CRPE
REINVENTING
PUBLIC EDUCATION

A Meta-Analysis of the Literature on the Effect of Charter Schools on Student Achievement

WORKING PAPER

August 2014

Julian R. Betts
Department of Economics,
University of California, San Diego,
and National Bureau of Economic Research

Y. Emily Tang
Department of Economics,
University of California, San Diego

Abstract

A meta-analysis is performed of the literature on charter schools and achievement, with a focus on lottery-based studies and rigorous value-added studies. Overall, for the limited set of charter schools, locations, and years that have been studied to date, charter schools are producing higher achievement gains in math relative to traditional public schools in most grade groupings. No significant differences emerge for reading achievement. However for both math and reading, the bulk of estimates are positive. For math, middle school studies tend to produce higher effect sizes than other grade groupings. For math, studies that use lotteries or propensity score matching tend to find higher effects than other methods. There is not a statistically significant link between the years covered by a study and the estimated effect size, but for both math and reading the trend is positive. A tiny but growing literature on nonachievement outcomes suggests positive influences of charter schools on educational attainment and behavioral outcomes.

Acknowledgments

We would like to thank the many researchers who provided supplementary information we needed to incorporate their papers into this or our earlier (2011 and 2008) literature reviews. Dale Ballou, Richard Buddin, Caroline Hoxby and Jonah Rockoff, Anna Nicotera and Mark Berends, Macke Raymond, Christopher Reicher and Larry McClure, Tim Gronberg, and Scott Imberman all provided key data we needed for one or more of our analyses. We are grateful to reviewers Betheny Gross and Brian Gill, and to Robin Lake for helpful suggestions.

Funding for this project comes from the Walton Family Foundation. We thank the Foundation for its support but acknowledge that the findings and conclusions presented here are those of the authors alone and do not necessarily represent the opinions of the Foundation.

Table of Contents

1. Introduction

Charter schools represent an increasingly important form of school choice in the United States. Charter schools are public schools but they are exempted from some of the state laws and regulations that govern traditional public schools. The intent is that charter schools can provide students with alternative curricula, teaching methods, and teachers who may differ in educational background and training from teachers in traditional public schools. This freedom to experiment and innovate comes with the threat that the charter authorizer may shut down charter schools should they fail to meet academic standards or to maintain financial viability.

The most important question to ask about charter schools is whether their students benefit academically and, increasingly, to understand whether the impact of charter schools is changing as the sector matures.[1] This report seeks to answer that question based on a survey of the literature on charter schools and achievement. The report updates and extends prior papers we have written (Betts and Tang 2008, 2011). Both of the earlier studies found that results varied by subject tested and grade span, but overall achievement in math and reading of students at charter schools was typically the same or higher than that of comparison groups of students at traditional public schools.

Following our earlier papers, we focus on charter school studies that adopt one of two methods. The first approach involves comparing students who win and lose lotteries to attend charter schools, which is essentially an experimental method. The second approach, known as value-added modeling, is not experimental, but takes into account a student's past academic achievement, unlike some of the weaker nonexperimental studies. The Charter School Achievement Consensus Panel (2006) argues that these approaches are superior to cross-sectional studies that examine the levels of achievement of students at one point in time, and to studies that compare successive cohorts of students in a given grade. Both of these approaches are likely to entail severe biases due to factors that the authors were not able to take into account, and so we exclude such studies.[2]

Only eight papers have used the lottery approach to date, including a total of 136 charter schools. Non-lottery-based studies that take the value-added approach while also constructing a comparison group against which to benchmark the academic progress of charter school students are far more abundant. In total, the present report includes in its analysis 52 value-added papers that use lottery-based or rigorous value-added approaches. (This consists of 38 studies used in

1. Some might argue that an equally important question is whether there is an effect on the test scores of students who remain in traditional public schools. See Betts (2009) for a review of a very small literature on whether charter schools improve academic achievement in traditional public schools through increased competition. Another mechanism through which charter schools could influence achievement in traditional public schools is by changing the average initial achievement of peers in traditional public schools.
2. Betts, Tang, and Zau (2010) use data from San Diego and show that models that do not measure individual students' achievement growth produce quite different results from the more sophisticated value-added models, and that the changes in estimated effects of charters are consistent with the idea that the weaker approaches fail to take into account the relatively disadvantaged backgrounds of students who attend charters.

our 2011 review, 12 new studies, and two studies that represent updates to two studies included in our 2011 report. These updates include Tuttle et al. (2013) and CREDO (2013d). Together, these studies produce 97 estimates of the charter school effects for a given grade span. (Some studies, for example, make three contributions by presenting separate estimates for elementary, middle, and high schools.)

Our analysis considers the overall impact of attending a charter school as well as variation of impacts by school type and context. Our specific goals are as follows:

1. To estimate the overall average impact of attending a charter school on math and reading achievement;
2. To measure the variation in this estimated mean impact across studies, and to infer what proportion of this variation reflects true variations in school quality, as opposed to mere statistical noise;
3. To estimate how charter school achievement effects vary by grade, subject tested, demographic background, and special needs background of students, and by urbanicity of the school;[3]
4. To compare our results to those in our prior review (Betts and Tang 2011) as a simple way of looking for changes in the impact of charter schools over time as the charter school movement has started to mature; and
5. To begin to introduce a new analysis of the relationship between charter school attendance and student outcomes other than math and reading achievement.

Overall, we find that charter schools are producing higher achievement gains in math relative to traditional public schools in most grade groupings, with the bulk of the math studies producing positive estimates. The results for reading, however, are less clear. Our meta-analysis revealed no significant differences between charter and traditional district schools yet most of the studies in our sample find significant positive results. Our analysis of non-achievement indicators, while still formative and based on limited studies, shows positive effects for charter schools.

2. Methods and Challenges for Meta-Analysis of the Literature, and an Assessment of Alternative Methods of Evaluating the Impact of Charter Schools

Our Methods of Analysis

We use several approaches to summarize the results. First, we perform a formal meta-analysis. This is a widely used method, especially in medical literature, of combining estimates from a variety of studies to draw overall conclusions. We use this approach to test whether the overall impact of charter schools on achievement is zero, to portray visually the variation across studies,

3. Items 2 and 3 focus on measuring and explaining variations in impact across schools. It is crucial to look for heterogeneous effects of charter schools because identification of outperforming and underperforming charter schools can potentially improve the average quality of charter schools overall. Outperforming schools can be expanded or replicated, while underperforming schools can be improved or, if necessary, closed.

and to estimate the degree to which this is real variation in the effectiveness of charter schools as opposed to statistical noise. Second, we illustrate the variation in the estimates using histograms. Finally, we use traditional vote-counting methods to show the number of studies that yield positive and significant, insignificant (either positive or negative), or negative and significant results. Though easy to understand, this method is criticized because it might wrongly interpret studies that find "no significant results" when in truth these studies were unlikely to yield an effect due to limited statistical power (often the result of a small sample size). This criticism is valid. However, as we will show, in the charter literature far more studies produce significant results than one would expect if small samples were biasing researchers toward concluding "no significant effects." The results of the vote count serve to accentuate our finding that charter schools are likely to outperform their traditional public school counterparts in some instances, and underperform in others.

Challenges for Meta-Analysis of the Literature

Table 10, in the appendix, shows the set of papers that we included in at least one of our three research methods, along with information on the geographic location and time span of the study.

These analyses present several challenges. Though improved from the set of studies covered in our 2008 analysis, the available studies offer limited geographic coverage, potentially leading us to overstate the generalizability of results. To provide readers with a sense of how broadly based a given result might be, in section 3 we report not only an overall effect size but also the number of studies and the number of geographic locations underlying a given estimate. Similarly, middle school studies far outnumber studies of elementary schools, especially for specific student groups, such as African American students attending elementary charter schools.

A second challenge is that charter schools and studies of them display different grade configurations. Fortunately, the number of studies estimating separate effects for estimates for elementary, middle, and high school levels has grown considerably over the last three years. This is a helpful development because it is not clear why one should expect charter school impacts to be the same across these various grade levels. However, many studies still aggregate elementary and middle schools (e.g., combining grades K–8). Others combine elementary, middle, and high schools, which we refer to as an "all grade-span."[4]

Another issue is how to weight the various studies. In the current paper we use a standard meta-analytic approach that assumes that variations across studies come from sampling error as well as random variation in the true effect size. Studies that produce more precise estimates will have higher weight than studies that are less precise. (By "a precise estimate" we mean that there is relatively little uncertainty about the size of the true underlying effect.) Thus, we assume that

4. We do not replicate section 3 of Betts and Tang (2011), which tested whether one could maintain the hypothesis of no negative effects of charter schools in the literature and, conversely, the hypothesis of no positive effects. That past analysis showed that for all but two combinations of grades studied and subjects tested, there is very strong evidence of both positive and negative effects in the literature.

variations in estimates across studies in part reflect true variation in the impact of charter schools on achievement. Because we typically find that well over 90 percent of the variation across studies is likely to be true variation, variations across studies in the precision of the estimates contribute only modestly to the weights for each study.

There are a number of locations in which multiple charter school effectiveness estimates exist because different authors have studied the same place. In the appendix, table 11 lists these studies.

We considered several methods of combining multiple estimates of the same place. Because the method of meta-analysis relies on independent estimates of the effect of a particular treatment, including presumably correlated effect sizes from multiple studies of a single place would introduce bias into the meta-analysis estimate of overall charter school effectiveness. We therefore sought to combine the information from the separate studies of a single place in some way. While taking an average of the effect sizes between studies is straightforward, combining the information from standard errors of the effect sizes requires making assumptions about the correlation between studies.

In most cases, the grade spans or time periods studied differ substantially between the studies. In these cases, we include each study in the appropriate grade span analysis without concerns about "double-counting" because while one study contributes to the understanding of the effectiveness of charter schools at the elementary level, the other may contribute to charter effectiveness at the combined elementary and middle school level, which our analysis treats as different real effects. (However, in this specific case we also try combining studies of elementary schools, studies of middle schools, and studies that combine the two.) In some cases, the years overlap but the start or end years differ between studies. These cases present a challenge because it is not clear how much correlation to assume between studies. For example, if the charter school operating environment is consistent over the years, then we would assume that correlation between an "earlier" study and a "later" study of the same place is relatively high. However, given the intentionally dynamic nature of the charter school environment, it is likely that there are also real differences in the effectiveness of charter schools, even those in the same place, between time periods. Put differently, we know there are real similarities, and also real differences, between the effects estimated in studies of the same place over different time periods, but we do not know how much similarity to assume.

After considering various assumptions about the correlation between studies, we concluded that the most transparent approach to handling these situations is to include all estimates, except in cases for which the estimates would be very closely correlated and therefore render the meta-analysis estimates unreliable.

An urgent need to standardize the reporting of results in charter school research

Much of the existing research does not present results in a way that allows readers to infer important information. One problem is that some studies present an effect size and indicate whether it is statistically significant, but do not present the standard error. This is a major problem because such studies cannot be included in a meta-analysis without knowledge of this measure of how precise the estimate might be. Another recent pattern is that some papers will report on tests for differences in estimated impacts across student subgroups, such as by race/ethnicity, but do not present the actual estimated impacts for the student subgroups. This omission is unfortunate because there is a genuine policy interest in knowing what the impact of charter schools is for various student subgroups, rather than knowing only whether the differences across groups are statistically significant. Finally, we continue to find that some papers do not report the exact number of charter schools being studied or the sample of charter school students. We urge researchers to report the following:

1. the effect sizes accompanied by standard errors,
2. the effect sizes and standard errors for subgroups, rather than just the difference in effect sizes between the given subgroup and another group, and
3. the number of charter schools and the number of charter school students in the study.

An Assessment of Alternative Methods of Evaluating the Impact of Charter Schools

Although it is clear that lottery-based and value-added models provide far more credible estimates than do the many cross-sectional studies that merely take a snapshot of schools at a single point in time, it is worth pointing out that none of the most popular methods used in the studies we cover is fail-proof. We present a brief summary of these issues below. Interested readers can find a more detailed explanation in Betts and Tang (2011).

Lottery Studies

The primary advantage of lottery studies is that, subject to some straightforward data checks, the studies will produce unbiased estimates of the impact of winning a lottery. This approach is useful because the only difference between those who are admitted and those who are not admitted is the luck of the draw. (In contrast, nonexperimental studies that compare students at charter schools with those at traditional public schools run a risk that there are very real differences between these two groups of students.)

The primary weakness of lottery-based studies is that by definition they focus solely on schools and grades for which the number of applicants exceeds the number of slots, which enables researchers to compare lottery winners to losers. Popular schools with lotteries are likely to outperform less popular charter schools, leading these studies to overstate the effect of charter schools overall. Thus the external validity of lottery-based studies may be quite low. Betts and

Atkinson (2012) point out that many of the lottery-based studies produce higher estimated impacts of attending a charter school than other methods. However, it also seems possible that lotteries are only used where the demand for charter schools exceeds the number of available spots, which implies that the lottery studies will mostly include charter schools that are recognized by parents and students to be high performing relative to nearby traditional public schools. Betts and Atkinson (2012) also point to a number of non-lottery-based value-added studies in these locations that replicate the lottery results quite closely, suggesting that it is not the method of analysis used in the lottery studies, so much as a genuinely higher effect in locations with highly oversubscribed schools.[5]

A second potential problem with lotteries is differential attrition from the data among the lottery losers. If students who do not gain admission to their preferred charter school are more likely to leave the district (and therefore the overall study), it seems plausible that highly motivated and concerned parents would be the ones most likely to leave the sample once their child loses a charter school lottery. This could induce an overstatement of the impact of attending a charter school. However, this same problem applies to nonexperimental studies as well.

An important aspect of lottery-based studies is that they can produce two distinct estimates: "intent to treat" and the impact of "treatment on the treated." The intent to treat refers to the causal impact of winning a lottery.

The impact of treatment on the treated provides an estimate of the impact on a student of actually attending a charter school after winning a lottery. There are methods that researchers can use to scale up the intent-to-treat estimate to account for lottery winners who do not attend a charter school and, conversely, for lottery losers who still end up gaining admission to a charter school. These latter "impact of treatment on the treated" estimates are the ones more comparable to the estimates from the value-added literature.[6]

---

5. The external validity of lottery studies is a major issue. For example, the path-breaking nationwide study by Gleason et al. (2010) reports that only 130 out of 492 charter middle schools nationwide in fact used admission lotteries, and further, only 77 of the 130 charter schools that were oversubscribed were willing to participate in the study. This raises concerns about how representative the schools that used lotteries and that were willing to participate might be.

6. The latter estimate will be higher in absolute value because the intent-to-treat estimate is a weighted average of what is presumably a zero effect of winning a lottery but not attending a charter school, plus the effect of winning a lottery and attending a charter school. For instance, suppose that the impact of winning a lottery and enrolling in a charter school is 0.4 test score points, but only one quarter of lottery winners attend a charter school. The overall impact of winning a lottery then, would be $\frac{1}{4} \times 0.4 + \frac{3}{4} \times 0 = 0.1$. Estimates of the impact of winning a lottery are referred to as "intent to treat" estimates, while estimates of the impact of actually attending a charter school are referred to as the impact of treatment on the treated.

Propensity Score Matching

The main weakness of non-lottery-based methods is that typically they compare students who attend and do not attend charter schools. These comparisons can lead to biased estimates of the impact of attending a charter school because there are many characteristics, observed and unobserved, that could vary between the two sets of students.

Propensity score matching is one method to control for the observed reasons why students elect to attend charter schools. These studies match charter school attendees with non-charter attendees who have similar estimated probabilities of attending a charter school. This approach is very useful, but it is subject to bias because the method cannot control for unobservable variables that might be related both to the chances of applying to a charter school and to the outcome being modeled. For instance, highly motivated students and families might be more likely to apply to charter schools. Because motivation is hard to measure, this creates the risk of an upward bias in the estimated effect of attending a charter school in these studies, because they cannot control for motivation, which may be correlated with both the probability of applying and test score growth.

The Center for Research on Education Outcomes (CREDO) has produced a string of studies of charter schools for a variety of states, using a matching method that is somewhat similar but not identical to propensity score matching (e.g., CREDO 2009a, 2011, 2012a, 2012b, 2013a, 2013b, 2013c). This approach is subject to the same issue as propensity score models: it could be that students who self-select into charter schools are different from students at traditional public schools for unobservable reasons. A particular concern about the CREDO approach that distinguishes it from other approaches is that it does not require a pre-treatment match (that is, a match between a charter student and a non-charter student made prior to the charter school student entering the charter school). Rather, a charter school student may be matched based on his or her achievement two or more years after starting at a charter school. This could lead to biased estimates if the true causal effects of attending a charter school are non-zero. There are other technical issues with the CREDO studies that we will discuss later. On the other hand, even though there are concerns about potential biases in the CREDO studies, they include extremely large samples of charter schools, and thus do not share issues about external validity to the same degree as smaller studies.

Student Fixed-Effect Models

Student fixed-effect models prevent the need to use students at traditional public schools as a comparison group, because the charter school student becomes his or her own comparison group. That is, we compare achievement growth during years enrolled in a charter for a given student with the growth for the same student in years not enrolled in a charter school.

However, this method has its own issues because identification is based on students who switch between charter and traditional public schools. In elementary schools, many students start in charters and do not switch, so it is hard to extrapolate fixed-effect results to such students. Thus there are issues about external validity in fixed-effect studies, especially at the elementary level.

Second, fixed-effect models can control for unobserved heterogeneity among students only to the extent that the heterogeneity is fixed over time. But students who switch between the two types of schools may have done so due to unobserved factors that evolve over time. For instance, if students sometimes transfer to charter schools after having had a bad year in a traditional public school, and their achievement would have improved regardless of whether they switched, then we would overstate the impact of charter schools on achievement.[7] This is a version of the so-called Ashenfelter's Dip issue, in which workers endogenously select into training programs (Ashenfelter 1978). Zimmer et al. (2009) test, in locations that provided sufficient data, whether student trajectories in the year preceding a student's switch into charter schools are significantly different from trajectories in earlier years, and find no evidence that pre-transfer dips may be biasing estimates in San Diego and Philadelphia. Due to lack of necessary data, they are unable to test whether this is also the case for the other locations they study, and therefore again argue that fixed-effect estimates must be interpreted with caution.

A third issue with fixed-effect models is that they treat students who switch from a charter school to a traditional public school symmetrically to a student who switches from a traditional public school to a charter school. This is sometimes referred to as the "reversibility" assumption. The potential problem here is best seen by way of an example. Suppose that a specific charter school has a positive impact on students' achievement growth both during the years they are at the charter school *and* the years after they leave. If this occurred, then in the subsample of students who switched back to a traditional public school, it would tend to bias downward the estimated effect of attending a charter school. (Conversely, if a charter school had a negative impact on students' achievement growth before and after attending, then the estimated effect of attending a charter school would be biased upwards toward zero for the subsample who later switched back to a traditional public school.)

In short, none of the methods utilized in the papers included in our meta-analysis is entirely accurate. But they represent the best methods available, and are likely to come much closer to estimating the true causal impact of charter schools than the less rigorous studies that compare mean outcomes at one point in time.

7. Conversely, a temporary dip in performance of a student at a charter school may induce the student's family to switch the student to a traditional public school the next year, which would bias downward the estimated impact of the charter school.

3. Meta-Analysis of Effect Size

We use a statistical approach to reviewing the literature that is known as meta-analysis. Details appear in the appendix. The central idea is that in combining multiple estimated effects of attending a charter school, we need to take into account the uncertainty in each estimate. In general, we should give less weight to the more uncertain estimates. However, if the underlying estimates are sufficiently different from each other, we are likely to conclude that most of the variation across studies is real, and not due to the uncertainty in the individual estimates. In this case, we are likely to give a fairly equal weight to each study.

In our analyses, the weights given to each study are fairly equal, indicating that most of the variation we see in effects across studies are likely to be real, rather than due to uncertainty in the estimates. (Below we will report the I^2 statistic introduced by Higgins et al. (2003) which provides an estimate of the percentage of the variation in effect sizes that reflects true underlying variation. The estimates suggest that close to 100 percent of the variation across the studies reflects true variations in the effects of attending different charter schools.)

We began by obtaining estimates of charter school effects for each grade span and the main grade spans found often in the literature.

