Exploring the Achievement-Attainment Disconnect in the Effects of School Choice Programs

Patrick J. Wolf
University of Arkansas

Collin Hitt
Southern Illinois University School of Medicine

Michael Q. McShane
EdChoice

Draft paper prepared for presentation at the conference "Learning from the Long-Term Effects of School Choice in America," Program on Education Policy and Governance, Kennedy School of Government, Harvard University, Cambridge, MA, April 19, 2018. An earlier draft of this paper appears as an AEI white paper at https://www.aei.org/publication/do-impacts-on-test-scores-even-matter-lessons-from-long-run-outcomes-in-school-choice-research/. We are grateful to the following individuals for useful comments on this project: Bradford White and participants in the Illinois Education Research Council's 2016 annual research symposium; Jay Greene and attendees of a presentation of this work to the University of Arkansas Department of Education Reform; and Paul Peterson, Marty West and Albert Cheng at the Harvard Program on Education Policy and Governance. We own any remaining errors.

Abstract

For the past 20 years, almost every major education reform has rested on a common assumption: Standardized test scores are an accurate and appropriate measure of program success and failure. School choice programs have been capped and specific schools of choice have been closed down when they have failed to generate clear positive achievement gains for their students. Recently, some school choice researchers have determined the effects of choice on longer-term educational attainment outcomes such as high school graduation, college enrollment and college graduation in addition to studying test score effects. In a surprising number of cases, the attainment effects reported by these scholars have not matched the achievement effects from the same program or set of choice schools. This study is a vote-counting meta-analysis on the extent to which the achievement effects from school choice evaluations, which usually are reported at the early stages of the evaluation, accurately predict the latter attainment effects from the same program. Across 27 different studies that produce 126 achievement-attainment results pairings, we find only a weak relationship between impacts on test scores and later attainment outcomes. Among the six major sets of associations tested, only achievement effects in English Language Arts (ELA) significant predict future attainment results, and then only for the outcome of college graduation. The ELA results from school choice evaluations do not reliably predict the high school graduation or college enrollment effects of the same programs. The achievement effects of school programs in math actually tend to be negative predictors of their subsequent attainment effects. Due to this empirical divide between achievement and attainment findings, we think 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 should not necessarily occupy a privileged place over parental demand and satisfaction as short-term measures of school choice success or failure, since test scores only occasionally predict attainment outcomes that matter greatly to students and their parents.

Exploring the Achievement-Attainment Disconnect in the Effects of School Choice Programs

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 (e.g. Lipsey 2012). 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 also has seen explosive growth in school choice programs, including inter-district and intra-district choice within the traditional public school system, public charter schools, and various programs to promote private school choice. These choice-based reforms have been evaluated largely based on their impacts on student test scores (Betts & Tang 2014; Shakeel, Anderson & Wolf 2016). Reading and math tests measure basic skills that almost everyone believes are important. Test scores are convenient to collect. They measure a slice of student knowledge with known levels of reliability. Yet even the most fervent believer in the power of standardized tests agrees that test scores are merely an interim measure or "intermediate outcome" (Hatry 2006). There is no point in increasing test scores for their own sake. Increased test scores are supposed to indicate progress toward more important long-term outcomes.

Perhaps the most convincing evidence of this supposed truth comes from seminal work that found connections between changes in individual-level student test scores and the likelihood that they would graduate from high school, avoid having children as teenagers, and earn more money in their late 20s (Chetty, Friedman & Rockoff 2014).

Other education research, especially involving evaluations of schools of choice, sows doubt surrounding 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 initial gains in student test scores (e.g. Wolf et al. 2013; Neild, Boccanfuso & Byrnes 2013). Conversely, studies of some school choice programs have found large short-term test score gains but no lasting benefits in terms of high school graduation rates or college attainment (e.g. Angrist et al. 2016). Improving test scores appears to be neither a necessary nor a sufficient condition for improving the later-life outcomes that truly matter.

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 reported to have different impacts on achievement and attainment.

We use an expansive definition of school choice, including private school voucher programs, public charter schools, and district-administered forms of public school choice such as open enrollment, early college high schools, magnet schools, and vocational schools. We define school choice expansively for two reasons. First, we wanted to gather the largest number of studies possible to generate greater statistical reliability in our analysis. Second, parental school choice is bigger than just private school voucher programs and public 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.

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. Our findings are based on 126 pairs of specific test score effects matched to specific attainment effects regarding more than 20 different school choice programs. 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. Our question is, across all findings and for policy-relevant subgroups of findings, do program impacts on test scores predict impacts on later outcomes?

We find that school choice program impacts just 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 (Shakeel, Anderson & Wolf 2016) and charters (Betts & Tang 2014). However, school choice impacts on attainment are much more consistently positive (Foreman 2017). This pattern itself implies that some programs have produced larger attainment impacts than achievement impacts. The pattern of findings, however, is actually more complicated. A school choice program's impact on test scores only significantly predicts its impacts on longer-term outcomes when the test score effects are in English Language Arts (ELA) and the longer-term outcome is graduation from a four-year college. The evidence that supports that one specific achievement-attainment connection comes from 11 pairs of findings, 10 of which come from evaluations of district-administered public school choice programs. The ELA findings generally are not predictive of high school graduation or college enrollment impacts. Shockingly, the math impacts of school choice

programs tend to negatively predict the attainment effects of those same programs, meaning that if the test score effects are positive we are more likely to observe negative program effects on attainment, and vice-versa.

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 often fade-out over time and sometimes are negatively associated with child behavior measures (e.g. Puma et al. 2012). Some of the preschool programs that have produced the most impressive improvements in later-life outcomes did so without producing lasting gains on test scores (Schanzanbach & Bauer 2016).

Studies of teacher impacts on student outcomes show a similar pattern of results. As with school choice, teacher impacts on test scores are used for policy purposes. In pilot projects and even entire districts, teachers who produce out-sized achievement gains, statistically estimated to be true "value added" by the teacher, are rewarded and promoted and those who do not are remediated or fired. As with school choice, those same methods are now being used to assess teacher impacts on other, noncognitive outcomes. It turns out that, in at least some studies, teacher impacts on test scores are almost entirely uncorrelated with teacher impacts on student classroom behavior, attendance, truancy, and grades (Jackson 2012; Gershenson 2016). Other studies find that teacher impacts on test scores are uncorrelated with teacher effects on self-reported noncognitive skills such as grit (Blazer & Kraft 2015; Kraft & Grace 2016). Teachers who possess higher noncognitive skills boost the noncognitive skills of their students but not student test scores (Cheng 2015; Cheng & Zamarro 2018). 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. School choice programs and specific types of schools of choice appear to specialize in producing either achievement gains or attainment benefits but rarely both.

Thus, 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 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, especially in math, may lead authorities to favor choice programs and schools whose benefits could easily fade over time and punish programs and schools that are poised to produce long-lasting gains.

In the following section we briefly describe our search strategy for gathering studies and 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, public charter schools, and various specific forms of district-administered public school choice. We then present our results in the form of aggregate findings, results broken out by type of school choice, and robustness checks. Finally, we discuss the important implications of our findings for current policy and future research.

Methods

Details regarding our literature search, quality screen, coding and analysis are provided in Appendix A. Here we briefly summarize those elements of our study.

