attendance, and graduating with a four-year college degree. Despite the insignificant effects on achievement and attainment, however, career academies had positive and significant impacts on employment earnings.

A more recent study of a tech-focused career academy comes from North Carolina. Researchers exploited enrollment lotteries, allowing for an experimental research design.[48] The academy has positive but insignificant impacts on ELA and math and positive and significant impacts on high school graduation. Impacts on college attendance were positive but insignificant.

Larger stand-alone career and technical schools provide high school students with vocational-training options in many cities. Philadelphia in particular has a large number of such schools, which admit students by lottery. Researchers at the Everyone Graduates Center at Johns Hopkins University have used admissions lotteries to conduct a random-assignment study of the schools' impacts on test scores and attainment.[49] Results were reported separately for the lottery cohorts of 2003, 2004, and 2005.

For the 2003 cohort, career and technical education (CTE) schools had negative but insignificant impacts on ELA and math scores, positive and significant effects on high school graduation and college attendance, and positive but insignificant effects on graduating with a four-year college degree. For the 2004 cohort, CTE schools had positive and significant impacts on ELA and math scores, high school graduation, college attendance, and graduating with a four-year college degree. For the 2005 cohort, CTE schools had positive but insignificant impacts on ELA and math scores, positive but significant impacts on high school graduation, positive and significant impacts on college attendance, and positive but insignificant impacts on graduating with a four-year college degree. The pattern of results from Philadelphia CTE schools speaks to the theme of our overall meta-analysis. Impacts on ELA and math scores varied substantially by year and were mixed overall. Findings on attainment were consistently positive.

Massachusetts is also home to several large, decades-old regional vocational schools. These are full-time schools of choice, permitted to admit students partly based on incoming test scores. A recent evaluation used multiple designs to examine the schools' impacts.[50] Some Massachusetts regional vocational schools use test score cutoffs to admit students; the impact of the schools on students near the cutoff was examined, using an RDD. The schools had positive but insignificant impacts on aggregate test scores and positive and significant impacts on high school graduation. The same study also used propensity score matching to examine impacts on a larger population of students, likewise finding positive and significant impacts on aggregate test scores and high school graduation. These findings are noted here but not included in our overall analysis, as ELA and math scores are not disaggregated.

Larger stand-alone career and technical schools provide high school students with vocational-training options in many cities.

Inclusive STEM (I-STEM) Schools. In the United States, there is a perceived shortage of workers in fields requiring knowledge in STEM. This perception has led to the creation of STEM-focused high school programs. However, this response has led to a subsequent concern that new opportunities to pursue studies in STEM fields are open mainly to students labeled as gifted or advanced. To offer the opportunities to study in STEM fields to a broader group of students, many states have launched initiatives to open "inclusive STEM schools"—nonselective schools of choice that provide a STEM-focused education.

The only publicly available study of I-STEM schools that examines attainment impacts comes from Texas, a state that emphasized creating I-STEM schools as part of a broader high school "redesign" effort. An evaluation of the redesign effort matched students at I-STEM schools to students at demographically similar Texas schools and used statistical

modeling to estimate the effects of attending an I-STEM school.[51] Texas I-STEM schools had positive but insignificant impacts on ELA and math scores and negative but insignificant impacts on high school graduation rates.

Small Schools of Choice (SSC). After charter schools, SSCs are perhaps the fastest-growing type of school choice in many districts. Chicago and especially New York City have embraced the idea that the district itself should facilitate creating new small high schools with total enrollments of around 400 students.

In New York City more than 100 SSCs were opened during the early 2000s. The schools were given operational autonomy and permitted to pursue unique academic missions, while their teachers were still covered by the same collective bargaining agreement as teachers in other district schools. Lotteries were used to admit students to schools that were oversubscribed. Multiple research teams have exploited these admissions lotteries to conduct random-assignment studies of SSC impacts on achievement and attainment and have reached the same conclusions about the effectiveness of the schools: New York City SSCs had positive and significant impacts on ELA scores, positive but insignificant impacts on math scores, and positive and significant impacts on high school graduation and college attendance.[52]

Chicago Public Schools followed a similar strategy as the New York City Department of Education. Dozens of small, autonomous schools were opened over a short period of time. The rapid appearance of SSCs created a natural research experiment. Some students suddenly had small schools of choice located close to their homes, whereas for other students the distance to a SSC was slightly farther away. Researchers have used this somewhat random variation in distance to SSC as an instrumental variable for SSC attendance to estimate the impact that SSCs have on student achievement and attainment.[53] SSCs in Chicago had a negative but insignificant impact on ELA and math scores and a significant and positive impact on high school graduation rates.