As in our previous study (Betts and Tang 2011), our main results in this section, in table 1, exclude the results for KIPP charter schools from both the middle school results and the results that combine elementary studies, elementary/middle studies, and middle school studies. (KIPP refers to the Knowledge is Power Program, a charter school operator. The KIPP estimates are often much larger than the estimates in studies that include all charter schools in a given region, and they would assume a disproportionate weight if we included them in the main analysis.) We later discuss the results when we add the KIPP studies into the analysis, and we also perform a meta-analysis of the KIPP studies themselves.

Tables 1 and 2 show the main results. Table 1 shows the results in terms of "effect sizes," that is, the predicted change in a student's achievement measured in terms of the number of standard deviations of achievement. Although this is the normal way of presenting results in education research, many readers may find it more understandable to read the results in terms of predicted changes in percentile rank for a student attending a charter school. Table 2 shows the results transformed into percentile rankings.[8] Below, we discuss the table 1 results for effect size in detail, and then briefly discuss how these estimated effects translate into percentiles.

8. The percentile ranking of a student indicates the number of students out of 100 that the student would score as highly as or higher than. For example a 99th percentile student scores as highly as or higher than 99 out of 100 students on average, while a 50th percentile student is in the middle of the achievement distribution.

Table 1. Effect Sizes and Significance From Meta-Analysis, by Grade Span and Subject Area

Grade Span	Reading Tests (# estimates-# locations), % true variation	Math Tests (# estimates-# locations), % true variation
Elementary	0.020 (17-15) 99.1%	0.045* (18-16), 99.2%
Middle	0.030 (18-15), 99.3%	0.084* (19-16), 99.5%
High	0.036 (15-12), 98.3%	0.051 (16-13), 99.3%
Combined Elementary/Middle	-0.001 (20-16), 98.8%	-0.002 (20-16), 99.7%
Elementary, Middle, and Combined Elementary/Middle	0.015 (50-26), 98.9%	0.044* (52-27), 99.2%
All	0.014 (21-17), 98.8%	0.034* (22-18), 99.6%

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less. The numbers in parentheses indicate the number of estimates included in the associated estimate of effect size and the number of locations. The percentage refers to the I^2 estimate of the percentage of the variation across estimates that reflect true variation in the effect of charter schools, rather than just statistical noise. Thus, for example, in the reading test result for elementary schools "(17-15), 99.1%" indicates 17 estimates covering 15 locations (with two studies each of New York City and San Diego schools), and that 99.1 percent of the variation across estimates in the literature may reflect true variation in the effect of charter schools. As mentioned in the text, we exclude a large number of studies of KIPP schools from the middle school tabulations as the number of studies greatly outweighs the share of these schools in the charter school population, while the effect sizes are also much larger than the average seen in other studies.

Table 2. Effect Sizes Expressed as Charter Students' Predicted Percentile After One Year, Starting at 50th Percentile, by Grade Span and Subject Area

Grade Span	Reading Tests	Math Tests
Elementary	50.8	51.8*
Middle	51.2	53.3*
High	51.4	52.0
Combined Elementary/Middle	50.0	49.9
Elementary, Middle, and Combined Elementary/Middle	50.6	51.8*
All	50.6	51.4*

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less. The numbers show the predicted test score percentile of a student who started at the 50th percentile, after one year of charter school attendance.

In table 1, results for each grade span for reading and math appear in the first and second columns respectively. For each grade span, the first number shows the estimated overall effect size. Effect sizes that are statistically significant (at the 5 percent level) are indicated with an asterisk. For elementary schools, the overall estimated effect sizes for reading and math achievement are 0.020 and 0.045, although only the latter estimate is significant at the 5 percent level. The corresponding estimates from Betts and Tang (2011) were 0.022 and 0.049 respectively, and both were significant. The drop in significance for reading derives from the new studies of Utah and Massachusetts, which unlike most other estimates were large and negative.

The second number for each grade span shows, in parentheses, the number of estimates contributing to the overall estimate, followed by the number of regions examined in the given studies. For example, in the meta-analysis of reading effects for elementary schools, "(17-15)" indicates that we found and used 17 separate estimates from 15 geographic areas in calculating the overall effect. Betts and Tang (2011) noted that the literature contains relatively few estimates of the impact on achievement of charter schools at either the elementary or high school level, although there are many estimates at the middle school level or at the combined elementary/middle school level. These important holes in the literature are starting to be filled in, in part due to more recent studies from CREDO, a research organization that now reports results separately by grade span.

The third number presented for each grade span shows an estimate of the percentage of the variation across estimates that reflects true variation in the impact of charter schools, as opposed to variation due to random noise. (This is the I^2 statistic referred to earlier.) For reading and math studies at the elementary level, we estimate that 99.1 percent and 99.2 percent of the variation reflects true variations in impact across studies. These are large percentages. Clearly, there appear to be important variations in charter school effects across studies and, implicitly, across geographic areas.

For middle schools, as for elementary schools, we find positive and significant effects of charter schools on math achievement, with a positive but insignificant effect on reading achievement.

The number of studies that focus specifically on charter high schools has grown over the last three years. As shown in the third row of table 1, no significant effect emerges overall in these studies.

A number of studies combine elementary and middle schools and, as shown in the fourth row of table 1, overall there is no significant effect of attending a charter school on reading or math achievement in these studies.

It is somewhat unusual to combine elementary and middle schools in this way. In a bid to find a representative portrait of the overall evidence on the impact of charter schools from studies of schools at the elementary, middle, and combined elementary/middle levels, the fifth row of table

1 combines all three of these study approaches. When pooling studies in this way, we find a positive overall estimated effect size for attending a charter school in these studies for both reading and math, but only the result for math is statistically significant.

Finally, some studies include test scores from elementary, middle, and high school grades together in one model. We refer to these as "All" grade span models. The sixth row of table 1 shows that the mean effect size in reading is not significant, but it is positive and significant for math.

In sum, none of the separate analyses by grade span shows a significant effect of attending a charter school in reading achievement, but in four of the six ways we combine studies by grade span, a positive and significant effect arises for math achievement. A second pattern is that in all cases, almost 100 percent of the variation across studies appears to be true variation.

Another way of gauging the size of the charter school effect sizes is to translate them into how a charter school student's academic ranking is predicted to change over time. As mentioned earlier, table 2 translates the effect sizes in table 1 into a student's predicted percentile after attending a charter school for one year, on the assumption that the student starts at the 50th percentile.

We do not discuss in any detail the percentiles predicted for reading because as we noted in table 1, the estimated charter school impacts on reading are not statistically significant. However, we note that the predicted gains in reading achievement from attending a charter school for one year are small, typically 1 percentile point or less.

The effects are bigger for math achievement. With an effect size of about 0.045 and 0.084 for math at the elementary and middle school levels, a student with median test scores—ranking 50th out of 100 students—would be predicted to move up to about the 52nd and 53rd percentiles, respectively, after attending a charter school for just one year. The math results from the analysis that combined elementary school studies, middle school studies, and combined elementary/middle school studies are identical to that for elementary school: The student is predicted to move up to the 52nd percentile. Finally, the studies that combine all grades predict a gain to the 51st percentile.

Consider the largest effect size, for middle school math, one more time. If the student began the year ranking 50th out of 100, after one year at a charter middle school he or she is predicted to rise to tie or outrank 53 out of 100 students. This is a meaningful change, and over several years of such gains, a student's gains could be quite large. For example, if a student experienced the same gains during all three years of middle school, he or she would move from the 50th percentile to just below the 60th percentile.

Another way of thinking about the estimated effects is to compare them to the estimated impact of other common educational interventions. The effect sizes for math range between 0.03 and

0.08, signifying that after one year of attending charter school a student's test score would increase, relative to those of other students, by 3 to 8 percent of one standard deviation. Clotfelter, Ladd, and Vigdor (2007) estimate that in North Carolina reducing class size by five students is associated with gains in achievement of 1.0 to 1.5 percent of a standard deviation.

Comparing the Overall Results to Those in Our Earlier Review: Reading Effects no Longer Significant, Math Effects Larger and More Significant

In just three years, the number of rigorous studies of charter school impacts on achievement has grown considerably. For example, Betts and Tang (2011, table 2) analyzed 59 separate math achievement effects, while in table 1 in the current report we analyze 95 estimates. [9]

Overall effect sizes for reading achievement in the current report are roughly the same (but higher for four grade spans, and lower for two). The only consequential change for reading achievement is that our overall estimate of the impact for reading achievement for elementary schools is no longer significant at the 5 percent level.

In contrast, both significance levels and effect sizes for the impact of charter schools on math achievement are generally larger in the current study than when we applied the same approach to the studies available for our 2011 report. Studies that combine all three grade spans now show a positive significant effect, but did not in Betts and Tang (2011). Apart from a slight drop in the effect size at the elementary school level (from 0.049 to 0.045), all of the other math effect sizes in table 1 are larger than the corresponding estimates in Betts and Tang (2011). For example, consider the fifth row of table 1, in which we combine all studies performed at the elementary level, the middle school level, or at the combined elementary/middle school level. Our estimated effect size is 0.044. In comparison, our corresponding estimate in Betts and Tang (2011) was only 0.020. (We include 52 studies in our current analysis for these grade spans compared to only 33 in our earlier study.)

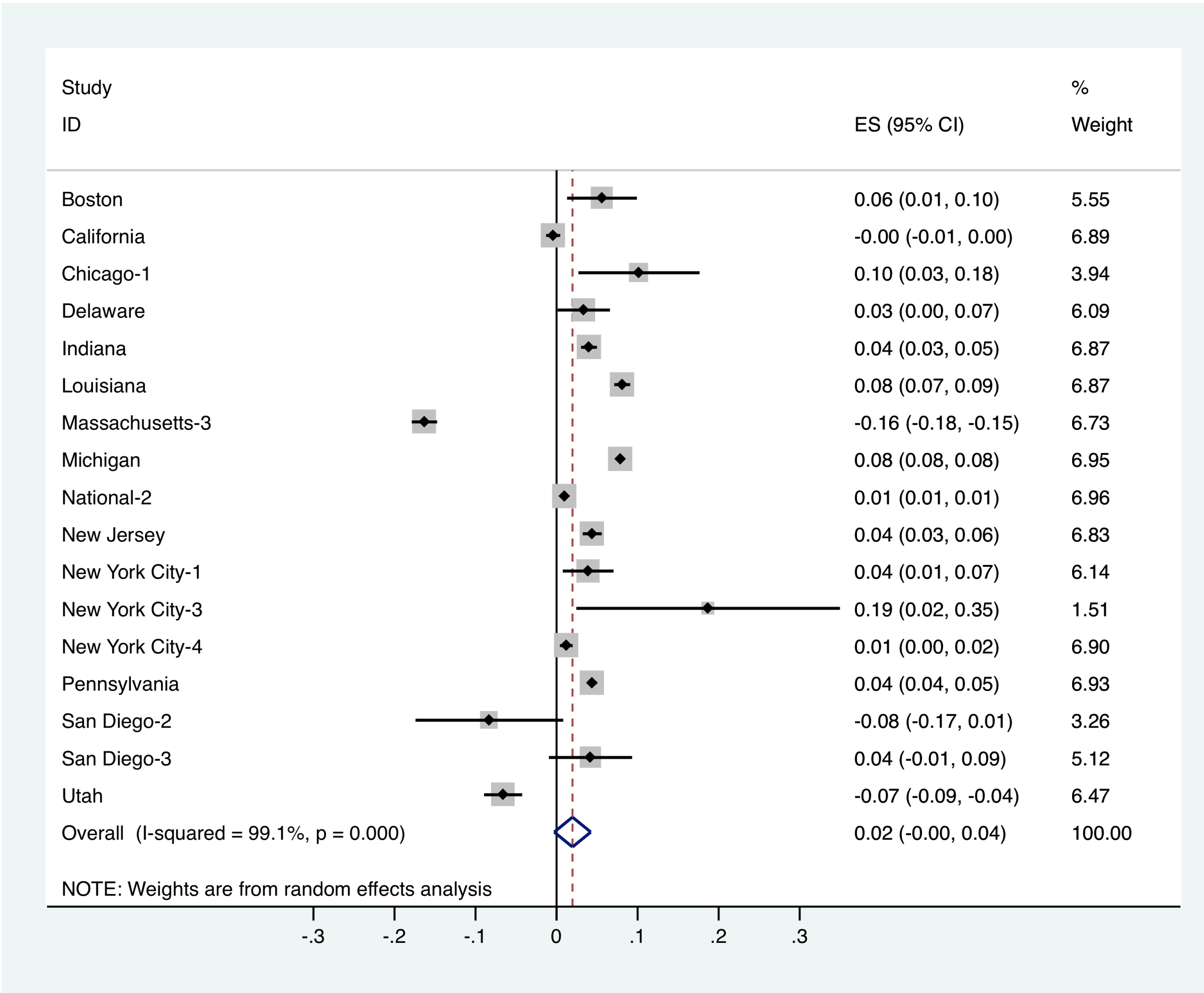
Differences in Effect Sizes Across Studies

It is useful to look at the effect sizes of individual studies and how they contribute to the overall estimates shown in table 1. Figures 1 and 2 provide an illustration of the variation in the effect sizes across studies of elementary schools for reading and math respectively. The figures use horizontal lines to indicate the 95 percent confidence interval for each estimate. The rightmost column shows the weight attributed to each study. (The size of each square is proportional to these weights.) The diamond at the bottom of each figure illustrates the overall estimated effect size, with the width of the diamond indicating the 95 percent confidence interval.

---

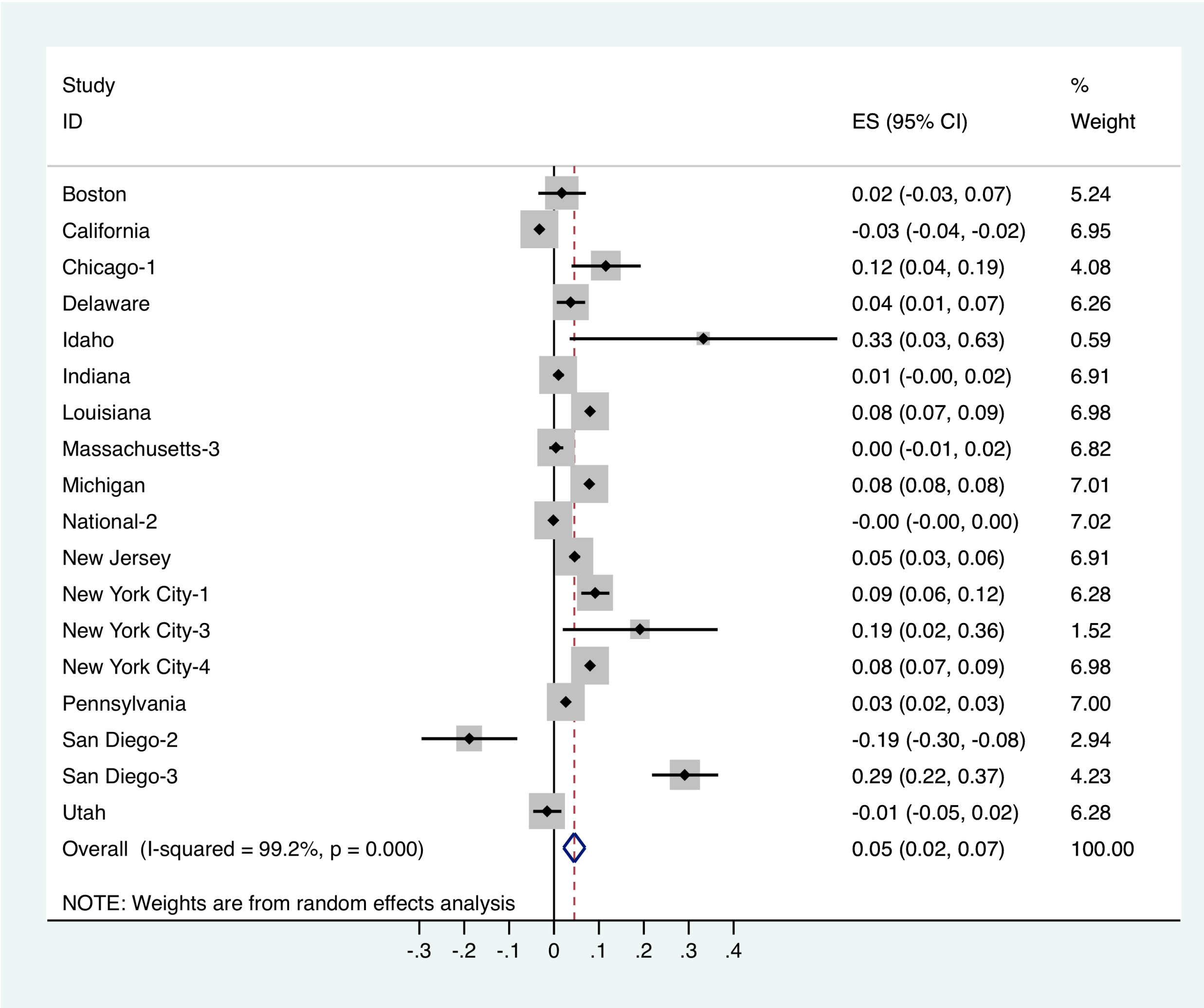
9 This corresponds to the total number of studies of elementary, middle, high, combined elementary/middle school, and all grade studies.

Figure 1. Elementary School Reading Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label.

Figure 2. Elementary School Math Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label.

The elementary school studies with the largest estimated reading effect size include studies of New York City, Boston, Michigan, Louisiana, and Chicago. Two studies show negative and significant results: a study by Ni and Rorrer of Utah (2012), and a CREDO (2013a) study of Massachusetts. A third study with a large negative (but in this case not quite significant) coefficient is a study of San Diego charters (Betts et al. 2005). A study of San Diego by Betts, Tang, and Zau (2010) using the same statistical approach but a later time frame produced a positive and, again, nearly significant coefficient. In math, the studies with the largest positive effect sizes for elementary charter schools were in Idaho, San Diego, New York City, and Chicago. (Again, a study of an earlier period in San Diego produced a negative and this time significant counterpoint. It seems likely that San Diego's charter schools have become more effective with regard to math and reading achievement over time.)

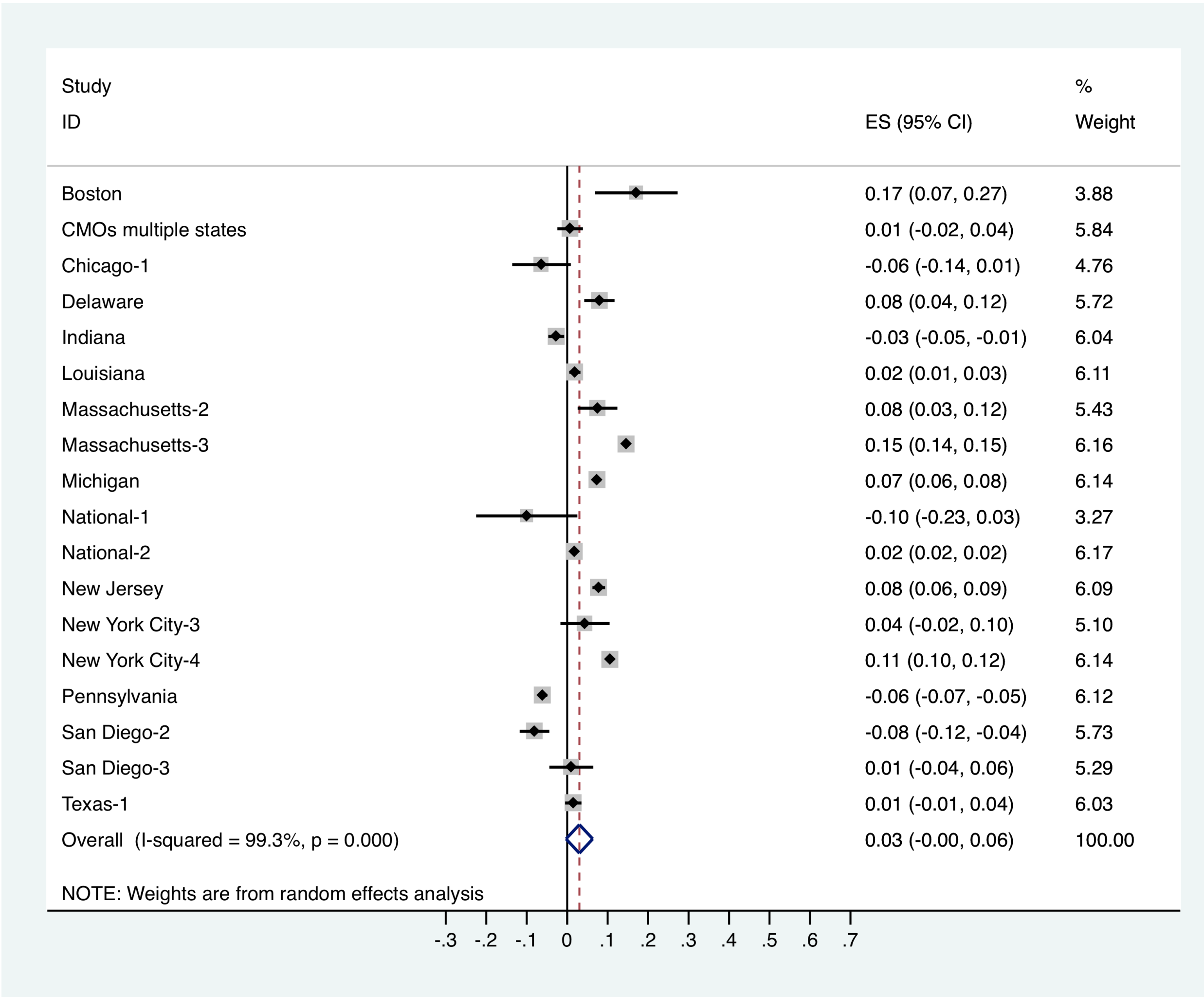
The bottom of the left-hand column in the figures reproduces the I^2 statistic along with the p value of a test for homogeneous effects across studies. The p values are essentially zero, which is what we typically found in our analyses of other samples. Put simply, this result indicates that the idea that charter schools have the same impact in all geographic areas is wrong.