The goal of our systematic literature search was to identify studies that examined the effect of school choice on the achievement outcomes of reading and math scores and the

attainment outcomes of high school graduation, college attendance, and completion of a four-year college degree. We limited our search to studies set in the United States.

Our search was a hybrid of systematic and networked search techniques. We searched Google Scholar using combinations of specific school choice and educational keywords. 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. Ultimately, a study (or a series of studies) was included only if it provided both achievement and attainment impacts of a school choice program or defined set of choice schools.

Studies were judged to be sufficiently rigorous if they employed quantitative analysis and either an experimental or quasi-experimental research design. Experimental designs randomly assigned students to treatment and control groups using lotteries. Quasi-experimental designs included Regression Discontinuity Design, Instrumental Variables, Propensity Score Matching, or statistical modeling using control variables. We defined study rigor generously in terms of overall research design, even including studies that merely added control variables to observational data, because we wanted to cast our search net broadly.

Any study not eliminated during the abstract review was given a full reading, whereupon we determined the research methods used and the outcomes of interest. 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.

The studies selected for inclusion based on our systematic search then formed the basis for our networked search. 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. In each instance that a new study was identified through the networked search, the entire network search process was repeated until no new studies were identified.

In total, we included information from 24 school choice evaluation projects (Table 1). Thirty-three distinct "studies" composed those 24 projects, as several of the projects reported achievement results in one study but attainment results in another. We linked multiple studies within a defined project if the data and project described in the studies matched up conclusively. In most if these cases, the authorship teams for the studies overlapped. The 24 projects produced 36 unique sets of estimates of the impact that school choice programs have had on achievement and attainment, since some projects examined multiple cohorts or groups. Those 36 sets of estimates generated 126 distinct achievement-attainment matched pairs, as indicated in the far-right column of Table 1, since all of the projects measured effects on both ELA and math and some of them also included multiple attainment measures. Two studies that used a composite measure of achievement instead of separate ELA and math scores (Barrow, Sartain & Torre 2016; Dougherty 2016) were excluded from this analysis. As we will see, the connection between achievement and attainment in school choice studies varies markedly depending on whether the achievement measure is ELA or math. Thus, composite achievement measures appear to be an improper mixture of oil and water for purposes of our study.

Table 1. Studies that Inform the Meta-Analysis

Authors	Year	Setting	Choice Type	Method	Achieve. Outcomes	Attain. Outcomes	Matched Pairs
Berger et al.	2013/14	National	ECHS	RCT	ELA, math	HS grad, col enroll, col grad	6
Kemple et al.	2000/01/08	National	CTE	RCT	ELA, math	HS grad, col enroll, col grad	6
Neild et al. (3 cohorts)	2013/14	Philadelphia	CTE	RCT	ELA, math	HS grad (3), col enroll (3), col grad (3)	18
Barrow et al.	2016	Chicago	Magnet	RDD	Composite	HS grad, col enroll	2
Abdulkadiroglu et al.	2014	Boston	Magnet	RDD	ELA, math	col enroll	2
Dobbie & Fryer (3 schools)	2014	New York	SEHS	RDD	ELA, math	HS grad (3), col enroll (3), col grad (3)	18
Bloom & Unterman, Abdulkadiriglu et al.	2012/13 2013	New York	SSC	RCT	ELA, math	HS grad, col enroll	4
Barrow et al.	2015	Chicago	SSC	IV	ELA, math	HS grad	2
Deming et al.	2014	Charlotte	OE	RCT	ELA, math	HS grad, col enroll, col grad	6
Cullen et al.	2006	Chicago	OE	RCT	ELA, math	HS grad	2
Young et al.	2010	Texas	STEM, ECHS, Charter	Match	ELA, math	HS grad (3)	6
Zimmer et al, Sass et al	2009/16	Chicago, Florida	Charter	Match	ELA, math	HS grad (2), col enroll (2)	8
Angrist et al	2016	Boston	Charter	RCT	ELA, math	HS grad, col enroll	4
Tuttle et al (2 cohorts)	2015	KIPP National	Charter	Match	ELA, math	HS grad (2)	4
Dobbie & Fryer	forthcoming	Harlem	Charter	RCT	ELA, math	HS grad, col enroll	4
Hoxby et al	2009	New York	Charter	RCT	ELA, math	HS grad	2
Unterman et al	2016	DC	Charter	RCT	ELA, math	HS grad	2
Wong et al	2014	Los Angeles	Charter	RCT	ELA, math	HS grad	2
Witte et al, Cowen et al	2014/13	Milwaukee	Voucher	Match	ELA, math	HS grad	2
Wolf et al	2013	DC	Voucher	RCT	ELA, math	HS grad	2
Mayer et al, Chingos & Peterson	2002/15	New York	Voucher	RCT	ELA, math	col enroll, col grad	4
Dobbie & Fryer (2 groups)	2017	Texas	Charter	Match	ELA, math	HS grad (2)	4
Ferguson et al (3 groups)	2012	CMO-National	Charter	RCT	ELA (3), math (2)	HS grad (3), col enroll (1)	6
Hemelt et al	2017	North Carolina	CTE	RCT	ELA, math	HS grad, col enroll	4
Edmunds et al	2017	North Carolina	ECHS	RCT	ELA, math	HS grad, col enroll, col grad	6

Notes: New York is New York City, ECHS is Early College High Schools, CTE is Career/Technical Enterprises, SSC is School Schools of Choice, RCT is Randomized Control Trial, RDD is Regression Discontinuity Design, IV is Instrumental Variables, HS grad is high school graduation, col enroll is college enrollment, col grad is college graduation.

Our goal is to compare the effects on reading and math scores of school choice policies to their effects on high school graduation, college attendance, and completion of a four-year college degree. 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. When multiple results are presented for a specific outcome – English Language Arts, math, high school graduation, college enrollment, or college graduation– we code the result from the authors' preferred statistical model or the most sophisticated model if the author does not signal a preference. We accept the authors' definition of the minimum level of statistical significance to conclude that the program effect was not due to chance. For most of our studies, that level is 95% confidence (a P-Value less than .05) but for some of them it is 90% confidence (a P-Value less than .10).

Our coding scheme may seem rigid. One can criticize the exactitude that we are imposing on the data, as achievement and attainment results must match regarding both direction and statistical significance. We believe this exactitude is justified by the fact that the conclusions that many policymakers and commentators draw about whether school choice "works" depends on the direction and significance of the effect parameter. As social scientists, we would prefer that practitioners rigidly adhere to the convention of treating all non-significant findings as null or essentially 0, regardless of whether they are positive or negative in their direction, but we also eschew utopianism. Outside of the ivory tower, people treat a negative school choice effect as a

bad result and a positive effect as a good result, regardless of whether it reaches scientific standards of statistical significance. Since one of our goals is to evaluate the prudence of those judgments, we adopt the same posture for our main analysis. As a robustness check, we collapse all positive and negative non-significant findings into a single middle category of null findings and re-compute our overall analysis. Our results do not change.

To analyze the extent to which the intermediate outcomes of ELA and math effects predict the end outcomes of our three attainment measures, we tabulate the paired results, creating a frequency matrix. The achievement results are arrayed on the vertical axis (left side), from statistically significant positive results in the top cell to statistically significant negative results in the bottom cell, with the non-significant positive and non-significant negative results in the two cells between those poles. The attainment results are arrayed on the horizontal axis (across the top), from statistically significant positive results in the far left cell to statistically significant negative results in the far right cell, with the non-significant positive and non-significant negative results in the two cells between those poles.