Summary of Results

We are examining a specific question: whether the relative impacts on achievement are correlated with the relative impacts on attainment for the population of empirical studies that examine both outcomes. It is important to note that we present results only from studies that examine impacts on *both* achievement and attainment. We exclude several school choice studies that examine only attainment impacts, and we exclude the large number of choice studies that examine only test score impacts. So we offer an important warning before digging into our main findings: Our results should in no way be interpreted as the overall effect that school choice has had on achievement test scores. Likewise, our findings do not represent the totality of evidence that school choice has had on attainment.

Overall, we identify 36 unique estimates of impacts on reading and math tests from studies that also examine impacts on attainment. Among those studies, 34 contain unique estimates of the impact on high school graduation, 19 contain unique estimates of the impact on college enrollment, and 11 contain unique estimates of the impact on completion of a four-year college degree.

Tables 1, 2, and 3 summarize the results by outcome. The impacts on achievement tests together lean slightly positive, though most findings are statistically insignificant. The impacts on attainment altogether lean more heavily positive.

Among ELA impact estimates, 11 (32 percent) are positive and significant, 13 (38 percent) are positive and insignificant, seven (21 percent) are negative and insignificant, and three (9 percent) are negative and significant. Among math impact estimates, 11 (33 percent) are positive and significant, 15 (46 percent) are positive and insignificant, six (18 percent) are negative and insignificant, and one (3 percent) is negative and significant.

Among attainment high school graduation impact estimates, 16 (47 percent) are positive and significant, 13 (38 percent) are positive and insignificant, three (9 percent) are negative and insignificant, and two (6 percent) are negative and significant. For estimates

Table 1A. High School Graduation Impacts Versus ELA Impacts: Number of Estimates by Sign and Statistical Significance

		High School Graduation Impacts				
		Negative Significant	Negative Insignificant	Positive Insignificant	Positive Significant	Total
ELA Impacts	Negative Significant	1	0	1	1	3
	Negative Insignificant	0	1	3	3	7
	Positive Insignificant	1	1	5	6	13
	Positive Significant	0	1	4	6	11
	Total	2	3	13	16	34

Pearson chi2(9) = 5.8434 Pr = 0.755, gamma = 0.2047 ASE = 0.223

Source: Authors' calculations.

Table 1B. High School Graduation Impacts Versus Math Impacts: Number of Estimates by Sign and Statistical Significance

		High School Graduation Impacts				
		Negative Significant	Negative Insignificant	Positive Insignificant	Positive Significant	Total
Math Impacts	Negative Significant	0	0	0	1	1
	Negative Insignificant	1	0	2	3	6
	Positive Insignificant	1	2	5	7	15
	Positive Significant	0	1	6	4	11
	Total	2	3	13	15	33

Pearson chi2(9) = 4.9008 Pr = 0.843, gamma = –0.1211 ASE = 0.237

Source: Authors' calculations.

of impact in enrolling in college, nine (47 percent) are positive and significant, seven (37 percent) are positive and insignificant, three (16 percent) are negative and insignificant, and none are negative and significant. Only 11 studies in our analysis examine the effects on college completion, three (27 percent) find significant positive effects, five (46 percent) find positive and insignificant effects, and three (27 percent) find negative and insignificant effects.