The right-hand column in the figures shows the weights assigned to each study when obtaining our overall estimated effect size. The statistical method uses the variation in effect sizes across studies that is above and beyond the mean estimated variances of the individual estimates to calculate the true underlying variance in effect sizes that reflects true variation. Smaller, less precise estimates get less weight than larger, more precise estimates, but because most of the variation is estimated to be "true," for the most part there is not much difference in the weight assigned to the various studies.[10]

Figures 3 and 4 show the estimated effects in middle school studies for reading and math respectively. For reading, estimates lie in a fairly narrow band centered at just above zero, with roughly two-thirds of estimates being positive. Positive results from Boston and Massachusetts exhibit the largest effect size in these studies. Figure 4 shows that most studies of math achievement produced positive effect sizes, often statistically significant. Again the biggest outlier is the result from Boston, with an effect size about double the size of the next biggest estimate (from New York City).

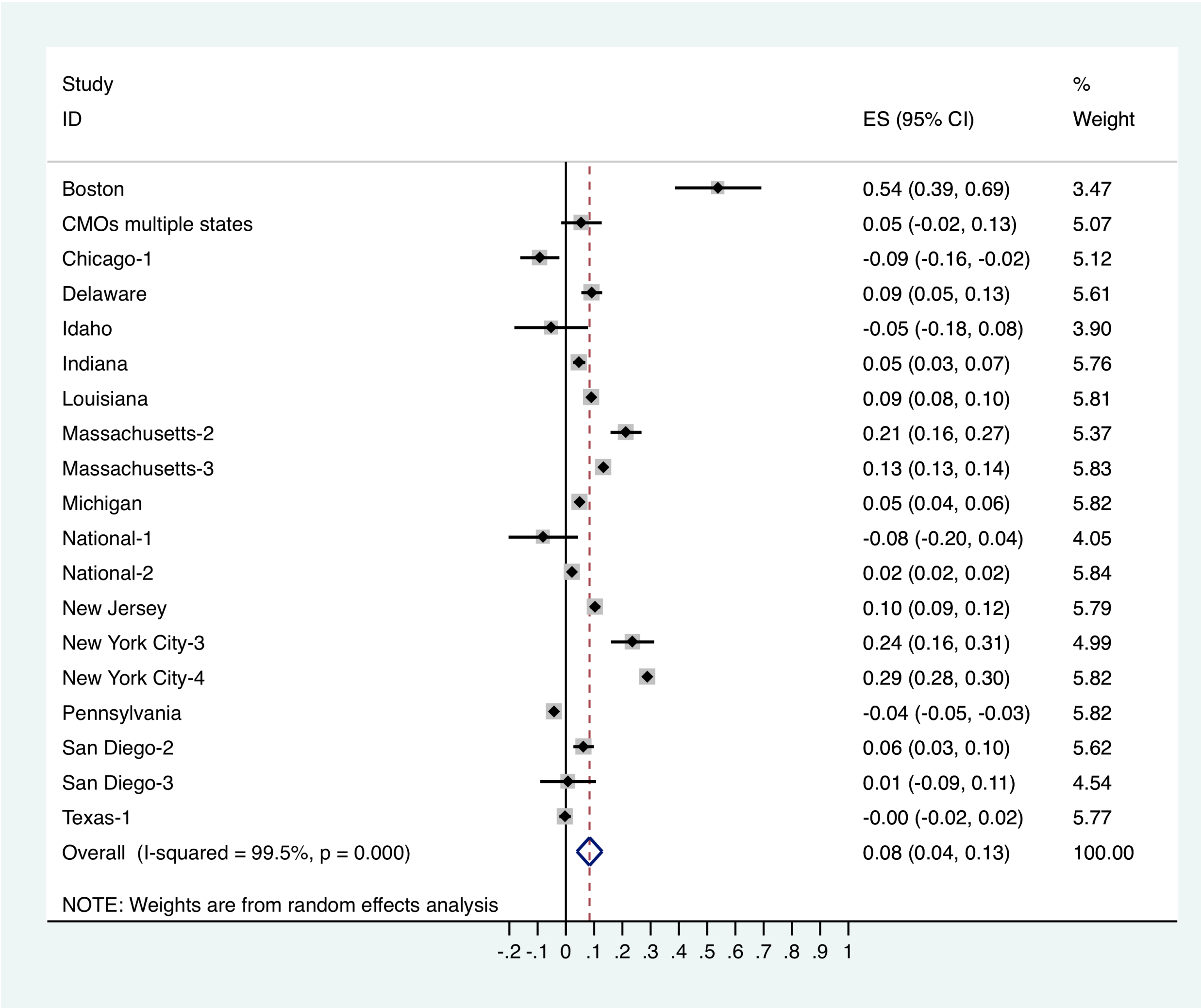
10. The weighting scheme here is optimal in that it produces the minimum variance estimate of the overall effect.

Figure 3. Middle School Reading Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label.

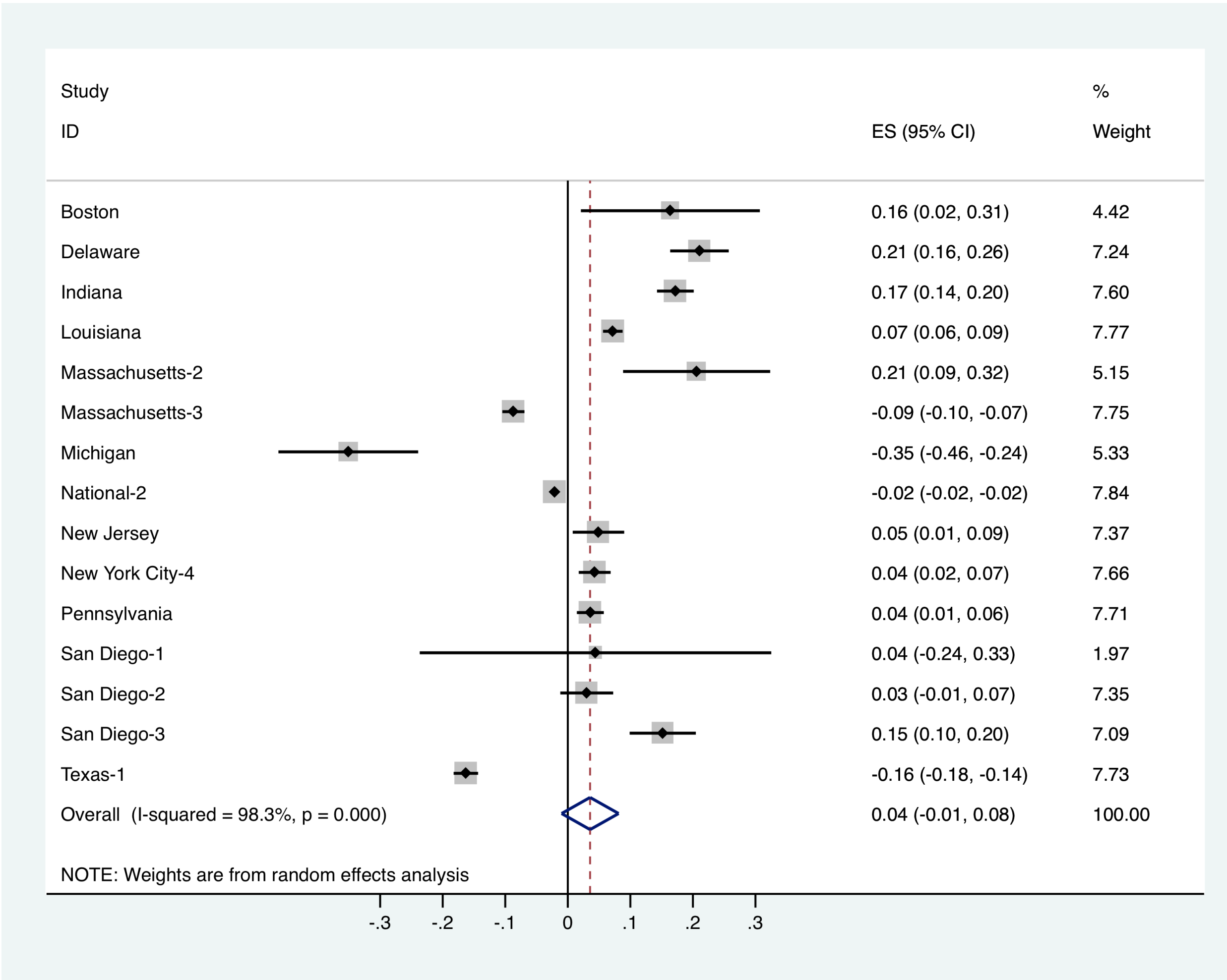
Figure 4. Middle School Math Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label.

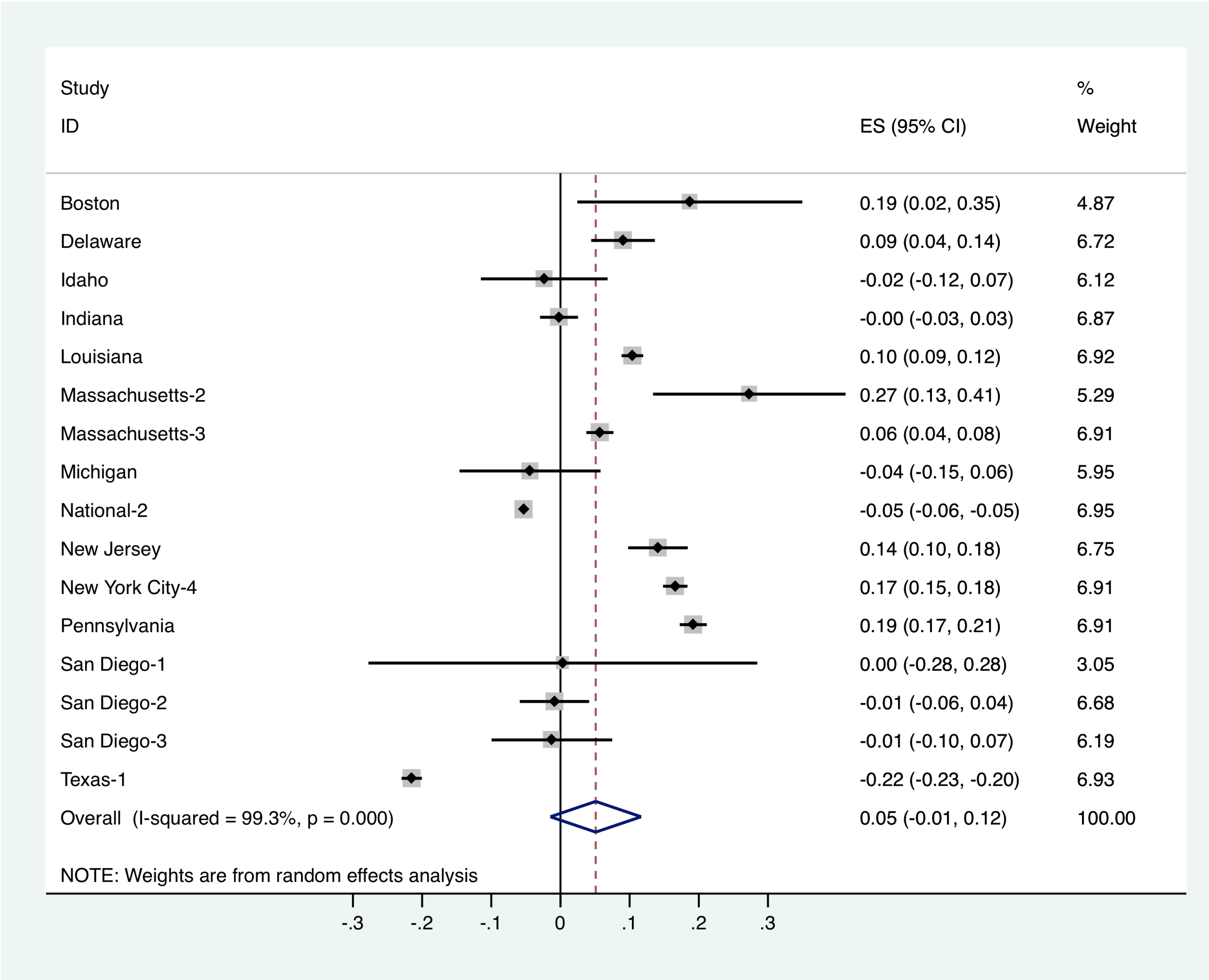
Figures 5 and 6 show corresponding figures for high school results. Although our overall estimates are insignificantly different from zero, there are a number of individual studies that find statistically significant positive and negative effects of attending a charter school. The overall estimated effect size for reading is positive, but three large negative estimates (from Michigan, Texas, and Massachusetts) counteract a large number of small positive effect size estimates from other studies. A similar pattern emerges for math achievement in figure 6, with far more positive and significant results than negative and significant results, but the overall estimated effect size of 0.05 is not statistically significant.

Figure 5. High School Reading Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label.

Figure 6. High School Math Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label.

Figures 7 and 8 show the results from the studies that combine elementary and middle schools, for which overall we find no significant effects. Considerable variation emerges for reading in figure 7, with a study of Washington, D.C., schools by Nichols and Özek (2010), a study of the Promise Academy in the Harlem Children's Zone in New York City by Dobbie and Fryer (2010), and another of Texas showing the largest positive effects, and studies of North Carolina, Ohio, and Texas showing the largest negative effects. It is interesting that Texas produces among the largest positive and largest negative effect sizes for reading achievement. The positive estimate comes from a fixed-effect model by Booker et al. (2004), covering the period 1995 to

2002. The estimates apply only for the subsample of charter schools that are two or more years old, and for students who did not switch schools in the current school year. The negative estimate comes from a fixed-effect estimate by Zimmer et al. (2009) covering the period 1996 through 2004, but does not distinguish between new charters and established charters, nor between students in their first year at a charter school and in later years. Zimmer et al. (2009) argue that because "newness is … an inherent part of the charter treatment," it is the latter number that is more representative of the performance of Texas charter schools.[11, 12]

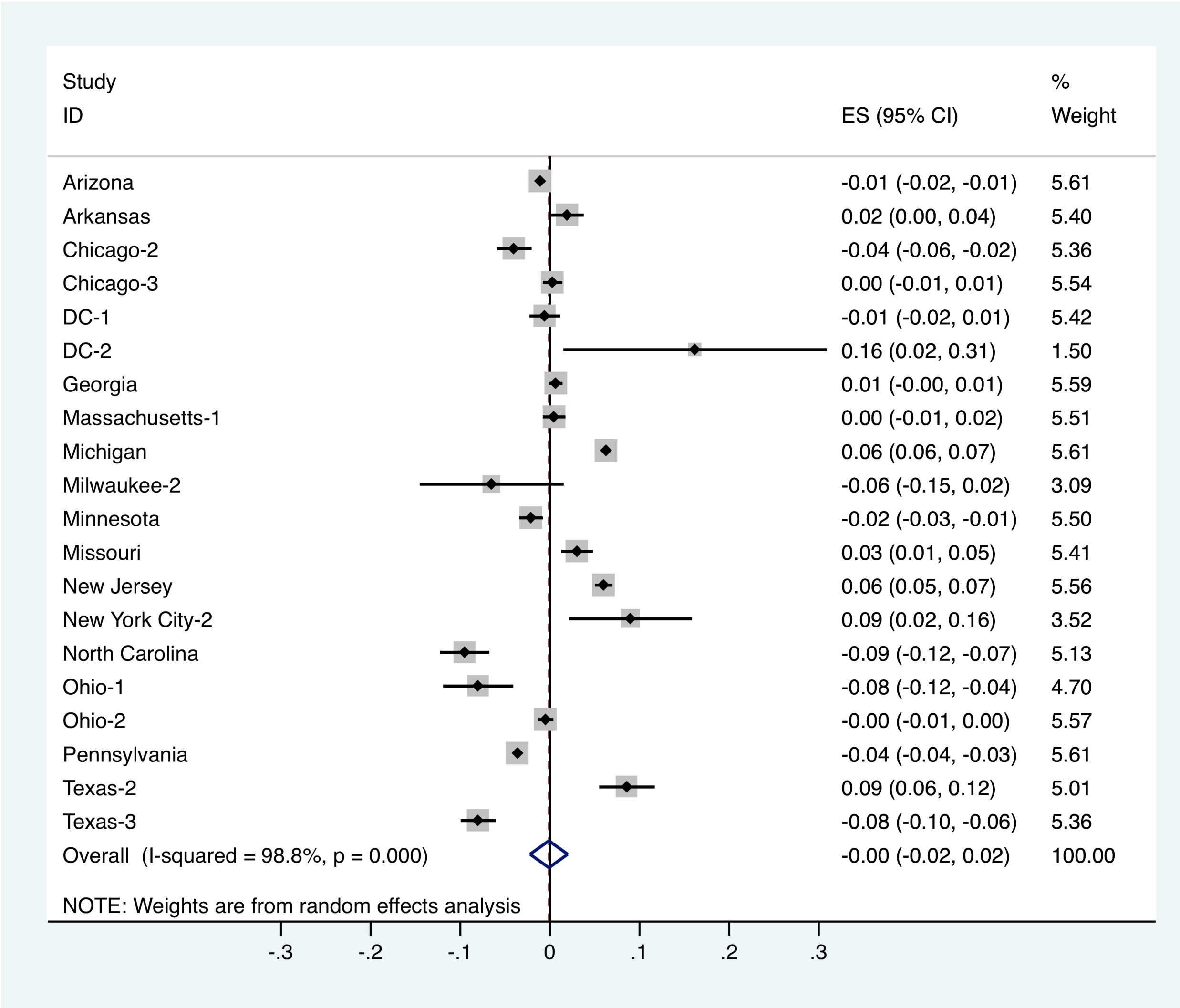
Figure 8 shows estimates for math achievement from studies that combine elementary and middle schools. Again, the insignificant overall estimate masks considerable variation. The studies with the largest estimated positive effects come from Washington, D.C., New York City, and Texas. The largest estimated negative effects come from studies in Ohio, North Carolina, and Texas. (The same pair of Texas studies that produces the contradictory estimates outlined above for reading also produces the quite large contradictory results for math.)

---

11. In this situation, we include both estimates because throughout this analysis we attempt to be as inclusive as is reasonable and methodologically sound. It may be the case that in some locations, strict governance quickly closes failing new schools, so that over time only experienced schools will continue to exist. It may also be the case that in other locations with looser governance biased towards experimentation, there will always be new schools emerging which are given freedom to struggle through potential "growing pains." Therefore, we think it is reasonable to include both estimates that include a significant portion of "new" schools and estimates that cover only experienced schools. In other cases where multiple authors study the same geographic location, we also choose to include all estimates, because choosing one estimate over others necessarily requires some value judgment over which estimate is the most reflective of the true effect of charter schools. Including all estimates judged to be methodologically sound seems to us to be the most transparent way of proceeding.