Each of the paired achievement-attainment findings is counted in the cell of this 4 x 4 (16-cell) matrix that corresponds with the direction and significance of its achievement result (vertical axis) and the direction and significance of its attainment result (horizontal axis). If every achievement effect from school choice studies (or a category of studies) perfectly predicted its attainment effects, then all of the observations would appear in the primary diagonal of the matrix, running from the upper left to the lower right, called the "trace." Counts of observations in cells outside of the trace represent errors in the achievement effects predicting the attainment effects. The cells above and to the right of the trace compose the "upper triangle." Observations in those cells represent cases where the attainment results of school choice are

more positive than the achievement results. The cells below and to the left of the trace compose the "lower triangle." Observations in those cells represent cases where the attainment results of school choice are more negative than the achievement results.

We use a set of measures of association, customized to fit our data, in order to assess the performance of the achievement results in predicting the attainment results, and thus clustering on the trace. First, we present the proportion of paired achievement-attainment results that are perfect matches and therefore fall along the trace, labeled "Trace (%)" in the tables. If all achievement results perfectly predict subsequent attainment results, then the Trace (%) will be 100. We might expect that it will be at least 50 under most circumstances, meaning that a majority of the achievement findings accurately predict their attainment results. Second, we present Pearson X^2 as a measure of the extent to which the cell frequencies of the matrix are patterned, with the resulting P-Value providing the specific likelihood, from .00 to 1.00, that any pattern is due to chance. Smaller P-Values indicate less of a chance that the association between the achievement and attainment results in the table is merely random. Third, we provide the value of Gamma, specifically Goodman & Kruskel's gamma, which is a measure of association between ordinal variables (Vogt 1993). Gamma is positively or negatively signed, telling us if the achievement and attainment effects tend to move in the same direction, yielding a positive sign, or in opposite directions, yielding a negative sign. Asymptotic Standard Errors approximate the error in the association captured by the Gamma statistic as the sample N, which also is provided in the tables, approaches infinity. Shaded rows in the tables highlight associations between achievement and attainment with statistically significant P-Values and positive, non-trivially sized Gammas, as such findings signal that achievement effects in a specific direction (positive or negative) tend to beget attainment effects in the same direction and

it is unlikely that the relationship is due to mere chance. If the achievement effects of school choice evaluations tend to be reliable predictors of the attainment effects from those same studies, we should see a lot of shaded rows in the tables that follow. Spoiler alert: We don't.

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 of school choice that examine both outcomes. Because we exclude the small number of choice studies that only examine attainment impacts and the much larger number of choice studies that examine only test score impacts, our results should not be interpreted as assessing the overall effect that school choice has had on achievement test scores or attainment. Other meta-analyses or systematic reviews focused specifically on those questions are available in the existing school choice literature (e.g. Foreman 2017; Shakeel, Anderson & Wolf 2016; Betts & Tang 2014).

Overall descriptive results

In total, we identify 126 pairs of estimates of impacts on English Language Arts (ELA) and math tests from studies that also examine impacts on high school graduation, college enrollment (either 2- or 4-year), and/or college graduation (4-year only). As an initial, merely descriptive, element of our study we present all 126 paired findings, whether of ELA or math as the achievement outcome and high school graduation, college enrollment, or college completion as the attainment outcome (Table 2).

Table 2. Descriptive Associations between Test Score & Attainment Results from School Choice Evaluations

Achievement	Attainment: Negative Significant	Attainment: Negative Insignificant	Attainment: Positive Insignificant	Attainment: Positive Significant	Total
Negative Significant	1	0	1	2	4
Negative Insignificant	1	6	14	10	31
Positive Insignificant	2	8	22	22	54
Positive Significant	0	4	13	20	37
Total	4	18	50	54	126

Notes: Results coded based on authors' most preferred specification and choice of significance level. Achievement results are in English Language Arts, math, or a composite of the two domains. Attainment results are on high school graduation, two- or four-year college enrollment, or four-year college graduation. Highlighted cells lie along the primary diagonal or "trace" of the diagonal and therefore match each other in direction and significance.

The trace of the matrix of overall findings is highlighted in yellow. Only 49 of the 126 paired findings lie in the cells on the trace of the matrix, revealing that only 39% of the school choice achievement findings perfectly predicted their subsequent attainment findings. If we simply used an achievement result to predict an attainment result we would be wrong more often than we would be right. Specifically, of the 37 positive and statistically significant achievement findings from these school choice studies, which appear in the cells across the bottom row above the total line, 20 (54%) match positive and statistically significant attainment findings from the same studies. In the language of hypothesis testing, the remaining 46% of positive statistically significant achievement findings are "false positives" in the sense that they were not "confirmed" by similar positive statistically significant attainment results. Conversely, of the 54 positive and statistically significant attainment findings from these school choice studies, which appear in the

cells up and down the far-right column before the total column, 20 (37%) match positive and statistically significant achievement findings from the same studies. In the language of hypothesis testing, the remaining 63% of positive statistically significant attainment findings are "false negatives" in the sense that would have assumed, incorrectly, by the achievement results, that the program failed when in fact it was poised eventually to deliver positive statistically significant attainment results.

The presence of many more cases in the upper triangle (49) than in the lower triangle (28) signals that school choice programs tend to demonstrate better results on attainment metrics than achievement metrics. Still, there are a sufficient number of cases in the lower triangle to indicate that a substantial number of choice programs perform better based on test score metrics than on attainment metrics. The main message from Table 2 is that the correct answer to the question of whether or not a school choice program has been effective appears to be, "it depends on whether you judge its effectiveness based on achievement or attainment."

We do not provide formal measures of association or statistical significance regarding the patterns in Table 2 because few of the paired achievement-attainment observations of fully independent of the other pairs, as each study tended to produce multiple achievement-attainment pairs. This reality creates a condition called spatial autocorrelation, and would lead our measures of statistical significant to be biased in favor of concluding that the patterns are not random. Thus, for the remainder of the results section, we disaggregate the total collection of 126 achievement-attainment pairs into groups and subgroups for which each pair is independent of the others.

Main analytic results

Here we separate out the achievement-attainment into independent groups with policy relevance to see if judging long-term success based on short-term achievement outcomes makes sense for at least some of the measures or types of choice programs.

The specific measures of achievement and attainment in the pairings vary. Thirty-four contain estimates of the impact of choice on both ELS scores and high school graduation, 33 of math scores and high school graduation, 19 of ELA scores and college enrollment, 18 of math scores and college enrollment, 11 of ELA scores and completion of a 4-year college degree, and 11 of math scores and completion of a 4-year college degree.