Broadly, the effects on attainment outcomes are skewed somewhat more heavily positive than are the impacts on achievement. While the difference in skewness is not major, it nevertheless implies that some studies have found positive effects on attainment without finding positive effects on achievement—as was the case with the evaluation of the DC Opportunity Scholarship Program.[54]

Multiple factors could lead to a significant finding for attainment to follow an insignificant finding on achievement. One possible explanation is measurement error: Test scores are a noisy, imperfect measure of reading and math ability, while graduation and

Table 2A. College Attendance Impacts Versus ELA Impacts: Number of Estimates by Sign and Statistical Significance

		College Attendace Impacts				
		Negative Significant	**Negative Insignificant**	**Positive Insignificant**	**Positive Significant**	**Total**
ELA Impacts	Negative Significant	0	0	0	0	0
	Negative Insignificant	0	2	3	1	6
	Positive Insignificant	0	1	3	4	8
	Positive Significant	0	0	1	4	5
	Total	0	3	7	9	19

Pearson chi2(4) = 5.0114 Pr = 0.286, gamma = 0.6883 ASE = 0.191

Source: Authors' calculations.

Table 2B. College Attendance Impacts Versus Math Impacts: Number of Estimates by Sign and Statistical Significance

		College Attendace Impacts				
		Negative Significant	**Negative Insignificant**	**Positive Insignificant**	**Positive Significant**	**Total**
Math Impacts	Negative Significant	0	0	0	0	0
	Negative Insignificant	0	1	1	2	4
	Positive Insignificant	0	1	4	4	9
	Positive Significant	0	1	2	2	5
	Total	0	3	7	8	18

Pearson chi2(4) = 0.6881 Pr = 0.953, gamma = –0.0625 ASE = 0.364

Source: Authors' calculations.

Table 2C. College Attendance Impacts Versus High School Graduation Impacts: Number of Estimates by Sign and Statistical Significance

		College Attendace Impacts				
		Negative Significant	**Negative Insignificant**	**Positive Insignificant**	**Positive Significant**	**Total**
High School Graduation Impacts	Negative Significant	0	0	0	0	0
	Negative Insignificant	0	0	2	0	2
	Positive Insignificant	0	3	2	1	6
	Positive Significant	0	1	2	8	10
	Total	0	3	6	9	18

Pearson chi2(4) = 13.3333 Pr = 0.010, gamma = 0.7778 ASE = 0.112

Source: Authors' calculations.

Table 3A. Impacts on Earning a Four-Year College Degree Versus Impacts on High School ELA: Number of Estimates by Sign and Statistical Significance

		Four-Year College Degree Impacts				
		Negative Significant	Negative Insignificant	Positive Insignificant	Positive Significant	Total
ELA Impacts	Negative Significant	0	0	0	0	0
	Negative Insignificant	0	2	3	0	5
	Positive Insignificant	0	1	2	0	3
	Positive Significant	0	0	0	3	3
	Total	0	3	5	3	11

Pearson chi2(4) = 11.0489 Pr = 0.026, gamma = 0.8065 ASE = 0.201

Source: Authors' calculations.

Table 3B. Impacts on Earning a Four-Year College Degree Versus Impacts on High School Math: Number of Estimates by Sign and Statistical Significance

		Four-Year College Degree Impacts				
		Negative Significant	Negative Insignificant	Positive Insignificant	Positive Significant	Total
Math Impacts	Negative Significant	0	0	0	0	0
	Negative Insignificant	0	1	2	1	3
	Positive Insignificant	0	2	3	1	6
	Positive Significant	0	1	0	1	2
	Total	0	3	5	3	11

Pearson chi2(4) = 3.1778 Pr = 0.529, gamma = –0.2800 ASE = 0.456

Source: Authors' calculations.

Table 3C. Impacts on Earning a Four-Year College Degree Versus Impacts on High School Graduation: Number of Estimates by Sign and Statistical Significance

		Four-Year College Degree Impacts				
		Negative Significant	Negative Insignificant	Positive Insignificant	Positive Significant	Total
High School Graduation Impacts	Negative Significant	0	0	0	0	0
	Negative Insignificant	0	0	0	0	0
	Positive Insignificant	0	2	3	0	5
	Positive Significant	0	1	1	3	5
	Total	0	3	4	3	10

Pearson chi2(2) = 4.3333 Pr = 0.115, gamma = 0.7000 ASE = 0.308

Source: Authors' calculations.

college attainment statistics are measured far more accurately. Random noise in measurement causes impact estimates to attenuate. The greater noise with which an outcome is measured, the more likely that estimated impacts on that outcome will be insignificantly different from zero.