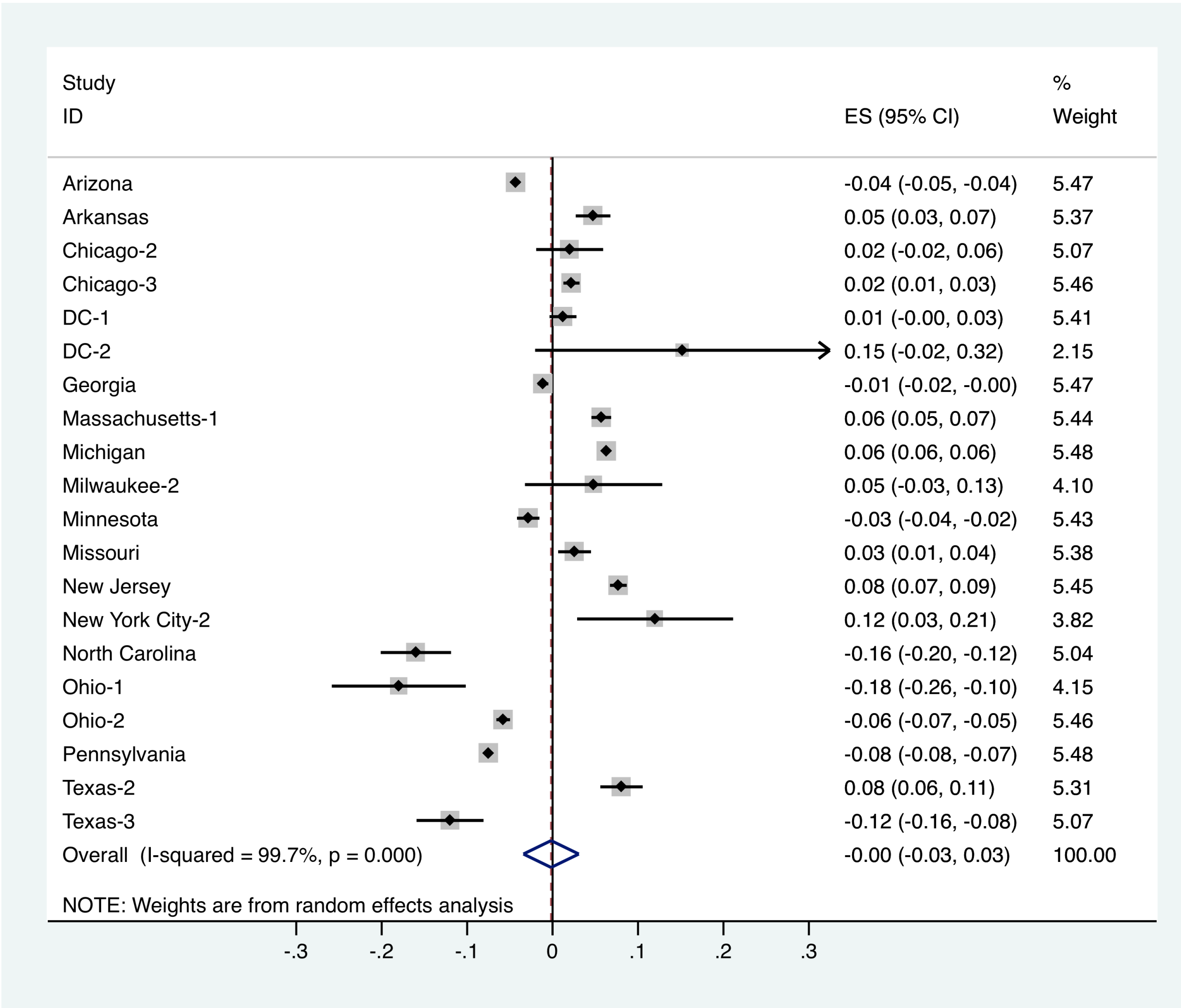
12. Zimmer et al. (2009) study newer and established schools separately and demonstrate that charter schools in most locations improve over time (i.e., the estimates of charter schools that are three or more years old are higher than estimate of charter schools that are younger). Charter schools either improve over time, or the less successful charter schools close quickly, or potentially both situations occur. They further note that of the locations they study, Texas is one of the states in which charter schools experience the most improvement over time (i.e., that has the most negative first-year charter school effects).

Figure 7. Reading Effect Sizes for Studies That Combine Elementary and Middle Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label. The studies in this figure are referred to in table 1 as "Combined Elementary/Middle Studies."

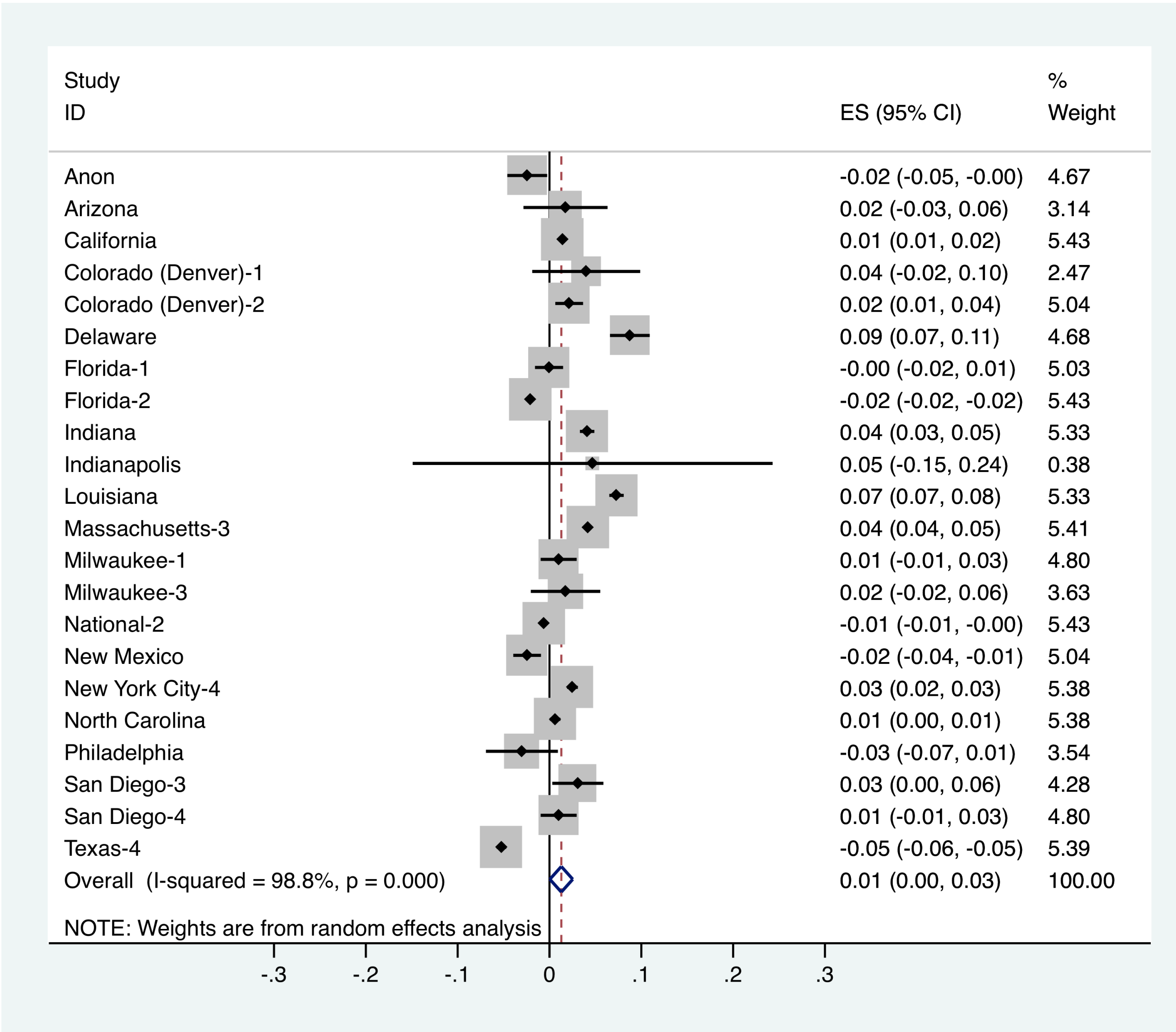
Figure 8. Math Effect Sizes for Studies That Combine Elementary and Middle Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label. The studies in this figure are referred to in table 1 as "Combined Elementary/Middle Studies."

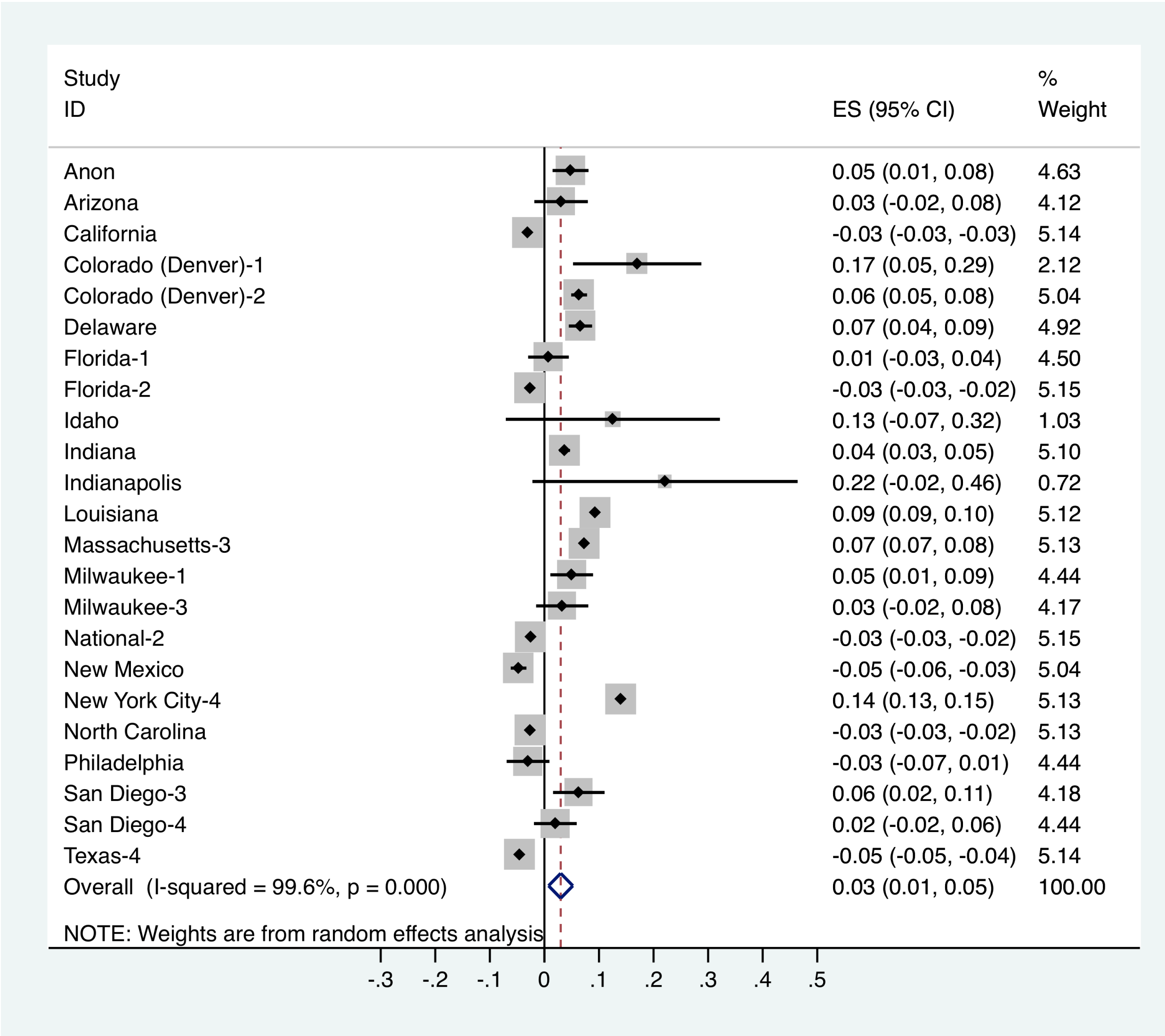
Figures 9 and 10 show reading and math results for the “All” grade span studies, which in the case of reading produced an overall effect size that was insignificantly different from zero. For reading, as shown in figure 9, most of the effect sizes are clustered in a narrow band on either side of zero. The main exceptions are positive effect sizes of 0.09 found for Delaware by Miron et al. (2007), and of 0.07 found for Louisiana by CREDO (2013d). For math, as shown in figure 10, the overall estimate is positive and significant at the 5 percent level. There are four large positive effect size estimates, for New York City, Indianapolis, Denver, and Idaho, but the latter three of these receive a small weight in the overall estimate because they are estimated quite imprecisely compared to the other studies that mostly have effect sizes near zero.

Figure 9. Reading Effect Sizes for Studies That Combine Elementary, Middle, and High Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label. The studies in this figure are referred to in table 1 as "All" grade spans.

Figure 10. Math Effect Sizes for Studies That Combine Elementary, Middle, and High Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



Notes: The horizontal lines show the 95 percent confidence interval, which is also indicated in the second column from the right. The rightmost column shows the weight ascribed to each study, with the size of the square proportional to these weights. The overall effect size estimate is shown at the bottom. Geographic locations with estimates from multiple studies have unique numbers appended to their labels to distinguish between studies. Table 11, in the appendix, indicates the author and year of the study referenced by each study ID label. The studies in this figure are referred to in table 1 as “All” grade spans.

Estimated Effects for KIPP Middle Schools Are Far Higher Than for Other Charter Middle Schools

The middle school results presented in table 1 and in figures 3 and 4 exclude the many estimates for individual KIPP schools. Table 3 shows the middle school meta-analysis when the KIPP studies are added back in. The reading and math effects are much more positive and both are statistically significant. However, these are not representative estimates of charter schools nationwide. For instance, about 25 percent of the weight in the middle school math meta-analysis in table 3 goes to the studies of individual KIPP schools. Yet our estimates suggest that nationwide KIPP schools account for only about 2 percent of all charter schools.

The bottom panel of table 3 shows the results of a meta-analysis that includes only the KIPP schools. This can be thought of as the second meta-analysis of the KIPP literature, following up on the similar analysis in Betts and Tang (2011). KIPP schools appear to have a statistically significant and positive influence on both reading and math achievement, with the effect size for math being twice as large as for reading.[13]

13. The effect sizes for KIPP schools, shown in the bottom panel of table 3, of 0.174 and 0.374 for reading and math respectively, are materially higher than the elasticities estimated by Betts and Tang (2011), which were 0.096 and 0.223 respectively. Since our earlier literature review, the preliminary report from a national study (Tuttle et al. 2010) has been replaced by a final report (Tuttle et al. 2013), and we have substituted the single effect size by subject in the later report for the many school-level estimates in the earlier report on the advice of one of the authors (Brian Gill, personal communication, February 2014).

Table 3. Results With and Without the KIPP School Estimates: Effect Sizes and Significance From Meta-Analysis, by Grade Span and Subject Area

Grade Span	Reading Tests (# estimates-# locations), % true variation	Math Tests (# estimates-# locations), % true variation
Including KIPP Schools		
Middle	0.054* (26-20), 99.0%	0.145* (27-21), 99.4%
Elementary, Middle, and Combined Elementary/Middle	0.025* (58-31), 98.7%	0.070* (60-32), 98.6%
Including Only KIPP Estimates		
Middle	0.174* (8-5), 85.2%	0.374* (8-5), 94.2%

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less. The numbers in parentheses indicate the number of estimates included in the associated estimate of effect size, and the number of locales, which in the case of KIPP schools is unknown due to the shielding of charter school identities in one study.

The estimates for KIPP middle schools are far higher than our average estimates in table 1, with estimated effect sizes for reading and math of 0.174 and 0.374 respectively. These effect sizes are enough to move a student initially at the 50th percentile to percentiles 56.9 and 64.6 in a single year of attendance at a KIPP school. These are very large effects, by any standard.

Our Results Are Not Sensitive to Inclusion of the CREDO Studies

Just as the KIPP studies would dominate the middle school analysis had they been included in table 1, the CREDO studies of individual states and cities constitute impressive proportions of the studies in each grade span. For instance, CREDO contributed 12 of the 20 studies that pool elementary and middle schools together. One can also examine the results once the many studies published by CREDO are removed. A concern about these CREDO estimates is that they all use the same nonexperimental method, which hinges upon how successfully the studies matched charter school students with counterparts at traditional public schools. Because many charter school students were matched with non-charter students using their characteristics and test scores once at the charter school, this could bias the results. It is not certain whether charter school effects would be over- or understated. But it is likely that charter effects, if anything, would be biased toward zero.[14] Table 4 shows the results when the grade span estimates shown in table 1 are repeated without the CREDO studies.

Table 4. Results When CREDO Studies Are Excluded: Effect Sizes and Significance From Meta-Analysis, by Grade Span and Subject Area

Grade Span	Reading Tests (# estimates-# locations), % true variation	Math Tests (# estimates-# locations), % true variation
Elementary	0.020 (9-7), 87.7%	0.059* (10-8), 94.6%
Middle	0.017 (10-9), 85.3%	0.082* (11-10), 93.4%
High	0.091 (7-5), 98.3%	0.032 (8-6), 97.5%
Combined Elementary/Middle	-0.016 (8-7), 94.2%	-0.013 (8-7), 95.8%
Elementary, Middle, and Combined Elementary/Middle	0.008 (27-15), 91.4%	0.046* (29-16), 94.7%
All	0.016 (11-9), 85.1%	0.039* (12-10), 64.6%

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less. The numbers in parentheses indicate the number of estimates included in the associated estimate of effect size, and the number of locales. For comparability with table 1, we also exclude the KIPP studies from the middle school category, the combined elementary/middle category and the combined elementary, middle, and combined elementary/middle school studies.

14. For instance, suppose a given charter school boosts test scores. A student at this charter school would be matched with a student who had similar achievement in spite of attending a relatively underperforming traditional public school. The latter student may have been on a steeper learning trajectory, holding constant other factors. This would tend to make the charter school student's achievement look worse. Conversely, at a charter school that was truly underperforming, a student could be matched with a student at a traditional public school who was in truth a weaker student. This would tend to lead to a less negative estimated charter effect than the true effect. Note that a separate criticism of CREDO's statistical approach, by Hoxby (2009), was rebutted by CREDO in CREDO (2009b).

The results with and without the CREDO studies are surprisingly similar. For reading, the same pattern, in which the overall effect size is not significantly different from zero, continues to hold across all grade spans studied. For math, exactly the same grade spans that showed positive and significant effects in table 1 continue to show positive and significant effects when the CREDO studies are dropped, and the magnitude of the effect sizes are quite similar.

Results for At-Risk Populations

Table 5 shows estimated effect sizes from meta-analyses for three at-risk populations: students in special education, English Language Learners (ELLs), and students eligible for federal meal assistance, the last of which is a commonly used proxy for poverty. Multiple studies for students in these three groups exist at only two grade spans, combined elementary/middle schools and all grades combined. However, for continuity with earlier tables we show the former twice, listed first as "Combined E/M" and then "E, M and Combined E/M."

Table 5. Effect Sizes for Studies of Selected Subsamples of Student Populations and Significance From Meta-Analysis, by Grade Span and Subject Area

	Grade Span		
Student Population	**Combined E/M (# estimates-# locations), % true variation**	**E, M and Combined E/M (# estimates-# locations), % true variation**	**All (# estimates-# locations), % true variation**
	READING TESTS		
Students in Special Education	-0.002 (12-12), 79.9%	-0.002 (12-12), 79.9%	0.025* (10-10), 82.7%
English Language Learners	0.005 (12-12), 73.5%	0.005 (12-12), 73.5%	0.032 (10-10), 87.7%
Students Eligible for Federal Meal Assistance	0.014 (12-12), 89.1%	0.014 (12-12), 89.1%	0.028* (10-10), 93.8%
	MATH TESTS		
Students in Special Education	0.002 (12-12), 79.4%	0.002 (12-12), 79.4%	0.017* (10-10), 0.0%
English Language Learners	0.027 (12-12), 81.3%	0.027 (12-12), 81.3%	0.015 (10-10), 58.6%
Students Eligible for Federal Meal Assistance	0.020* (12-12), 90.0%	0.020* (12-12), 90.0%	0.022* (10-10), 93.1%

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less.

Results for the three types of at-risk populations are mixed, and difficult to summarize simply. No subpopulation appears to do worse when attending charter schools, but beyond that the impact of charter schools varies by subject and grade span. ELLs show no significant differences for either subject for any of the available grade spans. Students in special education attending the charter schools included in the reviewed studies do as well as or, in the studies that pool all grades, better than their counterparts in district-run public schools in both math and reading.