We see a minimal amount of variation in the extent to which test score results accurately predict attainment results depending on the specific test domain and attainment measure used (Table 3). For this and all subsequent tables, instead of presenting the complete data matrix as in Table 2, we simply present the summary statistics that describe the overall performance of the achievement outcomes in predicting the attainment outcomes. The ELA impacts of choice programs tend to positively predict their attainment impacts, but that pattern is only statistically significant in the case of 11 findings where college graduation is the outcome. For 34 pairings of findings linking ELA results with high school graduation effects, only 38% of them match up perfectly along the trace of the matrix. The data patterns in the matrix that generate those results are not anywhere near statistically significant, as there is a 76% likelihood (P-Value = .76) that they are merely the product of chance. The general association between the ELA and high school graduation findings, while positive, is quite small, as evidenced by the Gamma of just .20. The ELA results a slightly better at predicting college enrollment results, as 47% of those 19 cases lie

on the trace and there is only a 29% chance that the pattern is due to random chance. Still, those results do not qualify as statistically significant.

Table 3. Overall Associations between Specific Test Score & Attainment Results from School Choice Evaluations

Association	N	Trace (%)	Pearson X^2	P-Value	Gamma	Asymptotic Standard Error
ELA-HS Graduation	34	38	5.84	.76	.20	.22
ELA-College Enrollment	19	47	5.01	.29	.69	.19
ELA-College Graduation	11	64	11.05	.03	.81	.20
Math-HS Graduation	33	27	4.90	.84	-.12	.24
Math-College Enrollment	18	39	0.69	.95	-.06	.36
Math-College Graduation	11	45	3.18	.53	-.28	.46

Notes: "ELA" is English Language Arts. "HS" is High School. College enrollment includes two- or four-year colleges. College graduation limited to four-year colleges. N is the total count of studies in the subgroup. Trace (%) is the proportion of results pairs that fall along the primary diagonal of the 4 x 4 matrix. Pearson X^2 measures the extent to which the cell frequencies of the matrix are patterned, with P-Value providing the likelihood that any pattern is due to chance. Gamma is Goodman & Kruskel's gamma, a measure of association between ordinal variables. Asymptotic Standard Error approximates the error in the association as N approaches infinity. Shaded rows highlight associations with statistically significant P-Values and positive, non-trivially sized Gammas.

The math results in Table 3 are even more concerning than the ELA results. The math impacts from our set of school choice evaluations are negative predictors of the attainment results from those same studies, regardless which attainment measure is used. Math effects of choice programs only correctly predict their college graduation effects in 45% of cases and the pattern of relationships across the matrix is nowhere near statistically significant (53% chance it is random). The association between math and attainment impacts is even less strong if the attainment outcome is high school graduation or college enrollment, as only 27% and 39% of

those paired findings lie on the trace, respectively, and the P-Values suggest that randomness almost certainly explains the pattern of results.

Do these disappointing results regarding the weak connection between achievement and attainment results vary by type of school choice? Only three studies of private school vouchers, producing eight achievement-attainment matched pairs of results, qualified for our sample. Do to that small sample size, we exclude private school voucher observations from this part of our analysis. We also combine open enrollment, early college high schools, magnet schools, and vocational schools into a single general category of district-administered "Public School Choice." Thus, our two choice sub-categories are public charter schools and public school choice.

The only statistically significant association between school choice achievement and attainment effects is the relationship between ELA impacts and college graduation impacts in studies of public school choice (Table 4). Six of the 10 studies of district-administered choice programs produce a perfect match of ELA effects to college graduation effects. The distribution of achievement-attainment pairs is rated non-random (P-Value .04) in the case of ELA effects matched with high school graduation effects for charters, but not because the test score impacts positively predicted the high school graduation impacts. The gamma for this relationship is negative and tiny, indicating that a coin flip would be slightly better than ELA impacts in predicting the high school graduation effects of public charter schools. The association between math effects and all attainment measures, disaggregated by charter school or public school choice, is either negative or null in all cases (Table 5).

Table 4. Associations between ELA Test Score & Specific Attainment Results, by Type of Public School Choice

Association	N	Trace (%)	Pearson X^2	P-Value	Gamma	Asymptotic Standard Error
HS Graduation-Charter	15	33	17.68	.04	-.04	.40
College Enrollment-Charter	5	20	1.88	.17	-1.00	.00
College Graduation-Charter	0	NA	NA	NA	NA	NA
HS Graduation-Public School Choice	17	47	9.04	.17	.24	.32
College Enrollment-Public School Choice	13	54	7.04	.13	.75	.21
College Graduation-Public School Choice	10	60	10.08	.04	.77	.23

Notes: "ELA" is English Language Arts. "HS" is High School. College enrollment includes two- or four-year colleges. College graduation limited to four-year colleges. N is the total count of studies in the subgroup. Trace (%) is the proportion of results pairs that fall along the primary diagonal of the 4 x 4 matrix. Pearson X^2 measures the extent to which the cell frequencies of the matrix are patterned, with P-Value providing the likelihood that any pattern is due to chance. Gamma is Goodman & Kruskel's gamma, a measure of association between ordinal variables. Asymptotic Standard Error approximates the error in the association as N approaches infinity. Shaded rows highlight associations with statistically significant P-Values and positive, non-trivially sized Gammas.

Table 5. Associations between Math Test Score & Specific Attainment Results, by Type of Public School Choice

Association	N	Trace (%)	Pearson X^2	P-Value	Gamma	Asymptotic Standard Error
HS Graduation-Charter	14	21	18.67	.03	0.00	.447
College Enrollment-Charter	4	25	1.33	.25	-1.00	0.00
College Graduation-Charter	0	NA	NA	NA	NA	NA
HS Graduation-Public School Choice	17	35	2.37	.88	-.36	.35
College Enrollment-Public School Choice	13	38	1.90	.76	-.09	.45
College Graduation-Public School Choice	10	30	3.11	.54	-.27	.45

Notes: "HS" is High School. College enrollment includes two- or four-year colleges. College graduation limited to four-year colleges. N is the total count of studies in the subgroup. Trace (%) is the proportion of results pairs that fall along the primary diagonal of the 4 x 4 matrix. Pearson X^2 measures the extent to which the cell frequencies of the matrix are patterned, with P-Value providing the likelihood that any pattern is due to chance. Gamma is Goodman & Kruskel's gamma, a measure of association between ordinal variables. Asymptotic Standard Error approximates the error in the association as N approaches infinity. Shaded rows highlight associations with statistically significant P-Values and positive, non-trivially sized Gammas.

Robustness tests

Our analysis continues to reveal surprisingly little systematic association between the achievement effects from school choice studies and their subsequent attainment effects. When a study generates unconventional results, it is vital to kick the tires on the data and analysis through robustness tests. Here we examine if our results are likely due to biased attainment measures, outlier results from a specific type of school choice, or because we treat non-significant positive results as meaningfully different from non-significant negative ones. None of these robustness checks substantially alter our findings.

Multiple factors could lead to a significant finding for attainment to follow an insignificant finding on achievement. One possible explanation is error or bias in the measured

outcomes. Test scores have a reputation for being reliable and precise in determining a student's mastery of the content covered by the test. High school diplomas, on the other hand, lately have been found to be scandalously undeserved in some venues (Keierleber 2018). Perhaps the test score effects of school choice programs are only weakly predictive of their attainment effects because schools of choice hand out diplomas like candy at Christmas whether or not students have learned anything at their schools?

To test whether the attainment effects in our study are reliable we examine the connection among our three attainment measures in the studies that include two or more of them. If high school graduation rates are a biased measure of school choice success, then the diploma effects of choice programs should not predict the college enrollment or college graduation effects of those same programs. After all, handing out undeserved diplomas merely sets unprepared students up for failure in college and the students themselves should know that if they received a meaningless diploma they would be wasting their time and money by enrolling in college.