However, measurement error cannot entirely explain the difference in trends across outcomes. For example, there are multiple instances in which school choice programs produced significant test score gains but no significant gains in high school graduation. Testing error can perhaps explain why estimated impacts would be larger for attainment than for achievement—but statistical noise can in no way explain why impacts would ever be larger for achievement than attainment.

This begs a larger question that we seek to examine. Test score impacts may not perfectly mirror attainment impacts, but it might be possible that a program's *relative* impact on achievement might mirror its *relative* impact on attainment. That is to say, if programs were separately ranked by their impacts on achievement and their impacts on attainment, perhaps the rank order of achievement effects would be correlated with the rank order of attainment effects. If this were the case, then achievement impacts could still serve as a useful indicator of a program's relative impact on longer-term outcomes, which in turn would support the strong use of achievement impacts as an accountability tool.

Tables 1, 2, and 3 show the cross tabs of the ordinal categories of study findings. If findings were perfectly and positively correlated across studies, one would expect to only see values in the cells running diagonal from the upper left to the lower right. That is not the pattern of actual findings.

In direction and significance, high school graduation impacts match ELA impacts in only 13 of 34 cases and math impacts in 9 of 33 cases. Impacts on college enrollment nominally match impacts on ELA in 9 of 19 cases and math in 7 of 18 cases. Granted, some values lie barely off the diagonal in these instances. Treating the values in Tables 1, 2, and 3 as ordinal, one could use Pearson's chi-squared test or Goodman and Kruskall's gamma to test whether findings were associated across studies. Treating them as nominal, one could use Goodman and Kruskall's lambda. Each approach produces the same results. Across the studies we examine, there is no significant or meaningful association between school choice impacts on math scores and high school graduation or college attendance. Nor are ELA impacts meaningfully associated with high school graduation rates. Under some tests, the relationship between ELA impacts and college attendance are significant—but the relationship is weak in magnitude, and the sample of studies is far narrower for college attainment than for high school graduation.

We found only 11 evaluations that examine the impacts on both achievement and completion of a four-year college degree. There is no apparent association between math impacts and impacts on the completion of a four-year college degree. The nominal impacts on math and college graduation match in 4 of 11 instances. The only exception to the overall pattern of results is the association between ELA impacts and college graduation, in which conclusions on the nominal impacts of the programs match in 6 of 11 instances. However, due to the small number of studies examined, any statistical test of association is extremely sensitive to the inclusion of even one additional study. This is an instance in which, literally, every additional study has the ability to affect how we understand this issue.

Overall, we have asked a simple question: Do test score impacts of school choice programs serve as a reliable predictor of attainment impacts? Across the existing literature, the answer is no. This pattern of findings may change as more studies are published on school choice. But if the pattern remains consistent, the implications for school choice policy are massive.

Conclusion and Policy Recommendation

Under current K–12 regulatory regimes, the growth of school choice is actively managed. Take the example of charter schools, where caps are commonly placed on the number of charters that can open. Because only a limited number of charter schools can operate at a given time, the authorities who grant charter

schools approval to open tend to focus on replicating schools with a demonstrated record of success. If that record of success is judged on test scores, authorities privilege schools that produce test score gains.

This "portfolio model" of regulation has grown substantially in popularity. The concept borrows its name from investment banking: Managers should begin with a diverse array of investments and thereafter transfer resources toward the assets producing the best results while dumping assets with poor returns. This method, or philosophy, of governance draws heavily on the principles of program evaluation—using social scientific methods to determine whether a program is producing benefits.

Program evaluation follows a simple logic model. Define the outcomes that a program is supposed to impact. Assess the extent to which the program has affected those outcomes. If the program produces positive impacts at an acceptable cost, recommend it for expansion or replication. If a program fails to do so, recommend it for reformation or elimination.