For students eligible for meal assistance, results are consistently positive for math.

Results for Racial/Ethnic Subgroups

Table 6 shows results from grade spans with multiple studies by race/ethnicity.[15] Interestingly, results for white students are negative and significant for both reading and math in all three grade spans for which results are available, except for math in studies that combined all grades, for which no significant effect emerges. Similarly, for Asian students, overall effect sizes are negative and statistically significant, with the exception of reading achievement in studies that combined elementary and middle schools, where no significant effect emerges.

15. The table does not separately show results for studies of E, M, and H grade spans as there are only one or two studies at these grade spans, and none of these are new since our 2011 literature review. Interested readers can find a summary of those results in table 6 of Betts and Tang (2011).

Table 6. Effect Sizes for Studies of Racial/Ethnic Subsamples of Student Populations and Significance From Meta-Analysis, by Grade Span and Subject Area

	Grade Span		
Race/Ethnicity	**Combined E/M (# estimates-# locations), % true variation**	**E, M and Combined E/M (# estimates-# locations), % true variation**	**All (# estimates-# locations), % true variation**
	READING TESTS		
White	-0.032* (15-13), 97.9%	-0.037* (17-14), 97.6%	-0.020* (14-13), 97.5%
Black	0.023 (16-13), 94.7%	0.020 (19-15), 93.8%	0.006 (14-13), 97.2%
Hispanic	-0.024 (16-13), 91.9%	-0.025 (19-15), 90.9%	-0.001 (14-13), 95.7%
Native American	-0.142 (9-9), 95.4%	-0.142 (9-9), 95.4%	-0.054 (9-9), 45.8%
Asian	-0.026 (12-12), 59.4%	-0.033* (14-13), 61.1%	-0.051* (10-10), 95%
	MATH TESTS		
White	-0.057* (15-13), 99.2%	-0.058* (17-14), 99.1%	-0.012 (14-13), 98.7%
Black	0.025 (16-13), 96.9%	0.028* (19-15), 96.4%	0.024 (14-13), 98.6%
Hispanic	-0.004 (16-13), 95.2%	-0.002 (19-15), 94.9%	0.019 (14-13), 98%
Native American	-0.034 (7-7), 67.6%	-0.034 (7-7), 67.6%	-0.077* (9-9), 57.9%
Asian	-0.46* (12-12), 78.3%	-0.058* (14-13), 77.8%	-0.037* (10-10), 64.2%

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less.

The results are quite different for black students. The effect sizes for black students are always positive but only significant for math when we combined all studies that included elementary and middle grades.[16]

Effect sizes for Hispanic students are negative in five out of six cases but are never significant, and are often very small. Effect sizes for Native Americans are generally negative, but are significant only for math achievement in studies that combine all grades.

Although the difference in results between black students on the one hand and white and Asian students on the other is quite striking, these results must be interpreted with caution for two reasons. First, the I^2 statistic, which estimates the percentage of the variation across studies that is real rather than statistical noise, is very high. For instance, the overall effect size in math for white students in studies that combine elementary and middle schools was -0.057, but 99.2 percent of the variation across the 15 contributing studies is likely to be real. This variation is not trivial: Nine studies showed negative and significant effects, two showed positive and significant effects, and four showed insignificant effects. The statistically significant effects ranged from a low of about -0.15 in Chicago and Pennsylvania to a high of about 0.09 in Missouri. Second, even if there had been uniformity of results by race across different studies, it does not necessarily mean that racial/ethnic group would experience the same impact in any charter school besides those that have already been studied by researchers.

Urban Districts and Schools

Table 7 shows the results when we focus on studies of urban districts or on individual schools in urban areas. The math effect size estimates are always higher in the urban subsample shown in Table 7 than in the overall sample shown in Table 2. In the case of reading, there are two cases (elementary and high schools) in which the effect sizes are higher in the urban subsample. Insignificant results in the overall sample (Table 2) become positive and significant in the urban school subsample in three cases: reading achievement in elementary schools and high schools, and math achievement in studies that combine elementary and middle schools.

16. (But the other three effect sizes at the E/M and “Combined E/M and E and M” levels for math and reading are weakly significant, at the 7- to 10-percent levels.)

Table 7. Effect Sizes for Studies of Urban Districts and Schools, by Grade Span and Subject Area

Grade Span	Reading Tests (# estimates-# locations), % true variation	Math Tests (# estimates-# locations), % true variation
Elementary	0.037* (7-4), 71.7%	0.085* (7-4), 90.7%
Middle	0.029 (6-4), 95.9%	0.167* (6-4), 98.2%
High	0.081* (5-3), 76.7%	0.070 (5-3), 92.7%
Combined Elementary/Middle	-0.002 (6-4), 82.6%	0.023* (6-4), 40.6%)
Elementary, Middle, and Combined Elementary/Middle	0.025 (19-6), 95.7%	0.100* (19-6), 99.1%
All	0.011 (10-7), 68.0%	0.062* (10-7), 96.2%

Note: Asterisks indicate effect size significantly different from zero at the 5 percent level or less.

Significant and positive charter school effects for urban students in particular have been noted in Angrist et al. (2013) and Gleason et al. (2010). There could be multiple reasons for the larger effects in urban settings. One obvious possibility is that charter schools have more value to add in large urban districts if the traditional schools in these areas are underserving their students to a greater extent than are their nonurban counterparts. Angrist et al. (2013) attributes the success of urban charter schools in Massachusetts to the "No Excuses" approach to education, which the authors describe as emphasizing "discipline and comportment, traditional reading and math skills, instruction time, and selective teacher hiring."

4. Histograms and Vote-Counting Analysis

We next show histograms of the effect sizes, to give a fuller picture of the distribution of effect sizes. We also show vote-counting results. Subject to the earlier warning that one cannot assume that a large set of insignificant effects together implies that the overall effect is insignificant, the vote-counting procedure provides a somewhat more transparent window onto the extent to which the literature produces heterogeneous results.

The histograms support the results of the overall findings and additionally offer a view of the distribution of effects. In the previous section we demonstrated that on average charter schools are serving students well, particularly in math. However, a positive overall effect may be the result of a few large positive results that obscure many more small negative results. Similarly, a moderate overall result may be the result of many small positive results negated by a few large negative results. Examining histograms allows us to consider the entire range of effects found across studies. We can use these pictures to pinpoint the upper and lower bounds of the effects found in each grade span.

The histograms present the percentage of studies finding effect sizes in each 0.05 unit range between effect sizes of -0.6 and 0.6. We created histograms for each grade span separately in order to examine the different effects according to the grade levels of the students studied.[17]

The picture at the elementary school level is in line with the formal meta-analysis results. Confirming the overall positive effects for both reading and math (which are significant for math, though not statistically significant in reading), we see that the effect sizes are more often greater than zero than less than zero in both subjects at the elementary school level. Substantially more studies report positive effects than negative effects in both subjects. Specifically, the first row (for elementary) of table 8 and table 9, column (1) shows that for reading 77 percent of unweighted estimates are positive, while for math 78 percent of unweighted estimates are positive. These percentages are very similar to the weighted percentages.

17. We generated both unweighted histograms, in which all studies receive equal weight regardless of whether it is a study of a single school or whether it is a study of an entire state, as well as weighted histograms, in which the weight used is the variable weight applied in the formal meta-analysis. The weighted histograms appear generally similar to the unweighted estimates, because the sample variance of individual studies was small relative to the true underlying variation in the effect sizes. Because the unweighted and weighted histograms look similar, we present only the weighted histograms. The weighted histograms incorporate information about the precision of the estimates by down-weighting estimates with large standard errors.

Table 8. Percentage of Reading Results by Level of Statistical Significance and by Method of Weighting Studies

	Sign and Significance	(1)	(2)
Gradespan (total # of studies)		Unweighted Excluding KIPP Percentage of Studies (# of studies)	Weighted by varying weights from formal meta-analysis Excluding KIPP Percentage of Studies
Elementary (17 studies)	-/Significant -/Insignificant +/Insignificant +/Significant	12% (2) 12% (2) 6% (1) 71% (12)	13 10 5 72
Middle (18 studies)	-/Significant -/Insignificant +/Insignificant +/Significant	17% (3) 11% (2) 22% (4) 50% (9)	18 8 22 52
High (15 studies)	-/Significant -/Insignificant +/Insignificant +/Significant	27% (4) 0% (0) 13% (2) 60% (9)	29 0 9 62
Combined Elementary/Middle (20 studies)	-/Significant -/Insignificant +/Insignificant +/Significant	35% (7) 15% (3) 15% (3) 35% (7)	37 14 17 32
Elementary, Middle, and Combined Elementary/Middle (50 studies)	-/Significant -/Insignificant +/Insignificant +/Significant	22% (11) 14% (7) 16% (8) 48% (24)	24 12 16 47
Studies of All Grades (21 studies)	-/Significant -/Insignificant +/Insignificant +/Significant	19% (4) 10% (2) 29% (6) 43% (9)	21 9 21 48

Note: Each number indicates the percentage of regression results for the given weighting method and combination of grade spans that fit the stated category of sign and statistical significance. The numbers within each cell may not sum to 100 due to rounding.

Table 9. Percentage of Math Results by Level of Statistical Significance and by Method of Weighting Studies

	Sign and Significance	(1)	(2)
Gradespan (total # of studies)		Unweighted Excluding KIPP Percentage of Studies (# of studies)	Weighted by varying weights from formal meta-analysis Excluding KIPP
Elementary	-/Significant	11% (2)	10
	-/Insignificant	11% (2)	13
(18 studies)	+/Insignificant	17% (3)	19
	+/Significant	61% (11)	58
Middle	-/Significant	11% (2)	11
	-/Insignificant	16% (3)	14
(19 studies)	+/Insignificant	11% (2)	10
	+/Significant	63% (12)	66
High	-/Significant	13% (2)	14
	-/Insignificant	31% (5)	32
(16 studies)	+/Insignificant	6% (1)	3
	+/Significant	50% (8)	51
Combined	-/Significant	40% (8)	42
Elementary/Middle	-/Insignificant	0% (0)	0
	+/Insignificant	20% (4)	17
(20 studies)	+/Significant	40% (8)	42
Elementary, Middle,	-/Significant	21% (11)	22
and Combined	-/Insignificant	8% (4)	7
Elementary/Middle	+/Insignificant	17% (9)	16
(52 studies)	+/Significant	54% (28)	55
Studies of All	-/Significant	23% (5)	26
Grades	-/Insignificant	5% (1)	5
	+/Insignificant	27% (6)	21
(22 studies)	+/Significant	45% (10)	48

Note: Each number indicates the percentage of regression results for the given weighting method and combination of grade spans that fit the stated category of sign and statistical significance. The numbers within each cell may not sum to 100 due to rounding.

Moreover, when considering statistical significance, the bulk of the estimates that are positive are also statistically significant, while the negative results are roughly mixed in significance—with half of the negative estimates being significant and half being insignificant for both subjects. Put differently, among the 14 elementary school-level studies that find statistically significant results in reading, 12 find positive effects and only two find negative effects. In the earlier formal meta-analysis section we discussed the role of the large negative elementary school reading results from Utah and Massachusetts in making the overall positive reading effect insignificant. With the exception of two large negative results bringing down the overall elementary reading effect, a generally similar story holds for elementary math—charters appear to be doing well. Among the 13 elementary school studies finding statistically significant math results, only 2 report negative results, while 11 report positive results.

The vote-counting explorations of the middle school math results reveal a very similar pattern to elementary school math and reading. The middle school level entry in column (1) of table 9 shows that 74 percent of the unweighted results are positive. Again, among the significant results, many more are positive than negative—among the 14 overall significant results, 12 are positive, and 2 are negative.

The results for middle school reading in column (1) of table 8 are remarkably similar overall, with 72 percent of results being positive and 28 percent negative. Among the 12 results here that are significant, three are negative. Charter middle schools appear to be performing more consistently in teaching math than in teaching reading, though they are doing well in both. This supports the formal meta-analysis results in which the overall effects are positive for both reading and math, but larger and only significant in the case of math.

Figures 11 through 18 show histograms of effect sizes, weighted by the meta-analytic weights. They arguably provide a clearer picture of the distribution of effect sizes than the earlier figures in the paper.