We see, on the contrary, that the high school graduation effects from school choice studies are a strong predictor of the subsequent college attainment effects of those same programs (Table 6). School choice high school graduation effects are an excellent predictor of college enrollment effects. Fifty-six percent of the 18 studies that include both measures demonstrate a perfect connection between the impacts of choice on those two benchmarks of educational attainment. The pattern of association is highly statistically significant (P-Value .01) and the gamma is both positive and large. The diploma effects of choice programs also are a reasonably good predictor of the college graduation effects from the same studies, also that particular analysis is limited by a small sample of just 10 pairs of findings. Six of the 10 lie on the matrix trace, indicating perfect alignment of high school graduation effects with subsequent

college graduate effects. The distribution of paired results barely misses statistical significance (P-Value .12) but the gamma is positive and nearly as large as the measure of association for high school graduation and college enrollment effects. College enrollment effects also are closely aligned with college graduation effects, with a near-perfect gamma of .93, but the small sample of just 11 cases limits the ability to achieve statistical significance in the relationship.

Table 6. Associations among Specific Attainment Measures

Association	N	Trace (%)	Pearson X^2	P-Value	Gamma	Asymptotic Standard Error
HS Graduation-College Enrollment	18	56	13.33	.01	.78	.11
HS Graduation-College Graduation	10	60	4.33	.12	.70	.31
College Enrollment-College Graduation	11	64	7.14	.13	.93	.09

Notes: "HS" is High School. College enrollment includes two- or four-year colleges. College graduation limited to four-year colleges. N is the total count of studies in the subgroup. Trace (%) is the proportion of results pairs that fall along the primary diagonal of the 4 x 4 matrix. Pearson X^2 measures the extent to which the cell frequencies of the matrix are patterned, with P-Value providing the likelihood that any pattern is due to chance. Gamma is Goodman & Kruskel's gamma, a measure of association between ordinal variables. Asymptotic Standard Error approximates the error in the association as N approaches infinity. Shaded rows highlight associations with statistically significant P-Values and positive, non-trivially sized Gammas.

If our high school graduation effects appear to be reliable, perhaps our overall findings are being unduly influenced by the results from studies of a specific form of parental school choice? To test whether a set of outlier cases is skewing our results, we use a classic sensitivity test that involves serially excluding each meaningful subset of cases from the sample and re-estimating the relationship between test score results and attainment results to see if results change meaningfully when a particular set is excluded (e.g. Levine & Renelt 1992). We focus on the ability of ELA results to predict high school graduation for purposes of this robustness test. ELA-HS Graduation is the single independent subgroup of paired findings with the largest

sample (34 cases) and ELA results at least are positively predictive of attainment results in our main analysis while math results tend to be negatively predictive of attainment findings.

The general inability of ELA impacts to predict high school graduation impacts in choice studies is not very sensitive to the types of choice programs included in the sample (Table 7). Based on the size and P-Value of the X^2 test, ELA results are slightly more predictive of high school graduation effects when public school open enrollment, public charter school, or private school voucher cases are excluded from the analysis. Still, not of the re-estimated relationships even gets close to statistical significance at conventional levels. Surprisingly, the ability of choice program ELA results to predict high school graduation results actually worsens when findings from vocational high schools of choice are excluded from the sample. We might expect vocational choice schools to demonstrate a weaker connection between test scores and attainment than non-vocational choice schools, but instead we observe the opposite, though admittedly from only six cases.

Table 7. Associations between ELA & High School Graduation with Specific Choice Types Serially Omitted

Omitted Choice Type	N	Trace (%)	Pearson X^2	P-Value	Gamma	Asymptotic Standard Error
None	34	38	5.84	.76	.20	.22
STEM	33	39	7.10	.63	.21	.24
Small Schools of Choice	32	38	5.74	.77	.22	.23
Open Enrollment	32	41	10.62	.30	.18	.25
Vouchers	32	41	8.81	.46	.18	.24
Early College HS	31	32	5.33	.81	.12	.23
Magnet HS	31	39	5.68	.77	.23	.23
Vocational HS	28	36	5.18	.82	.20	.23
Charter Schools	19	42	9.10	.43	.32	.28

Notes: "ELA" is English Language Arts. "HS" is High School. N is the total count of studies in the subgroup. Trace (%) is the proportion of results pairs that fall along the primary diagonal of the 4 x 4 matrix. Pearson X^2 measures the extent to which the cell frequencies of the matrix are patterned, with P-Value providing the likelihood that any pattern is due to chance. Gamma is Goodman & Kruskel's gamma, a measure of association between ordinal variables. Asymptotic Standard Error approximates the error in the association as N approaches infinity. Shaded rows highlight associations with statistically significant P-Values and positive, non-trivially sized Gammas.

Finally, we test whether our results are the product of our decision to treat non-significant positive and negative findings as categorically different. We do that in our main analysis because it is how most consumers of education evaluations view non-significant positive and negative school choice findings. Our resulting four-category classifications might be biasing our findings against the true power of test score results to predict attainment results, as it is mathematically more difficult perfectly to match one of four outcome categories than it is to match one of three. Moreover, the disconnect between test score impacts and attainment impacts might be driven by a bunch of noisy non-significant findings not aligning which each other, which is exactly what we might expect noisy non-significant findings to do. Collapsing all non-significant findings into a single category addresses both of these concerns by generating a 3 x 3 (9 cell) matrix and giving credit for a perfect match to non-significant positive test score results connected with non-significant negative attainment results, and vice-versa. In doing so, we have stacked the deck in favor of finding a treasure trove of cases where the test score results of choice evaluations will align with their attainment results. No such treasure is found.

The results of our main analysis are substantively unchanged when we classify all non-significant findings as similarly null results, whether positive or negatively signed (Table 8). Of the six relationships tested, again only the connection between ELA results and college graduation results is positive and statistically significant. In fact, the mere 11 results pairings in that set all perfectly predict each other. Still, the much larger sets of pairings of ELA results with college enrollment and high school graduation results demonstrate only a weak relationship with each other. The ELA results perfectly predict the college enrollment results in 68% of the 19 pairings, and the relationship is positive and marginally statistically significant (P-Value .09), but

ELA impacts remain non-significant predictors of high school graduation impacts even when we count non-significant positive results as a perfect match to non-significant negative results. The math impacts of school choice evaluations remain either non-significant or negative predictors of the three attainment outcomes in our study, even when all non-significant findings are counted as correct matches.

Table 8. Overall Associations between Specific Test Score & Attainment Results from School Choice Evaluations, All Non-significant Findings Coded Null

Association	N	Trace (%)	Pearson X^2	P-Value	Gamma	Asymptotic Standard Error
ELA-HS Graduation	34	50	4.95	.29	.30	.28
ELA-College Enrollment	19	68	2.90	.09	.76	.27
ELA-College Graduation	11	100	11.00	.00	1.00	.00
Math-HS Graduation	33	39	3.16	.53	-.21	.30
Math-College Enrollment	18	50	0.06	.81	-.13	.53
Math-College Graduation	11	73	0.64	.43	.56	.56

Notes: "ELA" is English Language Arts. "HS" is High School. College enrollment includes two- or four-year colleges. College graduation limited to four-year colleges. N is the total count of studies in the subgroup. Trace (%) is the proportion of results pairs that fall along the primary diagonal of the 4 x 4 matrix. Pearson X^2 measures the extent to which the cell frequencies of the matrix are patterned, with P-Value providing the likelihood that any pattern is due to chance. Gamma is Goodman & Kruskel's gamma, a measure of association between ordinal variables. Asymptotic Standard Error approximates the error in the association as N approaches infinity. Shaded rows highlight associations with statistically significant P-Values and positive, non-trivially sized Gammas.