So the portfolio model of school choice governance works as follows. Allow new schools to open. Evaluate their effectiveness. Identify schools that produce gains. Select those schools for replication and expansion. Close the schools that fail to produce gains. The process is intended, over time, to increase the overall quality of schools in a portfolio. This model of education governance is highly attractive, but its success depends first and foremost on one key factor: the appropriateness of the metric used to judge the success of the assets in the portfolio. This takes us back to the basics of program evaluation—the outcomes that researchers choose to focus on.

The outcome measure that is chosen by researchers and policymakers is a fundamental part of the evaluation process.[55] Ambitious programs such as school choice have ambitious goals: to leave children better off in the long run. But long-run outcomes naturally take a long time to observe. So program evaluators often chose to focus on short-term outcomes, at least in the early years of a program. If there is a disconnect between effects on test scores and later-life outcomes, particularly when evaluating schools of choice, the regulatory regime might need to be rethought.

Our meta-analysis shows that, at least for school choice programs, there is a weak relationship between impacts on test scores and later-life outcomes. This finding could have several possible explanations, but regardless of the cause, it has serious implications for school choice policy.

In the previous section we considered and ruled out measurement error as the likely cause of the test score–attainment disconnect. The more likely explanation, to us at least, is that test scores and measures of educational attainment capture different constructs that at the margin are only weakly related to each other. Test scores might measure changes in knowledge or intelligence. High school graduation, college attendance, or earnings later in life could capture habits such as perseverance and conscientiousness. Some schools might be better at promoting such noncognitive skills than others.

Test scores should be put in context and should not automatically occupy a privileged place over parental demand and satisfaction as short-term measures of school choice success or failure.

Now, perhaps these attainment measures do not matter as much as we might think. If schools have watered down graduation requirements so that students who should not be receiving diplomas are receiving them or if unprepared students are matriculating into college only to fail, perhaps even these

measures are imperfect looks into what we really care about. There is a robust literature on the impact of high school graduation on later-life outcomes,[56] but recent policy changes incentivizing increased graduation rates and college attendance might have watered down standards and weakened that relationship.

Even with these caveats in mind, the policy implications from this analysis are clear. The most obvious implication is that policymakers need to be much more humble in what they believe that test scores tell them about the performance of schools of choice. Test scores are not giving us the whole picture. Insofar as test scores are used to make determinations in "portfolio" governance structures or are used to close (or expand) schools, policymakers might be making errors. This is not to say that test scores should be wholly discarded. Rather, test scores should be put in context and should not automatically occupy a privileged place over parental demand and satisfaction as short-term measures of school choice success or failure.

We are not arguing that policymakers should give up trying to evaluate how well school choice programs or the students or schools that participate in them are performing. We are simply saying that, if test scores are their only barometer, policymakers might be receiving imperfect information. As it becomes easier to track later-life outcomes by linking student information to census or tax records, those looking to evaluate school choice policies have the opportunity to make a more holistic evaluation of the meaningful effects of these programs.

We are still a long way from fully understanding how schooling affects children. Our analysis points us in a direction of diversifying the data sources that we use to evaluate schools and school programs and investigating new metrics that might be better correlated with the types of life outcomes that we value.

The unprecedented growth in school choice begs two related questions. Are school choice programs improving student outcomes? And what is the best way for policymakers to manage the growth of school choice to maximize the benefits to students?

For both of these questions, researchers and policymakers have looked mainly at standardized tests to provide the answers. Our findings cast doubt on that approach.

About the Authors

Collin Hitt is an assistant professor of medical education and a research director of Continuing Professional Development at the Southern Illinois University School of Medicine. His research covers a wide array of educational topics including character skills, medical education, survey methods, school choice, and extracurricular activities.

Michael Q. McShane is director of National Research at EdChoice, a research fellow in the Economic and Policy Analysis Research Center at the University of Missouri, and an adjunct scholar in education policy studies at AEI. He is the author, editor, or coeditor of seven books on education reform.

Patrick J. Wolf is Distinguished Professor of Education Policy and 21st Century Endowed Chair in School Choice in the Department of Education Reform at the University of Arkansas. He has authored or coauthored four books and more than 140 journal articles, book chapters, and research reports primarily on school choice.