Figure 11. Histogram of Reading Effect Sizes From Elementary School Studies

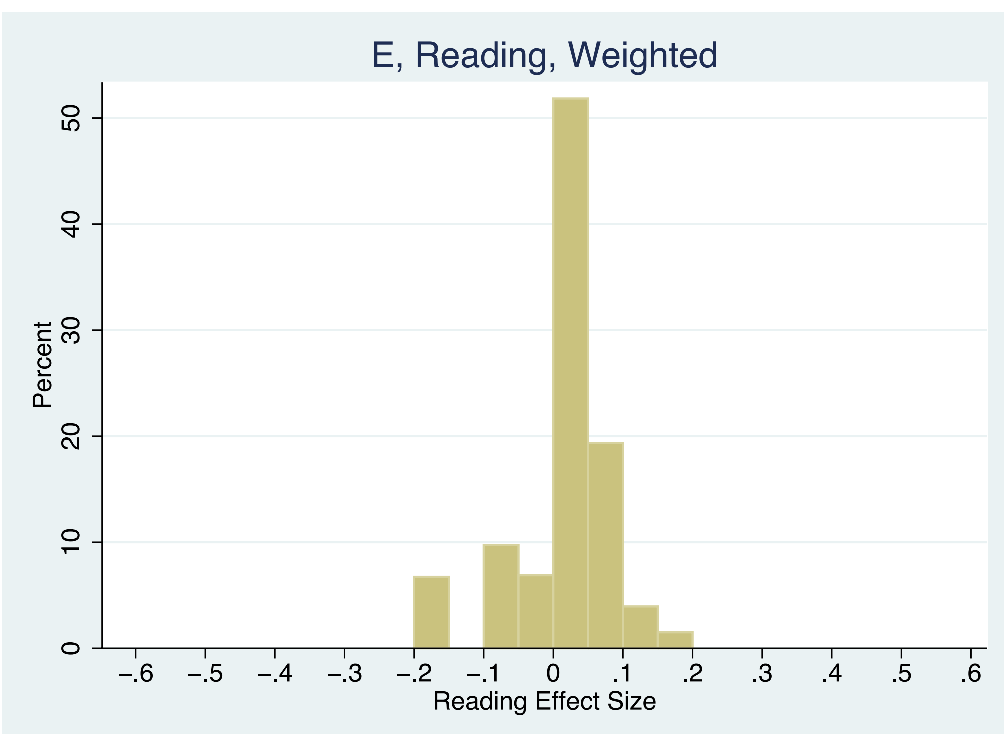


Figure 12. Histogram of Math Effect Sizes From Elementary School Studies

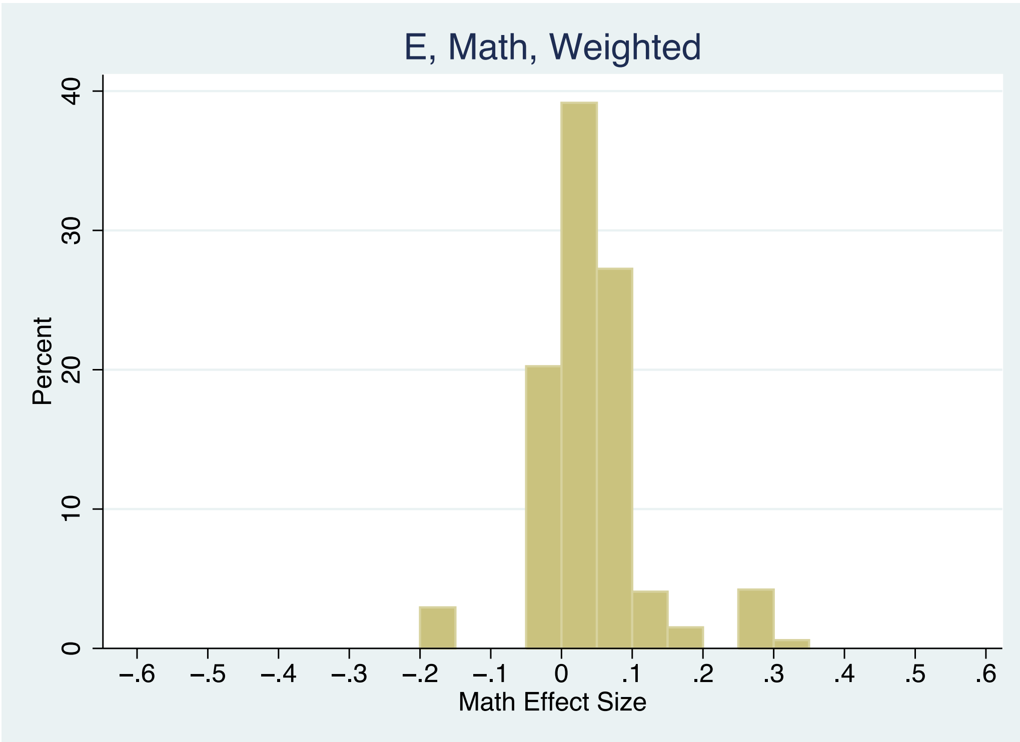


Figure 13. Histogram of Reading Effect Sizes From Middle School Studies

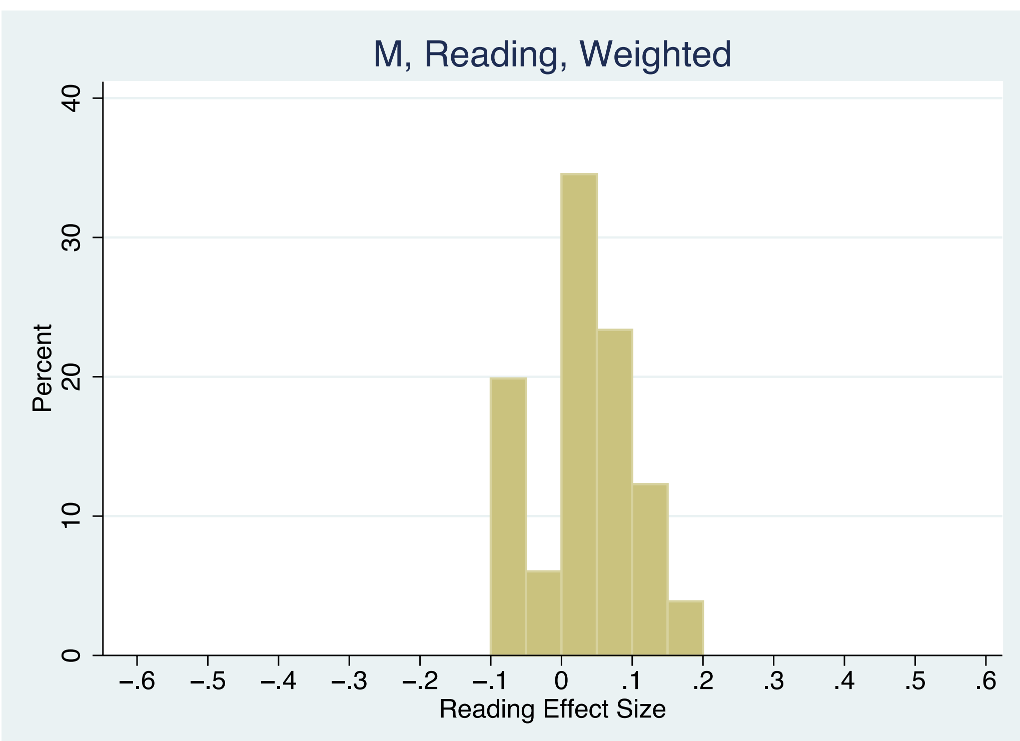


Figure 14. Histogram of Math Effect Sizes From Middle School Studies

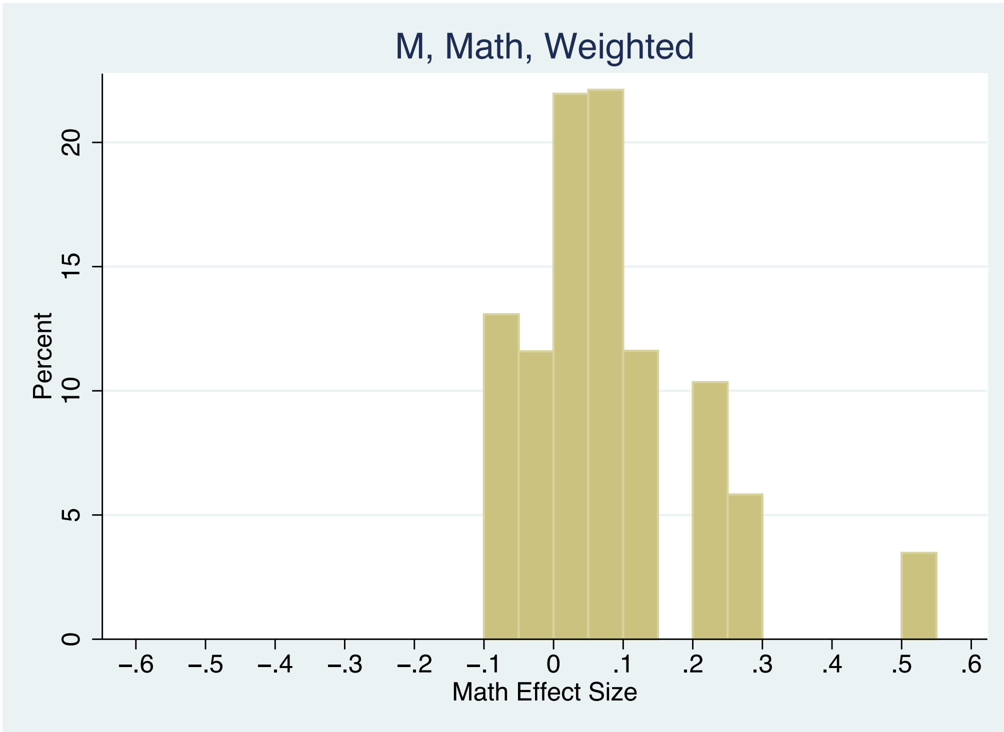


Figure 15. Histogram of Reading Effect Sizes From High School Studies

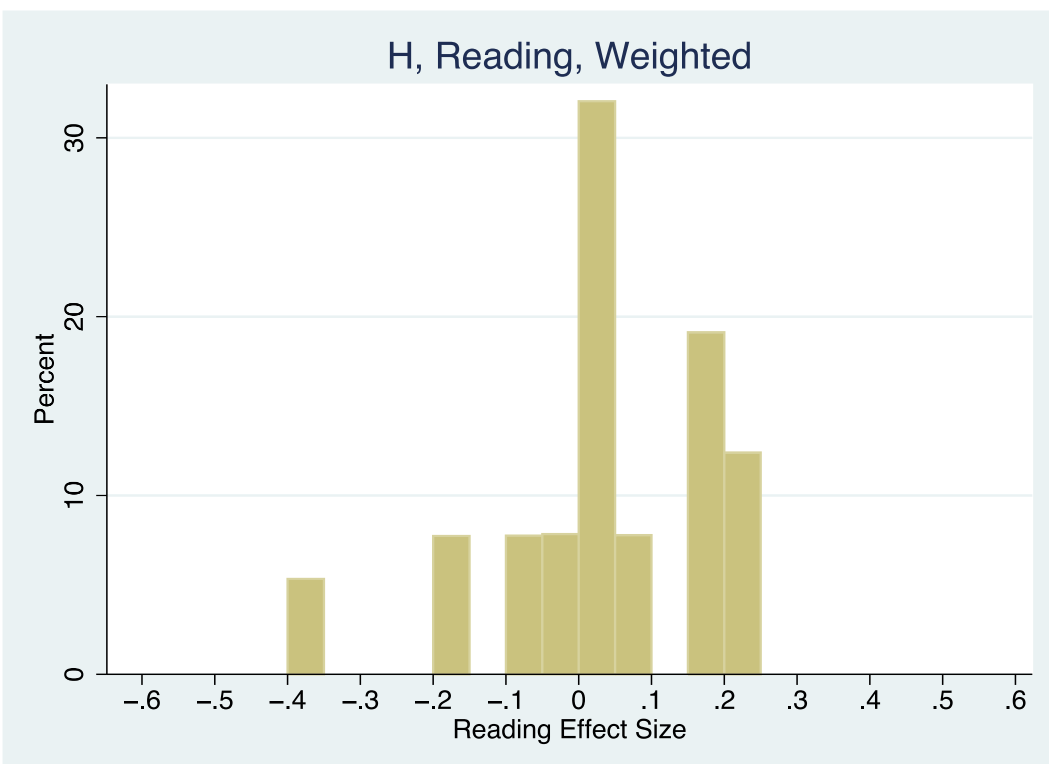


Figure 16. Histogram of Math Effect Sizes From High School Studies

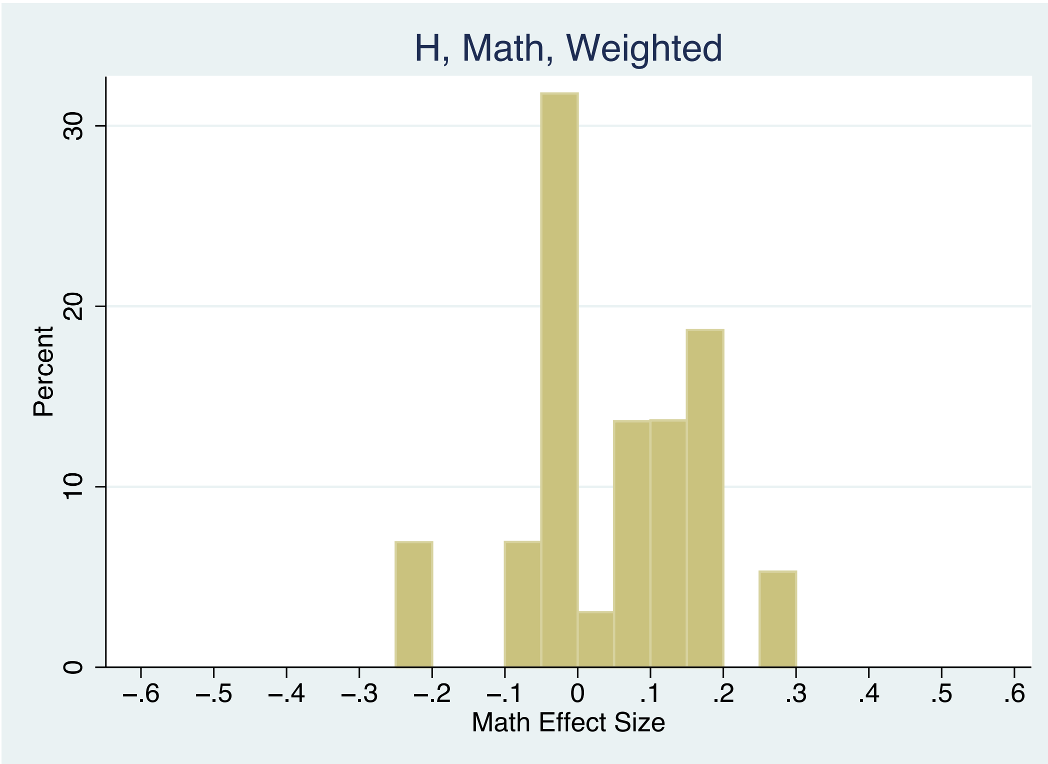


Figure 17. Histogram of Reading Effect Sizes From Studies That Combine Elementary and Middle Schools

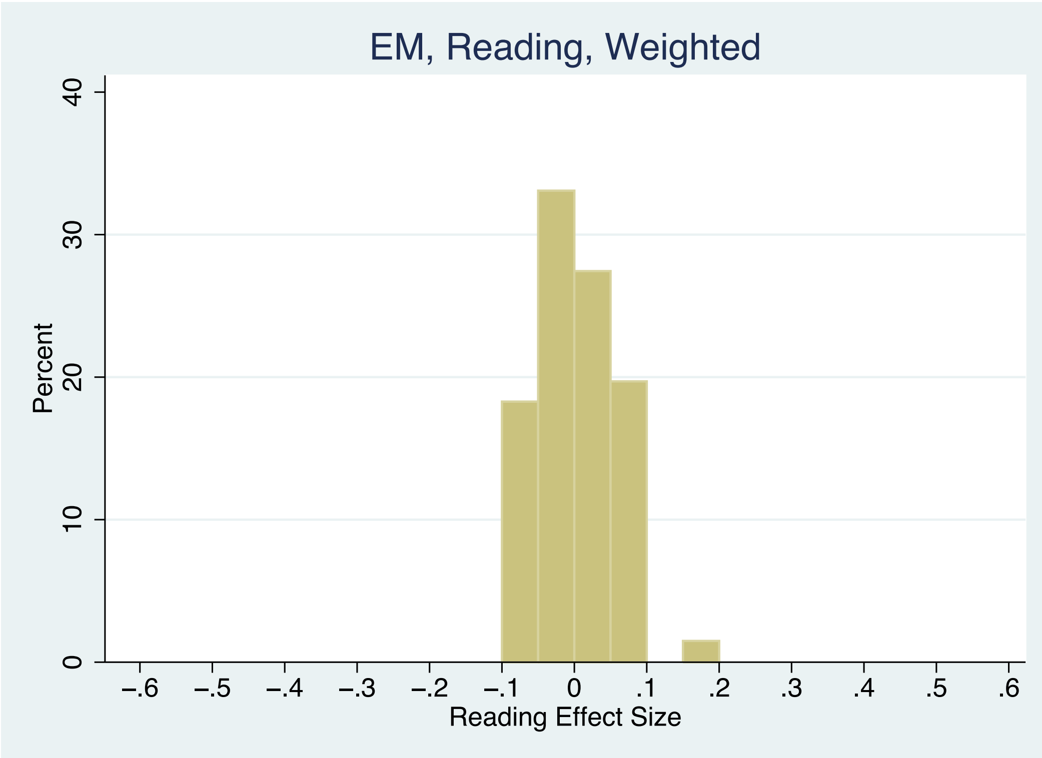
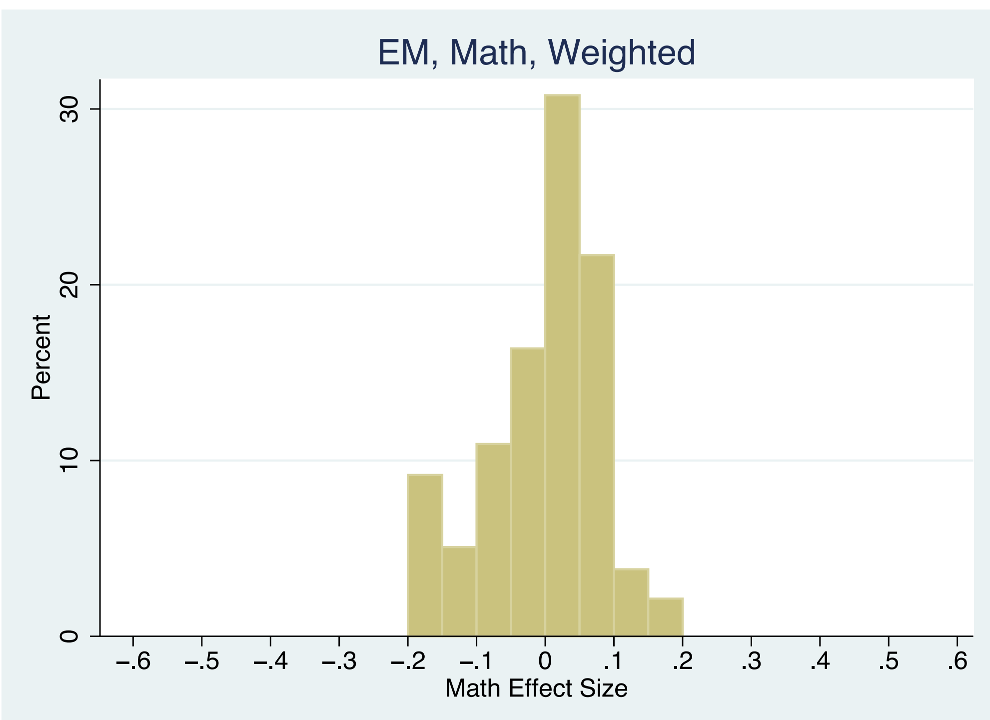


Figure 18. Histogram of Math Effect Sizes From Studies That Combine Elementary and Middle Schools



For both the elementary and middle school levels, the lower bounds are the same for reading and math results. The lower bound (defined as the lowest effect size in our sample of studies) for elementary school charter school effect sizes is -0.2 for both reading and math, while the lower bound for middle school charter school effect sizes is less negative, at -0.1 for both reading and math. The upper bounds for math, however, are larger in both the elementary and middle school levels. There is an estimate appearing in the 0.30 to 0.35 bin for elementary school math, and an estimate in the 0.50 to 0.55 bin for middle school math. The very large middle school estimate comes from a study in Massachusetts, using the lottery method. We discussed earlier that lottery methodology can only be used for oversubscribed schools, and therefore there is reason to believe a priori that these are estimates that apply to the most successful schools. Even without these two extreme estimates, we can see in figures 2 and 4 that the positive effects are greater in magnitude than the negative effects in both elementary and middle school math.

In our last report, we found that the histograms at the elementary school level looked generally favorable for charter schools, and more favorable for elementary school reading than for math. Because more evidence for positive charter school effects in elementary math has emerged since that report, and some evidence for negative effects in reading have emerged, the histograms plotting the distribution of effects shown in figures 11 and 12 now look fairly similar for reading and math, with perhaps more potential upside in math given the larger upper bound of the estimates. We also found in the last report that middle schools appeared to be serving charter school students well in both reading and math, with larger potential upsides in math. Figures 13 and 14 show that the same story holds true in this update. There are more estimates finding positive effects than negative effects, and the largest positive effects are larger in magnitude than the largest negative effects.

While there appears to be substantial improvement in charter high school performance, with several authors finding significant positive effects in recent years, we find again as we found in our last report that high school is still the single grade span at which results are most mixed. Column (1) of table 8 shows that 73 percent of estimates of high school reading effects are positive. However, we see that all of the remaining 27 percent of the estimates (which find negative effects) are in fact significant. The histogram in figure 15 shows that at least some of these negative estimates are also quite large in magnitude. There is one estimate in the -0.35 to -0.40 bin at this level, which comes from a CREDO study of Michigan. For high school reading, just as for middle school reading, of the 13 significant estimates, 4 are negative and 9 are positive.

High school math has fewer positive results than elementary or middle school math, but slightly more of the estimates are still positive than negative, as shown in figure 16. Table 9 shows that 56 percent of the high school math results are positive. While the rest of the results are negative, most of these negative results are not significant. Of the 10 studies finding significant results in high school math, 8 are positive and 2 are negative. Overall, as we also found in our last report, there appears to be more heterogeneity in charter high school effectiveness than in at the middle school or elementary school level.

Now we discuss the combined elementary/middle (EM) estimates. While we have five more estimates in this grade span than in our last report, the results from the updated vote-counting and histogram analysis are very similar to what we found previously. We again find that most of the studies in this level find significant results, with roughly equal numbers of negative results and positive results. In the last report we found that there is more variation in math effects than reading effects, and we see the same here. Figures 17 and 18 show a greater spread of math effects than reading effects: The reading estimates are mostly bounded between one-tenth of a standard deviation on either size of zero, while the math estimates stretch to two-tenths of a deviation in both the negative and positive direction.

If we combine the studies of elementary school students and the studies of middle school students with the studies of combinations of elementary and middle school students, we confirm the results above. Overall, there are many positive results found in elementary and middle school studies. Studies of combinations of elementary and middle school students are roughly balanced, finding positive and negative results. Therefore the unweighted E, M, and EM histograms show many positive estimates, some quite large, along with some negative results. We do not include these figures because they essentially replicate the information contained in the earlier figures.

Finally, we discuss those studies studying all grade spans. Similar to the changes in studies of the other grade spans, more new studies have found positive than negative results since our last report. Column (1) of the bottom row of tables 8 and 9 shows that around 72 percent of both reading and math results are positive. While the remaining estimates are negative and often significant, the histograms show that the lower bounds on these estimates are not large in magnitude. Figure 19 shows that the study finding the most negative estimate for reading is in the -0.1 to -0.05 bin. The picture looks more positive for math: Figure 20 shows that the study finding the most negative estimate is in the -0.05 to 0 bin (and therefore smaller than 5 percent of a standard deviation less than zero).

Figure 19. Histogram of Reading Effect Sizes From "All Grade" Studies That Combine Elementary, Middle, and High Schools

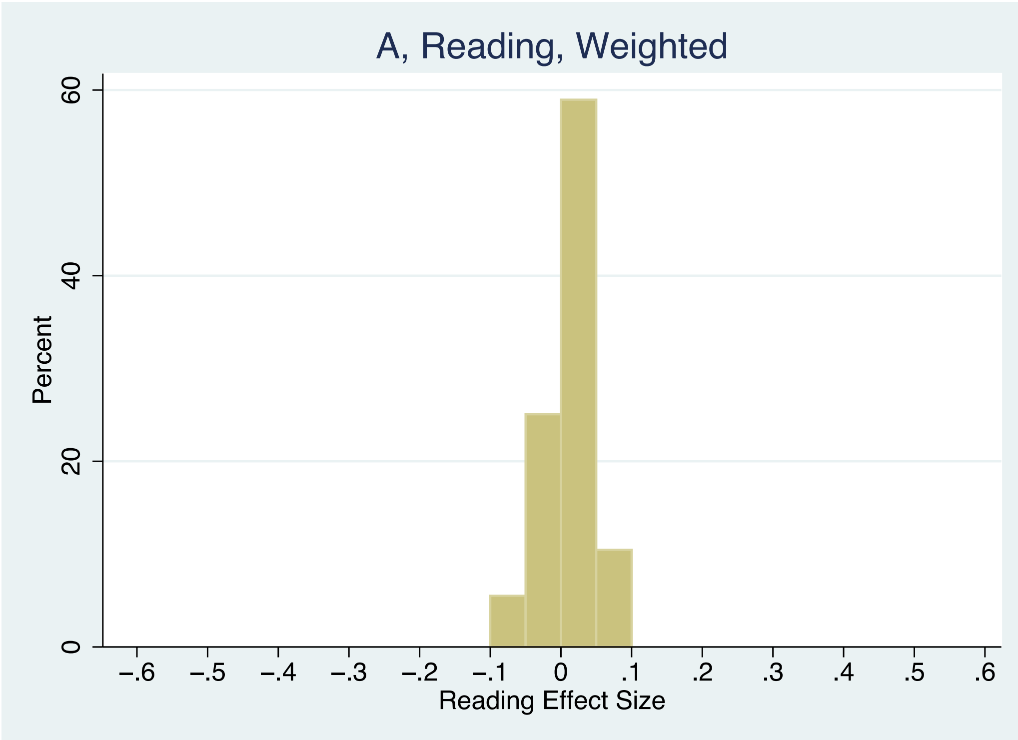
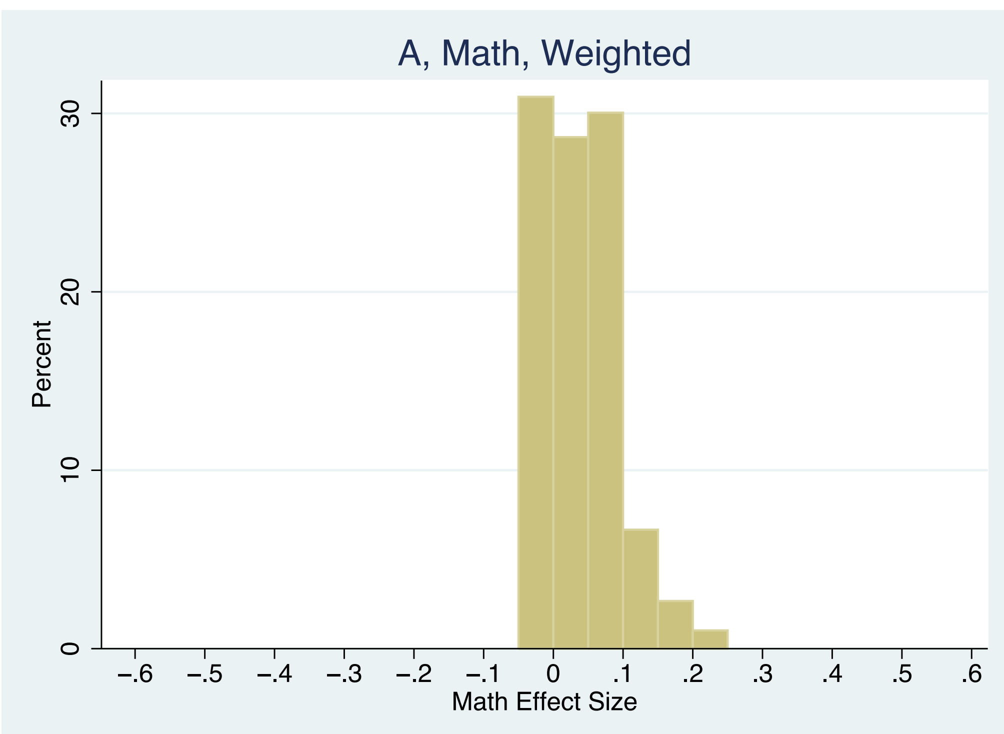


Figure 20. Histogram of Math Effect Sizes From "All Grade" Studies That Combine Elementary, Middle, and High Schools



Examining all of these results as separate parts of a whole, we conclude that, overall, charter schools appear to be serving students well, and better in math than in reading. The caveat here is that a substantial portion of studies that combine elementary and middle school students do find significantly negative results in both reading and math—35 percent of reading estimates are significantly negative, and 40 percent of math estimates are significantly negative. In general, there appears to be more variation in the results for math than reading, and more variation in results from studies at the high school level or studies that combine elementary and middle grade spans than in those studying elementary or middle schools on their own.