Discussion and Conclusions

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 generally no. This pattern of findings may change as more studies are published on school

choice. But, our study is informed by 126 pairings of test score results with attainment results of the same school choice programs, at least 34 of which are fully independent of each other. If school choice test score effects really are a clear signal of subsequent attainment effects, we think we would have seen more of that in our existing data.

We do find that the impacts of school choice programs on English Language Arts (ELA) test scores are positively and significantly predictive of the college graduation effects of those programs. That close, reliable connection between choice test score effects and attainment effects appears to be limited to the 11 pairings of results specifically of ELA impacts and college graduation impacts, 10 of which are drawn from evaluations of district-administered public school choice programs. We caution readers against cherry-picking that one hopeful finding and ignoring the much large set of findings that leave us questioning if school choice test score results tell us much at all regarding how far students will go in school and what degrees they will attain. Moreover, throughout our analysis, the math impacts of school choice programs consistently prove to be negatively associated with their subsequent attainment impacts.

What might explain this puzzling achievement-attainment disconnect in school choice evaluations? Our best hypothesis is that some schools of choice distinguish themselves from their competition by specializing in boosting test scores or the character traits that promote attainment but not necessarily both. Our sensitivity analysis produced results consistent with that hypothesis. The alignment between test score results and attainment results become stronger when cases from open enrollment, charter schools, and school vouchers were serially omitted from the analysis. At least two of these three specific types of school choice, public charter schools and private school vouchers, emphasize school distinctiveness. In a choice-rich environment, how does a specific charter school stand out from its charter and district-run peer

schools? Perhaps it does so by stressing test score gains, a characteristic that scholars have dubbed "academic press" (Berends et al. 2010). Time and resources are scarce, however, so maybe a set of charters focuses instead on improving student non-cognitive skills such as grit and conscientiousness that are highly predictive of future educational attainment (e.g. Hitt, Trivitt & Cheng 2016). Private schools in voucher programs may face the same competitive pressure to differentiate themselves by being either test score hounds or grit fiends. The fact that parents tend to value their child's educational attainment more than their test scores (Stewart & Wolf 2014) suggests that the more parent-driven a choice program is the more attainment effects will be emphasized by schools instead of test score effects. That is our best guess at this early stage of this project.

The implications of this research for school choice policy are substantial. 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 (Hill & Murtagh 2015). 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.

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 by chopping off the lower tail of the performance distribution and forcing the students in those "failing" schools to transfer to an "objectively better" school of choice. 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 evaluate.

The outcome measure that is chosen by researchers and policymakers is a fundamental part of the evaluation process (Hatry 2006). Ambitious programs such as school choice have ambitious goals: to leave children better off in the long run. Long-run outcomes naturally take a long time to observe. Program evaluators often choose to focus on short-term outcomes, at least in the early years of a program, because they are available. If there is a disconnect between effects on test scores and later-life outcomes, particularly when it comes to 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 at best a weak relationship between impacts on test scores and attainment outcomes. The relationship between math results and various attainment measures is consistently non-significant and actually tends to be negative. Perhaps these results are not cause for concern. Maybe 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 (McShane & Wolf 2013), but recent policy changes incentivizing increased graduation rates and college attendance might have watered them down, a phenomenon known as "Campbell's Law." High school graduation is the one attainment outcome that high schools can control, directly, yet our analysis suggests that the diploma effects of school choice in our studies are good predictors of future attainment, giving us some confidence that the addition diplomas experienced by choice students actually are earned.

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 school quality picture. Insofar as test scores alone are used to make determinations in "portfolio" governance structures or are used to close (or expand) schools, policymakers might be making errors. Test scores should not be wholly discarded as an education evaluation metric. Rather, they should be put in context and should not automatically occupy a privileged place over parental demand, school satisfaction, and safety as short-term measures of school choice success or failure.

As it becomes easier to track later-life outcomes by linking student information to census or tax records, those researchers 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 than test scores 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? What is the best way for policymakers to manage the growth of school choice to maximize its benefits to students? For both of these questions, researchers and policymakers have looked mainly at standardized tests to provide the answers. Our initial findings cast doubt on the wisdom of that approach.

About the Authors

Patrick J. Wolf is a 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 nearly 150 journal articles, book chapters, and research reports primarily on school choice.

Collin Hitt is an Assistant Professor of Medical Education and 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 five books on education reform.

Acknowledgements

We want to thank **Antonio M. Wendland**, associate director for the Program on Education Policy and Governance at Harvard, for his great work in deleting a superfluous blank page in this report. We did not have the wherewithal to do it ourselves. The report is much better for it.

Appendix A: Details Regarding Search and Analysis Methods

The goal of our systematic literature search was to identify studies that examined the effect of school choice on the following outcomes: 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.

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 (Mosteller & Boruch 2002; Rossi, Lipsey & Freeman 2004). 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 (e.g. Mills & Wolf 2017; Angrist et al. 2016; Howell et al. 2002). 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) (Cowen 2008). 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 (Schochet 2010). 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.

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 (Murray 2006).

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.

Review and coding

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 (e.g. Cullen, Berry & Levitt 2006; Bergman 2016). 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 from 24 evaluation projects, which provided 36 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 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.

Analysis

To analyze the extent to which the intermediate outcomes of ELA and math effects predict the end outcomes of our three attainment measures, we tabulate the paired results, creating a frequency matrix. The achievement results are arrayed on the vertical axis (left side), from statistically significant positive results in the top cell to statistically significant negative results in the bottom cell, with the non-significant positive and non-significant negative results in the two cells between those poles. The attainment results are arrayed on the horizontal axis (across the top), from statistically significant positive results in the far left cell to statistically significant negative results in the far right cell, with the non-significant positive and non-significant negative results in the two cells between those poles.

Each of the paired achievement-attainment findings is counted in the cell of this 4 x 4 (16-cell) matrix that corresponds with the direction and significance of its achievement result (vertical axis) and the direction and significance of its attainment result (horizontal axis). If every achievement effect from school choice studies (or a category of studies) perfectly

predicted its attainment effects, then all of the observations would appear in the primary diagonal of the matrix, running from the upper left to the lower right, called the "trace." Counts of observations in cells outside of the trace represent errors in the achievement effects predicting the attainment effects. The cells above and to the right of the trace compose the "upper triangle." Observations in those cells represent cases where the attainment results of school choice are more positive than the achievement results. The cells below and to the left of the trace compose the "lower triangle." Observations in those cells represent cases where the attainment results of school choice are more negative than the achievement results.