Notes

1. William G. Howell et al., "School Vouchers and Academic Performance: Results from Three Randomized Field Trials," *Journal of Policy Analysis and Management* 21, no. 2 (2002): 191–217; Joshua M. Cowen et al., "School Vouchers and Student Attainment: Evidence from a State-Mandated Study of the Milwaukee Parental Choice Program," *Policy Studies Journal* 41, no. 1 (2013): 147–67; John F. Witte et al., "High Stakes Choice: Achievement and Accountability in the Nation's Oldest Urban Voucher Program," *Education Evaluation and Policy Analysis* 36, no. 4 (2014): 437–56; and Jonathan N. Mills and Patrick J. Wolf, "Vouchers in the Bayou: The Effects of the Louisiana Scholarship Program on Student Achievement After Two Years," *Education Evaluation and Policy Analysis* 20, no. 10 (2017): 1–21.

2. Raj Chetty, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review* 104, no. 9 (2014).

3. M. Danish Shakeel, Kaitlin P. Anderson, and Patrick J. Wolf, "The Participant Effects of Private School Vouchers Across the Globe: A Meta-Analytic and Systematic Review" (working paper, Economics Research Network, Department of Education Reform, College of Education and Health Professions, University of Arkansas, Fayetteville, Arkansas, 2016).

4. Julian Betts and Y. Emily Tang, "A Meta-Analysis of the Literature on the Effect of Charter Schools on Student Achievement," Center on Reinventing Public Education, August 2014.

5. Patrick J. Wolf et al., "School Vouchers and Student Outcomes: Experimental Evidence from Washington, DC," *Journal of Policy Analysis and Management* 32, no. 2 (2013): 246–70; and Ruth Curran Neild, Christopher Boccanfuso, and Vaughan Byrnes, "The Academic Impacts of Career and Technical Schools: A Case Study of a Large Urban School District," Everyone Graduates Center, January 2013.

6. Joshua D. Angrist et al., "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics* 34, no. 2 (2016): 275–318.

7. Diane Whitmore Schanzenbach and Lauren Bauer, "The Long-Term Impact of the Head Start Program," Brookings Institution, August 2016.

8. Seth Gershenson, "Linking Teacher Quality, Student Attendance, and Student Achievement," *Education Finance and Policy* 11, no. 2 (2016); and C. Kirabo Jackson, "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from Ninth Grade Teachers in North Carolina," National Bureau for Economic Research, December 2012.

9. David Blazar and Matthew Kraft, "Teacher and Teaching Effects on Students' Academic Behaviors and Mindsets," Mathematica Policy Research, 2015; and Matthew A. Kraft and Sarah Grace, "Teaching for Tomorrow's Economy? Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies," Brown University, 2016.

10. Albert Cheng and Gema Zamarro, "Measuring Teacher Non-Cognitive Skill and Its Impact on Students: Insight from the Measures of Effective Teaching Longitudinal Database," *Economics of Education Review* (forthcoming); and Albert Cheng, "Like Teacher, Like Student: Teachers and the Development of Student Noncognitive Skills," University of Arkansas, College Department of Education Reform, 2015.

11. Wolf, "School Vouchers and Student Outcomes."

12. Angrist et al., "Stand and Deliver."

13. Mark W. Lipsey et al., "Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms," National Center for Special Education Research, November 2012.

14. Betts and Tang, "A Meta-Analysis of the Literature on the Effect of Charter Schools on Student Achievement"; and Shakeel, Anderson, and Wolf, "The Participant Effects of Private School Vouchers Across the Globe."

15. Chetty, Friedman, and Rockoff, "Measuring the Impacts of Teachers II."

16. Frederick F. Mosteller and Robert F. Boruch, *Evidence Matters: Randomized Trials in Education Research* (Washington, DC: Brookings Institution, 2002); and Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach* (Thousand Oak, CA: Sage Publications, 2004).

17. Mills and Wolf, "Vouchers in the Bayou"; Angrist et al., "Stand and Deliver"; and Howell et al., "School Vouchers and Academic Performance."