5. A Search for Variations Across Studies Related to Time, Grade Span, and Empirical Method

Given the growing number of studies, it becomes plausible that one could detect differences in impacts across studies due to the time the study covers, the grades examined, or even the method used by the empirical researchers. To do this we pooled all of the non-KIPP studies, across all grade spans, and then performed a type of statistical analysis called a metaregression.[18]

We tested three hypotheses about factors that might influence the effect size in a given study:

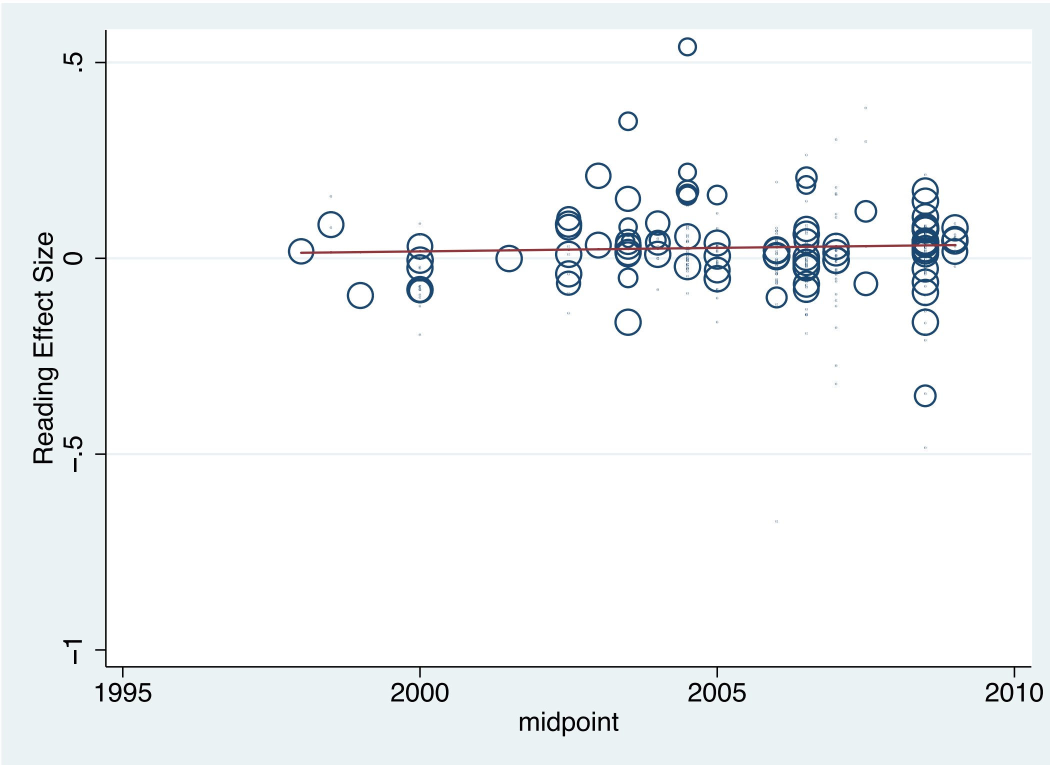
1. Is the time period, measured by the midpoint of the range of school years studied in a given report, related to the effect size?
2. Is the grade span studied linked to the effect size?
3. Is the statistical method used linked to the effect size?

Of these hypotheses the first is particularly relevant for policy because it addresses whether over time charter schools' impact on achievement has risen. We tested these hypotheses by regressing the effect sizes on three sets of explanatory variables individually and then together. Figures 21 and 22 show the fitted line between reading and math effect sizes and the midpoint of the years covered in the given study.[19] Although the slope was positive in both cases, suggesting a positive time trend, in neither case was the trend significantly different from zero.

---

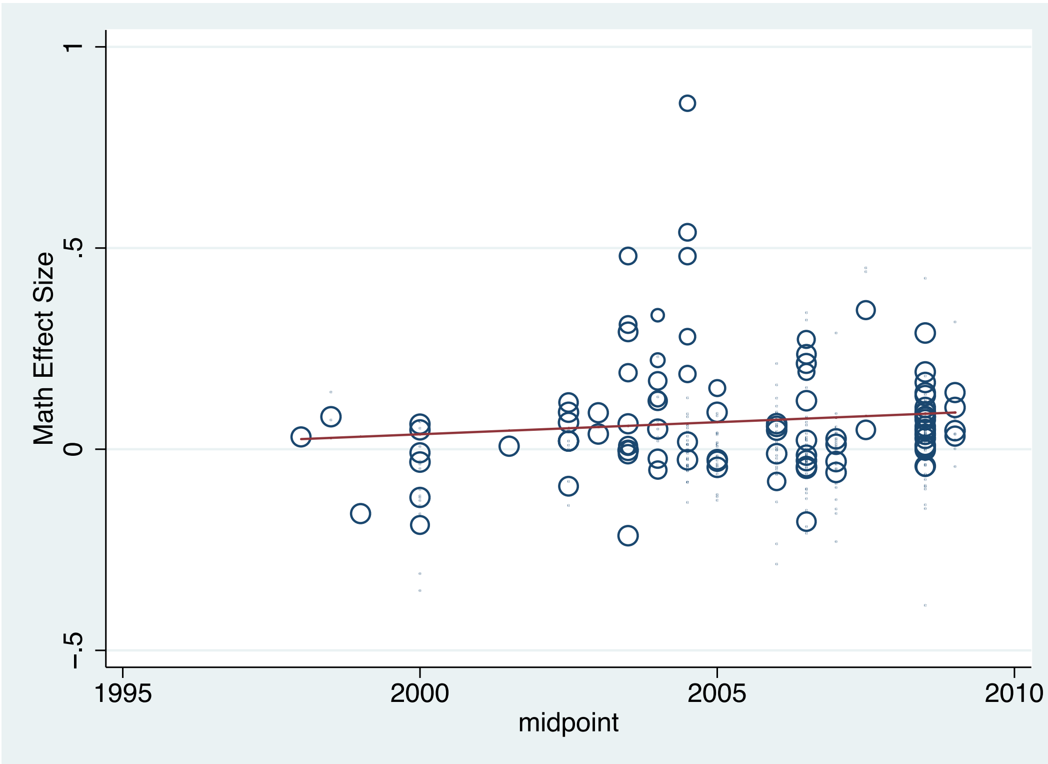
18. We avoided overcounting multiple estimates from the same study. This applied mostly to the most recent CREDO studies, which provide estimates not only for elementary, middle, and high schools, but also an overall estimate across all grades (or for elementary and middle grades in some papers). In cases such as these we used the results for the individual grade spans (elementary, middle, and high) rather than the overall estimate over multiple grade spans. The metaregression is a random-effects regression that assumes a two-part error term, the first being a homoscedastic error term, the variance of which reflects across study variation, and the second being an error term, the variance of which is equal to the square of the estimated standard error of the given effect size.
19. We coded start and end years of data using the calendar year in which a given school year ended. For example, if a study covered the school years 2007–2008 through 2009–2010, the midpoint is 2009.

Figure 21. Meta-Regression of Reading Effect Sizes at the Midpoint Year of Each Study



Note: The figure shows effect sizes plotted against the midpoint year of the study. Regression line is from a meta-regression of effect sizes on a constant and the midpoint year of the given study. The slope is not statistically significant at the 5 percent level. The size of the circle is proportional to the weight given to the estimate, which is inversely related to the total variance.

Figure 22. Meta-Regression of Math Effect Sizes at the Midpoint Year of Each Study



Note: The figure shows effect sizes plotted against the midpoint year of the study. Regression line is from a meta-regression of effect sizes on a constant and the midpoint year of the given study. The slope is not statistically significant at the 5 percent level. The size of the circle is proportional to the weight given to the estimate, which is inversely related to the total variance.

Tables 12 and 13 in the appendix show the regression results for reading and math respectively. Columns (1), (2), and (3) show results when we model effect sizes as a function of the midpoint of the years included in the study, the grade spans used in the study, and the statistical method used respectively. The rows list the explanatory variables: the midpoint of the years included in the study; dummy variables for studies of elementary schools, middle schools, high schools, and schools in all grades (with the comparison group being studies that combine elementary and middle schools); and dummy variables for studies based on student fixed effects, lotteries, and propensity score modeling (with the comparison group being the studies that use other methods). The studies that used other methods mostly consist of the CREDO virtual control method and one study that uses instrumental variables. Results do not change if we instead drop the single instrumental variable study. The model in column (4) includes all three sets of explanatory variables.

In the reading models, the only explanatory variable that rises to statistical significance is the indicator for middle school studies, which on average produce an effect size about 0.06 above that of the comparison group studies. However, in the final column (4) of appendix table 12, when we include all three sets of explanatory variables, the middle school indicator is no longer statistically significant.

In the math models shown in appendix table 13, three explanatory variables are statistically significant, both in the simpler models and in the combined model in column (4). As with reading models, middle school studies of math produce higher effect sizes than the omitted group, studies that combine elementary and middle schools. In addition, lottery-based studies and propensity score studies produce significantly higher estimates than do the "other method" studies. Each of these differences is quite large relative to the mean effect sizes shown in table 1. Middle school studies produce estimates of effect sizes that are 0.11 to 0.16 higher (depending on which model we estimate) than elementary/middle school studies. Lottery-based studies and propensity score studies produce effect sizes that are higher than the "other method" studies by 0.09 to 0.11 and 0.12 to 0.14 respectively. (In each case, the lower number in the ranges presented comes from the model in column (4) that combines all the explanatory variables.) Betts and Atkinson (2012) note the very high estimates from lottery-based studies and conjecture that this arises because only in areas where charter schools outperform traditional public schools are charter schools popular enough to be oversubscribed, and therefore use admission lotteries. They show three cases in which lottery-based and non-experimental estimates have been obtained for the same sets of schools. The results are similar although the latter are sometimes slightly lower.

The difference between results for the propensity score approach and the other methods, which consist mainly of the CREDO "virtual control record" approach, is somewhat puzzling, as the latter method appears to use a method somewhat similar to propensity score matching. One potential issue, mentioned earlier, is that the CREDO approach matches students using their current year test scores, which for charter students are endogenous outcomes. So if a charter school produces gains in achievement, students at that school will be matched with intrinsically stronger students in traditional public schools. On the other hand, an equally compelling argument is that the propensity score models and the CREDO studies have examined quite different sets of schools, and this alone could well explain observed differences.[20]

20. Davis and Raymond (2012) compare the CREDO approach to a more standard fixed effect approach using data from 14 states and two districts, and find broadly similar results. Although no formal comparison of the CREDO approach to a gold-standard lottery-based approach has been performed using a single data set, a CREDO (2010) analysis of charter schools in New York City comes to similar conclusions to a lottery-based study in that city by Hoxby, Murarka, and Kang (2009), with a similar estimated math effect and a reading effect that was about two-thirds the size of that obtained in the lottery-based study.

6. Outcomes Apart From Achievement

Accompanying the large literature we have reviewed above on charter schools' association with student achievement, there is a much smaller literature that examines the relation between attending a charter school and other outcomes, such as years of education completed and student behavior. There is little sense in performing a meta-analysis of the few papers in this literature, but a summary may still be useful.[21] Overall the studies appear to find positive effects of charter schools on non-achievement outcomes.

Educational Attainment

Lottery data are especially useful for analyzing college matriculation because we only observe this outcome once and thus cannot use models that rely on comparing a student over time as fixed effects models do. To date, three papers have used lottery methods to estimate the effects of charter school attendance on college enrollment.

Angrist et al. (2013) use lottery data from six Boston schools to examine a range of outcomes related to postsecondary success. The authors find that while overall postsecondary enrollment is similar among lottery winners and losers, there is a pronounced shift away from two-year colleges and toward four-year colleges among lottery winners. This effect is relatively large in magnitude, with charter attendees 17 percent more likely to be enrolled in a four-year college within 18 months of high school graduation than non-attendees.

The two other lottery studies of the impact of charter school attendance on college enrollment are from a single school in New York City and a single school in San Diego. Dobbie and Fryer (2013) use lottery data from Promise Academy in Harlem Children's Zone to find that lottery winners are 14 percent more likely to enroll in college than lottery losers. McClure et al. (2005) similarly find that lottery winners at the Preuss School at UCSD are more likely than lottery losers to report plans to enroll in a four-year college. Because a substantial portion of lottery losers did not respond to the survey used to measure college enrollment plans, the estimates of the size of the actual effect are not precise. The estimates range from lottery winners being 11.4 to 48.2 percent more likely than lottery losers to enroll in a four-year college, depending on assumptions made on non-respondent outcomes.[22]

21. For an extended discussion of this "non-achievement" literature, see Betts (2010) upon which this section is largely based. This section is very similar to our corresponding section from Betts and Tang (2011), but with the notable addition of the results in Dobbie and Fryer (2013) and Angrist et al. (2013).
22. One issue with the McClure et al. (2005) study is that it dropped a portion of the comparison group because the charter school expanded, and was able to enroll these students in the following year. This is not the approach that would be taken in a pure intent-to-treat study. Importantly, though, the authors show that the treatment group and the remaining comparison group is closely matched on baseline characteristics. Meanwhile, among the lottery winners, 90.3 percent were set to enroll in a four-year college in fall. Only 66.7% of respondents from the group of lottery losers planned to attend a four-year college in the fall, demonstrating a gap of about 23%. Just under two-thirds of students in the group that did not win the lottery replied to the survey. By assuming that none of the non-respondents or, alternatively, that all of these non-respondents, were intending to enroll in college, we obtain a range of 42.1% to 78.9% as the maximum range for the actual four-year college enrollment in this comparison group.

The remaining papers in this literature on educational attainment use non-lottery methods. Non-lottery methods may be regarded with more skepticism when used to study non-achievement outcomes than when they are used to study achievement outcomes. This is because while current evidence indicates that non-experimental methods can reliably replicate estimates derived from experimental methods in the test score literature when a baseline test score is included in the analysis, we cannot say the same for non-experimental methods of non-test outcomes. A baseline measure of educational attainment cannot be controlled for in these analyses because, as noted earlier, this outcome is only observed once per student.

Furgeson et al. (2012) use a matching method to study educational attainment outcomes in a select group of charter management organizations (CMOs). (CMOs are nonprofit organizations operating multiple charter schools.) The authors find that in two of the four CMOs with available data studied, college enrollment is substantially and significantly higher among charter students than the control group. The estimates are large: Charter students are 21 percent and 23 percent more likely to enroll in college than their respective matched non-charter counterparts. In the other two CMOs with available data studied, there is no significant effect on college enrollment. As might be expected, studies finding increases in college enrollment among charter students also often find increases in high school graduation rates. In the Furgeson et al. (2012) study, six CMOs had data available on graduation rates. In three of the six CMOs, charter students were significantly more likely to finish high school. The size of these estimates ranged from charter students being 12 to 23 percent more likely to graduate than their matched non-charter counterparts. In two of the six CMOs, there was no significant difference, and in one of the six, charter students were actually 22 percent *less* likely to graduate. The variation observed between schools in achievement outcomes also appears to carry over to non-achievement outcomes.

Booker et al. (2011) also find using a matching method that charter school students in Chicago and Florida are 8 to 10 percent more likely to attend college, and 7 to 15 percent more likely to graduate high school with a standard diploma. An update to this study, Booker et al. (2014), finds that these charter students in Florida are also more likely to stay enrolled in college, and earn more money annually at ages 23 to 25.

The papers discussed above also present a smattering of other findings, with varying statistical significance.[23] Again, the general picture that emerges is one suggestive of large positive impacts of charter schools on high school graduation and eventual college enrollment. It is important to note that this literature is still emerging, and currently covers only a limited number of geographic locations.

23. Angrist et al. (2013) find that lottery winners were more likely to pass the high school exit examination in Massachusetts, more likely to take an Advanced Placement (AP) exam, and scored higher on the AP Calculus exam. Lottery winners also scored higher on the SAT than lottery losers, but were not more likely to take the SAT. They were more likely to attend college overall (two- or four-year) but this effect is not significant. McClure et al. (2005) find that charter school attendees complete more college preparatory courses in high school.

Evidence on Attendance and Behavior

Imberman (2007) uses fixed-effect methods to study two behavioral outcomes in an anonymous large urban school district: students' attendance rates and suspensions from school (combined with more serious disciplinary actions). He finds significant reductions in student disciplinary infractions among those who attend charter high schools. While he finds no relation between charter school attendance and attendance rates in a baseline model, he does find a small positive relation between attending a charter two periods ago and attendance rates in the current period.

In addition to the college attendance results discussed earlier, Dobbie and Fryer (2013) also use lottery data to study impacts of winning a lottery on a number of behavioral outcomes at the Promise Academy in the Harlem Children's Zone. Lottery winners who were female were 12 percent less likely than lottery losers to become pregnant in their teens, and lottery winners who were male were 4 percent less likely to go to jail relative to males who lost the lottery.

This literature is obviously very small, but both papers find evidence that charter school attendance is associated with better noncognitive outcomes.

7. Conclusion

The overall tenor of our results is that charter schools are in some cases outperforming traditional public schools in terms of students' reading and math achievement, and in other cases performing similarly or worse. But there is stronger evidence of outperformance than underperformance, especially in math. Almost all of the variation across studies is likely to reflect true variation across locales in the average performance of charter schools. Our analysis of histograms also shows this variation clearly.

One conclusion that has come into sharper focus since our prior literature three years ago is that charter schools in most grade spans are outperforming traditional public schools in boosting math achievement. The average effect size has grown with the accumulation of three years of additional research. In the middle school studies, which produce the largest estimates, charter school students are predicted to gain 3.3 percentile points in a single year.

In contrast, our estimated reading impacts for charter schools, while almost always positive, do not reach statistical significance.

However, the reading results actually have more in common with the math results than meets the eye. The vote count analysis and histograms show that for reading and math alike a preponderance of effect sizes in the literature is positive and significant. (The important exception is the studies that combine elementary and middle schools, which show a quite even split between negative-and-significant versus positive-and-significant results.) The reason overall effect sizes for math are typically statistically significant, while the effect sizes for reading are never statistically significant, is due to a small number of studies that produce large negative effect sizes for reading.

One of the most important findings from our meta-analysis is the considerable heterogeneity in effect sizes across studies. Overall, our findings confirm that the impact of the charter sector on student outcomes varies considerably—especially across geographic areas. Urban areas account for strong positive effects.

It will always be the case that policymakers will want overall estimates of the average effect of charter schools on achievement, and this is perfectly understandable and reasonable. But to better understand which charter schools are outperforming or underperforming, policymakers deserve to see estimates of the effects of individual charter schools. With a few exceptions such as the lottery-based studies of a KIPP school in Lynn, Massachusetts, by Angrist et al. (2010), the study of the Promise Academy of the Harlem Children's Zone by Dobbie and Fryer (2010), and the Preuss School at UCSD by McClure et al. (2005), release of results on individual charter schools has not yet typically occurred. Academic journals may have little interest in publishing such detailed results. As suggested in Betts and Tang (2011), one alternative would be for a consortium of researchers knowledgeable in the field to begin building such a database, by vetting submissions of school-level findings, and including competently done value-added estimates into a database that would become publicly available. This database would not only serve a public purpose, but it would also allow for more nuanced meta-analyses of characteristics of charter schools that are truly making a positive or negative difference for student achievement.