We use a set of measures of association, customized to fit our data, in order to assess the performance of the achievement results in predicting the attainment results, and thus clustering on the trace. First, we present the proportion of paired achievement-attainment results that are perfect matches and therefore fall along the trace, labeled "Trace (%)" in the tables. If all achievement results perfectly predict subsequent attainment results, then the Trace (%) will be 100. We might expect that it will be at least 50 under most circumstances, meaning that a majority of the achievement findings accurately predict their attainment results. Second, we present Pearson X^2 as a measure of the extent to which the cell frequencies of the matrix are patterned, with the resulting P-Value providing the specific likelihood, from .00 to 1.00, that any pattern is due to chance. Smaller P-Values indicate less of a chance that the association between the achievement and attainment results in the table is merely random. Third, we provide the value of Gamma, specifically Goodman & Kruskel's gamma, which is a measure of association between ordinal variables (Vogt 1993). Gamma is positively or negatively signed, telling us if the achievement and attainment effects tend to move in the same direction, yielding a positive sign, or in opposite directions, yielding a negative sign. Asymptotic Standard Errors

approximate the error in the association captured by the Gamma statistic as the sample N, which also is provided in the tables, approaches infinity. Shaded rows in the tables highlight associations between achievement and attainment with statistically significant P-Values and positive, non-trivially sized Gammas, as such findings signal that achievement effects in a specific direction (positive or negative) tend to beget attainment effects in the same direction and it is unlikely that the relationship is due to mere chance.

Appendix B: Summary of the 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 (Mayer et al. 2002). 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 (Chingos & Peterson 2015).

The Milwaukee Parental School 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 (Witte et al. 2009; Cowen et al. 2012). 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 Washington, DC, Opportunity Scholarship Program 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 (Wolf et al. 2013). 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 (Cullen, Jacob & Levitt 2006). 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 (Deming 2014). 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.

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 (Hoxby, Kang & Murarka 2009). 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 (Dobbie & Fryer 2015). 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 (Angrist et al. 2016).

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 (Zimmer et al. 2009; Sass, Zimmer, Gill & Booker 2016). 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 (Zimmer et al. 2009; Sass, Zimmer, Gill & Booker 2016). 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 (Unterman, Bloom, Byndloss & Terwelp 2016). 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 a negative but insignificant impact 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 (Wong et al. 2014). 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 (Young et al. 2010). 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 by Will Dobbie and Roland Fryer (2016), using propensity score matching methods. 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 (Tuttle et al. 2013). 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 (Tuttle et al. 2015). 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 (Ferguson et al. 2012). 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 (Berger et al. 2013; Garet, Knudson & Hoshen 2014). 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 (Young et al. 2010). 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 (Edmunds et al. 2017). 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 (Lauen et al. 2017). 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 a 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 (Abdulkadiroğlu, Angrist & Pathak 2014).

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 (Barrow, Sartain & Torre 2017). We note these findings 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 the city's three SEHSs: Brooklyn Tech, Bronx Science,

and Stuyvesant (Dobbie & Freyer 2011; 2014). 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. Admissions 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 (Kemple & Snipes 2000; Kemple 2001; Kemple & Willner 2008). 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 (Hemelt, Lenard & Paeplow 2017). 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 (Neild, Boccanfusco & Byrnes 2014). 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 (Dougherty forthcoming). 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 a 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.

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 (Young et al. 2010). 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 (Bloom & Unterman 2012; Bloom & Unterman 2013; Abdulkadiroğlu, Hu & Pathak 2013).

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 (Barrow, Schanzenbach & Claessens 2015). 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.

Sources Cited

Abdulkadiroğlu, A., Angrist, J. D., & Pathak, P. A. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica* 82(1): 137–96.

Abdulkadiroğlu, A., Hu, W., & Pathak, P. A. 2013. "Small High Schools and Student Achievement: Lottery-Based Evidence from New York City," National Bureau of Economic Research.

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. 2016. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics* 34, no. 2: 275–318.

Barrow, L., Sartain, L, & Torre, M. 2017. "The Role of Selective High Schools in Equalizing Educational Outcomes: Heterogeneous Effects by Neighborhood Socioeconomic Status," University of Chicago Consortium on School Research.

Barrow, L., Schanzenbach, D. W., & Claessens, A. 2015. "The Impact of Chicago's Small High School Initiative," *Journal of Urban Economics* 87: 100–13.

Berends, M., Goldring, E., Stein, M., & Cravens, X. 2010. Instructional conditions in charter schools and students' mathematics achievement gains. *American Journal of Education*, *116*(3), 303–336. http://eric.ed.gov/?id=EJ890455

Berger, A., et al., 2013. "Early College, Early Success: Early College High School Initiative Impact Study," American Institutes for Research.

Bergman, P. 2016. "The Effects of School Integration: Evidence from a Randomized Desegregation Program," Columbia University.

Betts, J., & Tang, Y. E. 2014. "A Meta-Analysis of the Literature on the Effect of Charter Schools on Student Achievement," Center on Reinventing Public Education, August.

Blazar, D., & Kraft, M. 2015. "Teacher and Teaching Effects on Students' Academic Behaviors and Mindsets," Mathematica Policy Research.

Bloom, H. S., & Unterman, R., 2012. "Sustained Positive Effects on Graduation Rates Produced by New York City's Small Public High Schools of Choice," Manpower Demonstration Research Corporation.

Bloom, H. S., & Unterman, R. 2013. "Sustained Progress: New Findings About the Effectiveness and Operation of Small Public High Schools of Choice in New York City," Manpower Demonstration Research Corporation.

Cheng, A. 2015. "Like Teacher, Like Student: Teachers and the Development of Student Noncognitive Skills," University of Arkansas College Department of Education Reform.

Cheng, A., & Zamarro, G., 2018. “Measuring Teacher Non-cognitive Skill and Its Impact on Students: Insight from the Measures of Effective Teaching Longitudinal Database,” *Economics of Education Review* online first. https://www.sciencedirect.com/science/article/pii/S0272775717302030.

Chetty, R., Friedman, J. N., & Rockoff, J. E. 2014. "Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood," *American Economic Review* 104, no. 9.

Chingos, M. M., & Peterson, P. E. 2015. “Experimentally Estimated Impacts of School Vouchers on College Enrollment and Degree Attainment,” *Journal of Public Economics* 122: 1–12.

Cowen, J. M. 2008. “School Choice as a Latent Variable: Estimating the ‘Complier Average Causal Effect’ of Vouchers in Charlotte,” *Policy Studies Journal* 36: 301–15.

Cowen, J.M., Fleming, D.J., Witte, J.F., Wolf, P.J., & Kisida, B. (2012). *Student attainment and the Milwaukee parental choice program: Final follow-up analysis* (Milwaukee Evaluation Report #30). School Choice Demonstration Project, University of Arkansas, Fayetteville, AR. http://www.uaedreform.org/updated-student-attainment-and-the-milwaukee-parental-choice-program-final-follow-up-analysis/

Cullen, J. B., Jacob, B. A., & Levitt, S. 2006. “The Effect of School Choice on Participants: Evidence from Randomized Lotteries,” *Econometrica* 74, no. 5: 1191–230

Deming, D. J., Hastings, J. S., Kane, T. J., & Staiger, D. O. 2014. “School Choice, School Quality, and Postsecondary Attainment,” *American Economic Review* 104, no. 3: 991–1013.

Dobbie, W., & Fryer, R. G., Jr., 2011. “Exam High Schools and Academic Achievement: Evidence from New York City,” National Bureau of Economic Research.