18. Joshua M. Cowen, "School Choice as a Latent Variable: Estimating the 'Complier Average Causal Effect' of Vouchers in Charlotte," *Policy Studies Journal* 36 (2008): 301–15.

19. P. Schochet et al., "Standards for Regression Discontinuity Designs: Version 1.0 (Pilot)," What Works Clearinghouse, June 2010.

20. Michael P. Murray, "Avoiding Invalid Instruments and Coping with Weak Instruments," *Journal of Economic Perspectives* 20, no. 4 (2006): 111–32.

21. Julie Berry Cullen, Brian A. Jacob, and Steven Levitt, "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica* 74, no. 5 (2006): 1191–230; and Peter Bergman, "The Effects of School Integration: Evidence from a Randomized Desegregation Program," Columbia University, 2016.

22. Daniel P. Mayer et al., "School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program," Mathematica Policy Research, February 19, 2002; Alan B. Krueger and Pei Zhu, "Another Look at the New York City School Voucher Experiment," *American Behavioral Scientist* 47 (2004): 658–98; Paul E. Peterson and William G. Howell, "Efficiency, Bias, and Classification Schemes: A Response to Alan B. Krueger and Pei Zhu," *American Behavioral Scientist* 47, no. 5 (2004): 699–717; and Marianne Bitler et al., "Distributional Effects of a School Voucher Program: Evidence from New York City," *Journal of Research on Education Effectiveness* 8 (2015): 419–50.

23. Matthew M. Chingos and Paul E. Peterson, "Experimentally Estimated Impacts of School Vouchers on College Enrollment and Degree Attainment," *Journal of Public Economics* 122 (2015): 1–12.

24. John F. Witte et al., "The MPCP Longitudinal Educational Growth Study Second Year Report," University of Arkansas, Department of Education Reform, School Choice Demonstration Project, 2009; and Joshua M. Cowen et al., "Student Attainment and the Milwaukee Parental Choice Program: Final Follow-Up Analysis," School Choice Demonstration Project, University of Arkansas, 2012.

25. Wolf et al., "School Vouchers and Student Outcomes."

26. Cullen, Jacob, and Levitt, "The Effect of School Choice on Participants."

27. David J. Deming et al., "School Choice, School Quality, and Postsecondary Attainment," *American Economic Review* 104, no. 3 (2014): 991–1013.

28. Caroline M. Hoxby, Jenny Lee Kang, and Sonali Murarka, "Technical Report: How New York City's Charter Schools Affect Achievement," National Bureau of Economic Research, September 2009.

29. Will Dobbie and Roland G. Fryer Jr., "The Medium-Term Impacts of High-Achieving Charter Schools," *Journal of Political Economy* 123, no. 5 (2015): 985–1037.

30. Angrist et al., "Stand and Deliver."

31. Ron Zimmer et al., "Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition," Rand Corporation, 2009; and Tim R. Sass et al., "Charter High Schools' Effects on Long-Term Attainment and Earnings," *Journal of Policy Analysis and Management* 35, no. 3 (2016): 683–706.

32. Zimmer et al., "Charter Schools in Eight States"; and Sass et al., "Charter High Schools' Effects on Long-Term Attainment and Earnings."

33. Rebecca Unterman et al., "Going Away to School: An Evaluation of SEED DC," Social Science Research Network, October 13, 2016.

34. Mitchell D. Wong et al., "Successful Schools and Risky Behaviors Among Low-Income Adolescents," *Pediatrics* 134, no. 2 (2014): e389-e396.

35. Viki Young et al., "Evaluation of the Texas High School Project: Third Comprehensive Annual Report," SRI International, 2010.

36. Will S. Dobbie and Roland G. Fryer Jr., "Charter Schools and Labor Market Outcomes," National Bureau of Economic Research, 2016.

37. Christina Clark Tuttle et al., "KIPP Middle Schools: Impacts on Achievement and Other Outcomes," Mathematica Policy Research, February 2013.

38. Christina Clark Tuttle et al., "Understanding the Effect of KIPP as It Scales: Volume I, Impacts on Achievement and Other

Outcomes," Mathematica Policy Research, September 17, 2015.