Appendix. Details on Meta-Analysis

We assume that the effect of charter schools on achievement is not fixed across studies. Given that charter schools are afforded considerable freedom to experiment, and that the regulatory framework for charter schools varies across states, and surely across individual districts as well, it would seem untenable to make the alternative assumption that there is a single fixed impact of charter schools on achievement.[24]

In a random effects meta-analysis, we take a weighted average of the effect sizes across studies. If Y_i is the effect size for the i^{th} of k studies and W_i is the weight for each study, our overall estimated effect size M is :

$$(1) \qquad M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}$$

The weight for each study is the inverse of the sum of the within-study variance (based on the standard error) and an estimate of the true between-study variance, T^2:

$$(2) \quad W_i = \frac{1}{V_{Y_i} + T^2}$$

The between-studies variance estimate T^2 is based on a method of moments estimate of the variance of true effect sizes. Note that as T^2 becomes large relative to the average within-study variance estimate, then we will tend towards equal-weighting across studies, whereas as T^2 becomes relatively small the weights can become highly unequal with heavier weight given to studies with the lowest sampling variance.

We report the I^2 statistic introduced by Higgins et al. (2003), which provides an estimate of the percentage of the variation in effect sizes that reflects true underlying variation.

24 For a review of the random-effects approach to meta-analysis and measures of heterogeneity, see Borenstein et al. (2009), chapters 12 through 16.

Table 10. Details on the Studies Used in Any of Our Approaches

Authors	Year Published	Name of State or City	First Year of Data	Final Year of Data	Grade Span(s) Studied	Included in Meta-Analysis of Effect Size, Vote Counting Study, and Histograms
Abdulkadiroglu et al.	2009	Boston	2002	2007	E, M, H	E, M, H
Angrist et al.	2010	Boston (1 KIPP school)	2006	2009	M	M
Angrist, Pathak, and Walters	2013	Massachusetts	2002	2011	M	M
Ballou et al.	2006	Idaho	2003	2005	E, M, H, A	E, M, H, A
Betts et al.	2005	San Diego	1998	2002		E, M, H
Betts, Tang, and Zau	2010	San Diego	2001	2006	E, M, H, A	E, M, H, A
Bifulco and Ladd	2006	North Carolina	1996	2002	EM	EM
Booker et al.	2004	Texas	1995	2002	EM	EM
Buddin and Zimmer	2003	California	1998	2002	E	E
CREDO	2009a	National	2001	2008	E, M, H	E, M, H
CREDO	2009a	Arizona, Arkansas, California, Chicago, Colorado (Denver), DC, Florida, Georgia, Massachusetts, Minnesota, Missouri, New Mexico, North Carolina, Ohio, Texas	varies	varies	EM (9 locations), A (7 locations)	EM (9 locations), A (7 locations)

CREDO	2013	Louisiana, Massachusetts, Michigan, New York City	varies	varies	E, M, H, A (LA, MA, NYC) E, M, EM (MI)	E, M, H, A (MA, NYC) E, M, EM (MI)
CREDO	2012	Indiana, New Jersey	varies	varies	E, M, H, A (IN) E, M, EM (NJ)	E, M, H, A (IN) E, M, EM (NJ)
CREDO	2011	Pennsylvania	varies	varies	E, M, EM	E, M, H, A (IN) E, M, EM (NJ)
Dobbie and Fryer	2009	NYC (1 school, Promise Academy in Harlem Children's Zone)	2004	2009	E, M	E, M
Furgeson et al.	2012	CMOs multiple states	varies	varies	M	M
Gleason et al.	2010	National (29 schools)	2004	2008	M	M
Gronberg and Jansen	2005	Texas	2003	2004	M, H	M, H
Hoxby and Murarka	2007	NYC	2004	2006	E	E
Hoxby, Murarka, and Kang	2009	NYC	2000	2008	EM	EM
Hoxby and Rockoff	2004	Chicago	2001	2004	E, M	E, M
Imberman	2007	Anonymous	1995	2005	A	A
McClure et al.	2005	San Diego	2003	2004	H	H
Miron et al.	2007	Delaware	2000	2005	E, M, H, A	E, M, H, A
Ni and Rorrer	2012	Utah	2004	2009	E	E
Nichols and Özek	2010	DC	2001	2009	EM	EM
Nicotera, Mendiburo, and Berends	2009	Indianapolis	2002	2006	A	A

Nisar	2012	Milwaukee	2006	2009	EM	EM
Solmon, Paark, and Garcia	2001	Arizona	1997	1999	A	A
Sass	2006	Florida	2000	2003	A	A
Tuttle et al.	2010	Anonymous (22 KIPP schools)	varies	varies	M	(Not used: Superseded by Tuttle et al. 2013)
Tuttle et al.	2013	KIPP multiple states	2002	2011	M	M
Witte, Wolf, Carlson, and Dean	2012	Milwaukee	2007	2011	A	A
Woodworth et al.	2008	Bay Area (3 KIPP schools)	2003	2005	M	M
Zimmer et al.	2009	Chicago, Colorado (Denver), Milwaukee, Ohio, Philadelphia, San Diego, Texas	varies	varies	EM (3 locations), A (4 locations)	EM (3 locations), A (4 locations)

Note: E, M, H and A stand for analyses of elementary, middle, high schools, and all grades, respectively, and EM stands for combined elementary and middle.

Table 11. Author and Year of Study Referenced by Study ID Label (for Cases with More Than One Study)

Study ID Label used in Figures	Author	Year Published
Chicago-1	Hoxby and Rockoff	2005
Chicago-2	Zimmer et al.	2009
Chicago-3	CREDO	2009a
Colorado (Denver)-1	Zimmer et al.	2009
Colorado (Denver)-2	CREDO	2009a
DC-1	Sass	2006
DC-2	Nichols and Ozek	2010
Florida-1	Sass	2006
Florida-2	CREDO	2009a
Massachusetts-1	CREDO	2009a
Massachusetts-2	Angrist, Pathak, and Walters	2013
Massachusetts-3	CREDO	2013
Milwaukee-1	Zimmer et al.	2009
Milwaukee-2	Nisar	2012
Milwaukee-3	Witte et al.	2012
National-1	Gleason et al.	2010
National-2	CREDO	2009a
National-3	CREDO	2013
New York City-1	Hoxby and Murarka	2007
New York City-2	Hoxby, Murarka, and Kang	2009
New York City-3	Dobbie and Fryer	2010
New York City-4	CREDO	2013
Ohio-1	Zimmer et al.	2009

Ohio-2	CREDO	2009a
San Diego-1	McClure et al.	2005
San Diego-2	Betts et al.	2005
San Diego-3	Betts et al.	2010
San Diego-4	Zimmer et al.	2009
Texas-1	Gronberg and Jansen	2005
Texas-2	Booker et al.	2005
Texas-3	Zimmer et al.	2009
Texas-4	CREDO	2009a

Table 12. Meta-Regression Estimates for Determinants of Reading Effect Sizes

	(1)	(2)	(3)	(4)
VARIABLES				
Elementary		0.0275		0.0173
		(0.0312)		(0.0321)
Middle		0.0643*		0.0383
		(0.0290)		(0.0333)
High		0.0449		0.0266
		(0.0333)		(0.0352)
All Grades		0.0193		0.0176
		(0.0290)		(0.0294)
Midpoint Year of Study	0.00180			-0.000603
	(0.00335)			(0.00380)
Fixed			-0.0451	-0.0379
			(0.0256)	(0.0315)
Lottery			0.0459	0.0368
			(0.0312)	(0.0340)
Propensity			0.0409	0.0368
			(0.0327)	(0.0388)
Constant	-3.578	-0.00452	0.0259*	1.216
	(6.715)	(0.0216)	(0.0117)	(7.631)
Observations	92	93	93	92

Notes: Standard errors in parentheses. * $p < 0.05$. ** $p < 0.01$.

Table 13. Meta-Regression Estimates for Determinants of Math Effect Sizes

	(1)	(2)	(3)	(4)
VARIABLES				
Elementary		0.0612		0.0478
		(0.0443)		(0.0438)
Middle		0.162**		0.110*
		(0.0407)		(0.0447)
High		0.0626		0.0372
		(0.0463)		(0.0465)
All Grades		0.0440		0.0428
		(0.0415)		(0.0405)
Midpoint Year of Study	0.00601			0.00603
	(0.00510)			(0.00522)
Fixed			-0.0430	-0.00528
			(0.0351)	(0.0401)
Lottery			0.112*	0.0928*
			(0.0435)	(0.0454)
Propensity			0.138**	0.116*
			(0.0450)	(0.0513)
Constant	-11.98	-0.00419	0.0498**	-12.09
	(10.22)	(0.0310)	(0.0170)	(10.46)
Observations	96	97	97	96

Notes: Standard errors in parentheses. * $p < 0.05$. ** $p < 0.01$.

References

Abdulkadiroglu, Atila, Joshua D. Angrist, Sarah R. Cohodes, Susan M. Dynarski, Jon Fullerton, Thomas J. Kane, and Parag A. Pathak. 2009. *Informing the Debate: Comparing Boston's Charter, Pilot and Traditional Schools.* Boston, MA: The Boston Foundation.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2010. "Who Benefits From KIPP?" NBER Working Paper 15740. Cambridge, MA: National Bureau of Economic Research.

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. 2013. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry and Choice." NBER Working Paper 19275. Cambridge, MA: National Bureau of Economic Research.

Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters. 2013. "Explaining Charter School Effectiveness." *American Economic Journal: Applied Economics* 5 (4): 1–27.

Ashenfelter, Orley. 1978. "Estimating the Effects of Training Programs on Earnings." *Review of Economics and Statistics* 60: 47–57.

Ballou, Dale, Bettie Teasley, and Tim Zeidner. 2006. "Charter Schools in Idaho." Nashville, TN: National Center on School Choice. Paper prepared for the National Conference on Charter School Research, Vanderbilt University, September 29. http://www.vanderbilt.edu/schoolchoice/conference/papers/Ballouetal_2006-DRAFT.pdf.

Betts, Julian R. 2009. "The Competitive Effects of Charter Schools on Traditional Public Schools." In *Handbook of Research on School Choice*, edited by Mark Berends, Matthew G. Springer, Dale Ballou, and Herbert Walberg, chap. 11. New York: Routledge.

Betts, Julian R. 2010. "The Effect of Attending Charter Schools on Achievement, Educational Attainment and Behavioral Outcomes: A Review", Chapter 4 in Paul T. Hill and Julian R. Betts (Eds.), **Taking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools**, Lanham, MD: Rowman & Littlefield Publishers, Inc.

Betts, Julian R., and Richard C. Atkinson. 2012. "Better Research Needed on the Impact of Charter Schools." *Science* 335 (6065): 171–172.

Betts, Julian R., Lorien A. Rice, Andrew C. Zau, Y. Emily Tang, and Cory R. Koedel. 2005. *Does School Choice Work? Effects on Student Integration and Achievement.* San Francisco: Public Policy Institute of California.

Betts, Julian R. and Y. Emily Tang. 2008. "Value-Added and Experimental Studies of the Effect of Charter Schools on Student Achievement: A Literature Review," National Charter School Research Project, Seattle, WA: Center on Reinventing Public Education.

Betts, Julian R. and Y. Emily Tang. 2011. "The Effect of Charter Schools on Student Achievement: A Meta-Analysis of the Literature," National Charter School Research Project, Seattle, WA: Center on Reinventing Public Education.

Betts, Julian R., Y. Emily Tang, and Andrew C. Zau. 2010. "Madness in the Method? A Critical Analysis of Popular Methods of Estimating the Effect of Charter Schools on Student Achievement." In *Taking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools*, edited by Paul T. Hill and Julian R. Betts, chap. 2. Lanham, MD: Rowman & Littlefield.

Bifulco, Robert, and Helen F. Ladd. 2006. "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance and Policy* 1 (1): 50–90. http://www.mitpressjournals.org/toc/edfp/1/1.

Booker, Kevin, Scott M. Gilpatric, Timothy J. Gronberg, and Dennis W. Jansen. 2004. "Charter School Performance in Texas." Unpublished paper, last modified May 25. Knoxville: University of Tennessee. http://web.utk.edu/~sgilpatr/charterperf.pdf.

Booker, Kevin, Tim R. Sass, Brian Gill, and Ron Zimmer. 2011. "The Effects of Charter High Schools on Educational Attainment." *Journal of Labor Economics* 29 (2): 377–415.

Booker, Kevin, Brian Gill, Tim Sass, and Ron Zimmer. 2014. "Charter High Schools' Effects on Long-Term Attainment and Earnings." Mathematica Policy Research Working Paper. http://www.mathematica-mpr.com/~/media/publications/PDFs/education/charter_long-term_wp.pdf

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester, United Kingdom: John Wiley and Sons.

Buddin, Richard and Ron Zimmer. 2003. "Appendix C: Academic Outcomes," in Charter School Operations and Performance: Evidence from California. Santa Monica, CA: RAND Corporation.

Charter School Achievement Consensus Panel. 2006. "Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines." National Charter School Research Project White Paper Series, No. 2. Seattle, WA: Center on Reinventing Public Education.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. "How and Why Do Teacher Credentials Matter for Student Achievement?" NBER Working Paper 12828. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w12828.

CREDO. 2009a. "Multiple Choice: Charter School Performance in 16 States." Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2009b. “Fact vs. Fiction: An Analysis of Dr. Hoxby’s Misrepresentation of CREDO’s Research.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2009c. “CREDO Finale to Hoxby’s Revised Memorandum.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2011.” Charter School Performance in Pennsylvania.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2010. “Charter School Performance in New York City.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2012a. “Charter School Performance in Indiana.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2012b. “Charter School Performance in New Jersey.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2013a. “Charter School Performance in Massachusetts.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2013b. “Charter School Performance in Michigan.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2013c. “Charter School Performance in New York City.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

CREDO. 2013d. “Charter School Performance in Louisiana.” Stanford, CA: CREDO. Available at http://credo.stanford.edu.

Davis, Devora H., and Margaret E. Raymond. 2012. “Choices for Studying Choice: Assessing Charter School Effectiveness Using Two Quasi-Experimental Methods.” *Economics of Education Review* 31: 225–236.

Dobbie, Will, and Roland Fryer Jr. 2010. “Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence From the Harlem Children’s Zone.” Working Paper, Harvard University, May.

Dobbie, Will, and Roland Fryer Jr. 2013. “The Medium-Term Impacts of High-Achieving Charter Schools on Non-Test Score Outcomes.” NBER Working Paper 19581. Cambridge, MA: National Bureau of Economic Research.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Bing-ru Teh, Natalya Verbitsky-Savitz, Melissa Bowen, Allison Demeritt, Paul Hill, and Robin Lake. 2012. "Charter-School Management Organizations: Diverse Strategies and Diverse Student Impacts." Princeton, NJ & Seattle, WA: Mathematica Policy Research and Center on Reinventing Public Education.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. 2010. "The Evaluation of Charter School Impacts: Final Report." NCEE 2010-4029. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gronberg, Timothy J., and Dennis W. Jansen. 2005. "Texas Charter Schools: An Assessment in 2005." Austin: Texas Public Policy Foundation. http://www.texaspolicy.com/pdf/2005-09-charterschools-rr.pdf.

Hanushek, Eric A., John F. Kain, Steven G. Rivkin, and Gregory F. Branch. 2007. "Charter School Quality and Parental Decision Making With School Choice." *Journal of Public Economics* 91: 823–848.

Heckman, James, Robert LaLonde, and Jeff Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David E. Card, 3A: 1865–2097. Amsterdam: North Holland.

Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. London: Academic Press.

Higgins, J., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. "Measuring Inconsistency in Meta-Analyses." *British Medical Journal* 327 (7414): 557–560.

Hoxby, Caroline M. 2009. "A Statistical Mistake in the CREDO Study of Charter Schools." Unpublished paper. Palo Alto, CA: Stanford University. http://credo.stanford.edu/reports/memo_on_the_credo_study%20II.pdf.

Hoxby, Caroline M., and Sonali Murarka. 2007. "New York City's Charter Schools Overall Report." Cambridge, MA: New York City Charter Schools Evaluation Project. http://users.nber.org/~schools/charterschoolseval/nyc_charter_schools_report_july2007.pdf.

Hoxby, Caroline M., Sonali Murarka, and Jenny Kang. 2009. "How New York City's Charter Schools Affect Achievement." Cambridge, MA: New York City Charter Schools Evaluation Project. http://users.nber.org/~schools/charterschoolseval/how_NYC_charter_schools_affect_achievement_sept2009.pdf.

Hoxby, Caroline M., and Jonah E. Rockoff. 2004. "The Impact of Charter Schools on Student Achievement." Unpublished paper, last modified March 2005. Cambridge, MA: New York City Charter Schools Evaluation Project. http://www.rand.org/content/dam/rand/www/external/labor/seminars/adp/pdfs/2005hoxby.pdf.

Imberman, Scott. 2007. "Achievement and Behavior in Charter Schools: Drawing a More Complete Picture." Unpublished paper, last modified October 30. University of Houston. http://www.ncspe.org/publications_files/OP142.pdf.

McClure, Larry, Betsy Strick, Rachel Jacob-Almeida, and Christopher Reicher. 2005. "The Preuss School at UCSD: School Characteristics and Students' Achievement." University of California, San Diego, The Center for Research on Educational Equity, Assessment and Teaching Excellence. http://create.ucsd.edu/Research_Evaluation/PreussReportDecember2005.pdf.

Miron, Gary, Anne Cullen, Brooks Applegate, and Patricia Farrell. 2007. "Evaluation of the Delaware Charter School Reform: Year One Report." Kalamazoo, MI: The Evaluation Center, Western Michigan University. http://www.doe.k12.de.us/files/pdf/sbe_decseval.pdf.

Ni, Yongmei, and Andrea K. Rorrer. 2012. "Twice Considered: Charter Schools and Student Achievement in Utah." *Economics of Education Review* 31 (5): 835–849.

Nichols, Austin, and Umut Özek. 2010. "Public School Choice and Student Achievement in the District of Columbia." Working Paper 53. Washington, DC: National Center for Analysis of Longitudinal Data. http://www.edweek.org/media/publicschoolchoiceandstudentachievement-calder.pdf.

Nicotera, Anna, Maria Mendiburo, and Mark Berends. 20011. "Charter School Effects in Indianapolis," in School Choice and School Improvement, edited by Berends, Mark, Marisa Cannata, and Ellen B. Goldring, Cambridge, Mass.: Harvard Education Press pp. 35-50.

Nisar, Hiren. 2012. "Do Charter Schools Improve Student Achievement?" New York: National Center for the Study of Privatization in Education. http://www.ncspe.org/publications_files/OP216.pdf.

Sass, Tim R. 2006. "Charter Schools and Student Achievement in Florida." *Education Finance and Policy* 1 (1): 91–122. http://www.mitpressjournals.org/toc/edfp/1/1.

Solmon, Lewis C., Kern Paark, and David Garcia. 2001. "Does Charter School Attendance Improve Test Scores? The Arizona Results." Phoenix, AZ: Goldwater Institute's Center for Market-Based Education. http://www.goldwaterinstitute.org/pdf/materials/111.pdf.

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

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

Witte, John F., Patrick J. Wolf, Deven Carlson, and Alicia Dean. 2012. "Milwaukee Independent Charter Schools Study: Final Report on Four-Year Achievement Gains." SCDP Evaluation Report #31. Fayetteville, AR: School Choice Demonstration Project. Available at http://www.uaedreform.org/milwaukee-parental-choice-program-evaluation/.

Witte, John F., Patrick J. Wolf, Alicia Dean, and Deven Carlson. 2011. "Milwaukee Independent Charter Schools Study: Report on Two- and Three-Year Achievement Gains." SCDP Evaluation Report #25. Fayetteville, AR: School Choice Demonstration Project. http://files.eric.ed.gov/fulltext/ED518594.pdf.

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

Zimmer, Ron, Richard Buddin, Derrick Chau, Glenn Daley, Brian Gill, Cassandra Guarino, Laura Hamilton, Cathy Krop, Dan McCaffrey, Melinda Sandler, and Dominic Brewer. 2003. "Charter School Operations and Performance: Evidence from California." Santa Monica, CA: Rand. http://www.rand.org/publications/MR/MR1700/.

Zimmer, Ron, Brian Gill, Kevin Booker, Stephane Lavertu, Tim R. Sass, and John Witte. 2009. "Charter Schools in Eight States: Effects on Achievement, Attainment, Integration, and Competition." Santa Monica, CA: Rand.