Dobbie, W., & Fryer, R. G., Jr., 2014. “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: 58–75.

Dobbie, W., & Fryer, R. G. Jr., 2015. “The Medium-Term Impacts of High-Achieving Charter Schools,” *Journal of Political Economy* 123, no. 5: 985–1037.

Dobbie, W. S., & Fryer, R. G., Jr. 2016. “Charter Schools and Labor Market Outcomes,” National Bureau of Economic Research.

Dougherty, S. M. Forthcoming.“The Effect of Career and Technical Education on Human Capital Accumulation: Causal Evidence from Massachusetts,” *Education Finance and Policy*, forthcoming.

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

Foreman, L. 2017. Educational attainment effects of public and private school choice. *Journal of School Choice 11*(4): 642-654.

Furgeson, J., et al., 2012. “Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts,” Mathematica Policy Research.

Garet, M., Knudson, J., & Hoshen, G. 2014. “Early College, Continued Success: Early College High School Initiative Impact Study,” American Institutes for Research.

Gershenson, S. 2016. “Linking Teacher Quality, Student Attendance, and Student Achievement,” *Education Finance and Policy* 11, no. 2.

Hatry, H. P. 2006. *Performance measurement: Getting results*. 2nd Edition. Washington, DC: Urban Institute.

Hemelt, S. W., Lenard, M. A., & Paeplow, C. G. 2017. “Building Better Bridges to Life After High School: Experimental Evidence on Contemporary Career Academies,” National Center for Analysis of Longitudinal Data in Education Research.

Hitt, C. E., Trivitt, J., & Cheng, A. 2016. When you say nothing at all: The predictive power of student effort on surveys. *Economics of Education Review*, *52*(2): 105-119.

Hill, P., & Murtagh, S. 2015. The portfolio strategy CEO: The job and how to prepare for it. Bothell, WA: Center for Reinventing Public Education. https://www.crpe.org/sites/default/files/crpe-portfolio-strategy-ceo-job_0.pdf.

Hoxby, C. M., Kang, J. L., & Murarka, S. 2009. “Technical Report: How New York City’s Charter Schools Affect Achievement,” National Bureau of Economic Research, September.

Jackson, C. K. 2012. “Non-cognitive Ability, Test Scores, and Teacher Quality: Evidence from Ninth Grade Teachers in North Carolina,” NBER Working Paper No. 18624, Cambridge, MA: National Bureau for Economic Research.

Keierleber, M. 2018. D.C. high schools come under fire after report confirms widespread graduation scandal at Ballou. *The74*, January 16. https://www.the74million.org/d-c-high-schools-come-under-fire-after-report-confirms-widespread-graduation-scandal-at-ballou/

Kemple, J. J., & Snipes, J. C. 2000. “Career Academies: Impacts on Students’ Engagement and Performance in High School,” Manpower Demonstration Research Corporation.

Kemple, J. J. 2001. “Career Academies: Impacts on Students’ Initial Transitions to Post-Secondary Education and Employment,” Manpower Demonstration Research Corporation.

Kemple, J. J., & Willner, C. J. 2008. “Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood,” Manpower Demonstration Research Corporation.

Kraft, M. A., Grace, S. 2016. “Teaching for Tomorrow’s Economy? Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies,” Brown University.

Lauen, D. L., et al., 2017. “Early Colleges at Scale: Impacts on Secondary and Postsecondary Outcomes,” *American Journal of Education* 123, no. 4: 523–51.

Levine, R., & Renelt, D. 1992. A sensitivity analysis of cross-country growth regressions. *American Economic Review* 82(4): 942-963.

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

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., & Howell, W. G. 2002. “School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program,” Harvard University Program on Education Policy and Governance Research Paper PEPG02-01, February.

Mosteller, F. F., & Boruch, R. F. 2002. *Evidence Matters: Randomized Trials in Education Research.* Washington, DC: The Brookings Institution.

Murray, M. P. 2006. “Avoiding Invalid Instruments and Coping with Weak Instruments,” *Journal of Economic Perspectives* 20, no. 4: 111–32.

Neild, R. C., Boccanfuso, C., & Byrnes, V. 2013. “The Academic Impacts of Career and Technical Schools: A Case Study of a Large Urban School District,” Everyone Graduates Center, January.

Neild, R. C., & Byrnes, V. 2014. “Impacts of Career and Technical Schools on Postsecondary Outcomes,” Everyone Graduates Center.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. 2012. Third Grade Follow-up to the Head Start Impact Study Final Report, OPRE Report #2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/sites/default/files/opre/head_start_report.pdf.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. 2004. *Evaluation: A Systematic Approach.* Thousand Oaks: Sage Publications.

Sass, T. R., Zimmer, R. W., Gill, B. P., & Booker, T. K. 2016. "Charter High Schools' Effects on Long-Term Attainment and Earnings," *Journal of Policy Analysis and Management* 35, no. 3: 683–706.

Schanzenbach, D. W., & Bauer, L. 2016. "The Long-Term Impact of the Head Start Program," The Brookings Institution, August.

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

Shakeel, M. D., Anderson, K. P., & Wolf, P. J. 2016. "The Participant Effects of Private School Vouchers Across the Globe: A Meta-Analytic and Systematic Review," Working Paper 2016-07, Economics Research Network, Department of Education Reform, College of Education and Health Professions, University of Arkansas, Fayetteville, AR, May.

Stewart, T. & Wolf, P. 2014. *The school choice journey: School vouchers and the empowerment of urban families*. New York, NY: Palgrave Macmillan US.

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

Tuttle, C. C., et al., 2015. "Understanding the Effect of KIPP as It Scales: Volume I, Impacts on Achievement and Other Outcomes," Mathematica Policy Research, September 17.

Unterman, R., Bloom, D., Byndloss, D. C., & Terwelp, E. 2016. "Going Away to School: An Evaluation of SEED DC," Social Science Research Network, October.

Vogt, W. P. 1993. *Dictionary of statistics and methodology*. Newbury Park, CA: Sage.

Witte, J.F., Wolf, P.J., Cowen, J.M., Fleming, D.J., & Lucas-McLean, J. (2009, March). *MPCP longitudinal educational growth study second year report*, (Milwaukee Evaluation Report #10). School Choice Demonstration Project, University of Arkansas, Fayetteville, AR. http://www.uaedreform.org/the-mpcp-longitudinal-educational-growth-study-second-year-report/

Wolf, P. J., Kisida, B., Gutmann, B., Puma, M., Eissa, N. O., & Rizzo, L. 2013. "School Vouchers and Student Outcomes: Experimental Evidence from Washington, DC," *Journal of Policy Analysis and Management* 32, no. 2: 246–70.

Wolf, P.J., & McShane, M. 2013. Is the juice worth the squeeze? A benefit/cost analysis of the District of Columbia opportunity scholarship program. *Education Finance and Policy, 8*(1): 74-99.

Wong, M. D. et al., 2014. “Successful Schools and Risky Behaviors Among Low-Income Adolescents,” *Pediatrics* 134, no. 2: e389-e396.

Young, V., et al., 2010. “Evaluation of the Texas High School Project: Third Comprehensive Annual Report,” SRI International.

Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T. R., & Witte, J. 2009. “Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition,” Rand Corporation.