39. Joshua Furgeson et al., "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts," Mathematica Policy Research, 2012.

40. Andrea Berger et al., "Early College, Early Success: Early College High School Initiative Impact Study," American Institutes for Research, 2013; and Michael Garet, Joel Knudson, and Gur Hoshen, "Early College, Continued Success: Early College High School Initiative Impact Study," American Institutes for Research, 2014.

41. Young et al., "Evaluation of the Texas High School Project."

42. Julie A. Edmunds et al., "Smoothing the Transition to Postsecondary Education: The Impact of the Early College Model," *Journal of Research on Educational Effectiveness* 10, no. 2 (2017): 297–325.

43. Douglas L. Lauen et al., "Early Colleges at Scale: Impacts on Secondary and Postsecondary Outcomes," *American Journal of Education* 123, no. 4 (2017): 523–51.

44. Atila Abdulkadiroğlu, Joshua D. Angrist, and Parag Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica* 82, no. 1 (2014): 137–96.

45. Lisa Barrow, Lauren Sartain, and Marisa de la Torre, "The Role of Selective High Schools in Equalizing Educational Outcomes: Heterogeneous Effects by Neighborhood Socioeconomic Status," University of Chicago Consortium on School Research, 2017.

46. Will Dobbie and Roland G. Fryer Jr., "Exam High Schools and Academic Achievement: Evidence from New York City," National Bureau of Economic Research, 2011; and Will Dobbie and Roland G. Fryer Jr., "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools," *American Economic Journal: Applied Economics* 6, no. 3 (2014): 58–75.

47. James J. Kemple and Jason C. Snipes, "Career Academies: Impacts on Students' Engagement and Performance in High School," Manpower Demonstration Research Corporation, 2000; James J. Kemple, "Career Academies: Impacts on Students' Initial Transitions to Post-Secondary Education and Employment," Manpower Demonstration Research Corporation, 2001; and James J. Kemple and Cynthia J. Willner, "Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood," Manpower Demonstration Research Corporation, 2008.

48. Steven W. Hemelt, Matthew A. Lenard, and Colleen G. Paeplow, "Building Better Bridges to Life After High School: Experimental Evidence on Contemporary Career Academies," National Center for Analysis of Longitudinal Data in Education Research, 2017.

49. Neild, Boccanfuso, and Byrnes, "The Academic Impacts of Career and Technical Schools"; and Ruth Curran Neild and Vaughan Byrnes, "Impacts of Career and Technical Schools on Postsecondary Outcomes," Everyone Graduates Center, 2014.

50. Shaun M. Dougherty, "The Effect of Career and Technical Education on Human Capital Accumulation: Causal Evidence from Massachusetts," *Education Finance and Policy*, forthcoming.

51. Young et al., "Evaluation of the Texas High School Project."

52. Howard S. Bloom and Rebecca Unterman, "Sustained Positive Effects on Graduation Rates Produced by New York City's Small Public High Schools of Choice," Manpower Demonstration Research Corporation, 2012; Howard S. Bloom and Rebecca Unterman, "Sustained Progress: New Findings About the Effectiveness and Operation of Small Public High Schools of Choice in New York City," Manpower Demonstration Research Corporation, 2013; and Atila Abdulkadiroğlu, Weiwei Hu, and Parag A. Pathak, "Small High Schools and Student Achievement: Lottery-Based Evidence from New York City," National Bureau of Economic Research, 2013.

53. Lisa Barrow, Diane Whitmore Schanzenbach, and Amy Claessens, "The Impact of Chicago's Small High School Initiative," *Journal of Urban Economics* 87 (2015): 100–13.

54. Wolf et al., "School Vouchers and Student Outcomes."

55. Harry P. Hatry, *Performance Measurement: Getting Results* (Washington, DC: Urban Institute Press, 2006).

56. Summarized in Patrick J. Wolf and Michael Q. McShane, "Is the Juice Worth the Squeeze? A Benefit/Cost Analysis of the District of Columbia Opportunity Scholarship Program," *Education Finance and Policy* 8, no. 1 (Winter 2013): 74–99.